Joachim Melcher

Process Measurement in Business Process Management

Theoretical Framework and Analysis of Several Aspects

KIT Scientific Publishing

Joachim Melcher

Process Measurement in Business Process Management

Theoretical Framework and Analysis of Several Aspects

Process Measurement in Business Process Management

Theoretical Framework and Analysis of Several Aspects

by
Joachim Melcher

KIT Scientific Publishing

Dissertation, Karlsruher Institut für Technologie
Fakultät für Wirtschaftswissenschaften
Tag der mündlichen Prüfung: 8. Februar 2011
Referenten: Prof. Dr. Detlef Seese, Prof. Dr. Andreas Oberweis

Impressum

Karlsruher Institut für Technologie (KIT)
KIT Scientific Publishing
Straße am Forum 2
D-76131 Karlsruhe
www.ksp.kit.edu

KIT – Universität des Landes Baden-Württemberg und nationales
Forschungszentrum in der Helmholtz-Gemeinschaft

KIT Scientific Publishing 2012
Print on Demand

ISBN 978-3-86644-789-9

Process Measurement in Business Process Management

Theoretical Framework and Analysis of Several Aspects

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für
Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Inform. Joachim Melcher

Tag der mündlichen Prüfung: 8. Februar 2011
Referent: Prof. Dr. Detlef Seese
Korreferent: Prof. Dr. Andreas Oberweis
2011 Karlsruhe

ACKNOWLEDGMENTS

This thesis was written during my occupation as research assistant at the *Institut für Angewandte Informatik und Formale Beschreibungsverfahren* (Institute for Applied Informatics and Formal Description Methods) at the *Universität Karlsruhe (TH)* and the *Karlsruher Institut für Technologie (KIT)* (Karlsruhe Institute of Technology [KIT])[1]. Many people have contributed to the success of the thesis.

First of all, I want to thank Professor Detlef Seese for the supervision of this thesis and his willingness to let me choose this topic. So, I was able to transfer knowledge acquired during my studies of computer science at the *Universität Karlsruhe (TH)* in the area of empirical software engineering (Professor Walter F. Tichy).

Furthermore, I am indebted to the following persons and organizations for their contribution to the thesis:

Agnes Koschmider for the possibility to conduct experiment 1 in the "Workflow Management" lecture (Chapter 6), Markus Kress for many fruitful discussions about different aspects of the thesis, Roland Küstermann for his technical assistance with Apache Tomcat (Chapter 7), Christine Melcher for procuring some publications from libraries in Heidelberg and Mannheim, Jan Mendling for his cooperation in experiment 2 (Chapter 6), Sanaz Mostaghim for her fruitful remarks on heatmaps (Chapter 5), Kerstin Schmidt for her technical assistance with the online questionnaire for experiment 2 (Chapter 6), Oliver Schöll for his administrative support, Hajo A. Reijers for his cooperation in experiment 2 (Chapter 6), Jana Weiner for her administrative support, the students from the *Humboldt-Universität zu Berlin*, the *Technische Universiteit Eindhoven* and the *Universität Karlsruhe (TH)* who participated in the experiments (Chapters 6 and 7), the anonymous reviewers for their fruitful remarks on the papers which this thesis is partially based upon and—finally—the *Deutsche Forschungsgemeinschaft* (German Research Foundation) for the financial support of my SYNASC 2008 conference trip to Timişoara (Romania).

Additionally, my thanks go to all my colleagues from the research group (Časlav Božić, Hagen Buchwald, Tobias Dietrich, Matthes Elstermann, Jörg Janning, Markus Kress, Roland Küstermann, Andreas Mitschele, Amir Safari, Oliver Schöll and Andreas Vogel) for the pleasant and cooperative working atmosphere, our weekly lunches and the exchange of experiences and information about writing a PhD thesis.

I also want to thank the colleagues from the other research groups of our institute, our administrative and technical staff, our tutors, Kerstin Schmidt as well as Frederic Toussaint and the whole CIP pool team for their support in the daily work as research assistant.

1 The KIT was founded on October 1, 2009 by a merger of the *Universität Karlsruhe (TH)* and the *Forschungszentrum Karlsruhe* (Research Center Karlsruhe).

Furthermore, I want to express my gratitude to the members of the university's debating club *Debatte Karlsruhe e. V.* and judo club. They offered me the possibility for intellectual and physical diversion from the work on this thesis.

Last but not least, I want to thank my sister Christine Melcher for proof-reading the manuscript of this thesis. However, for any errors which may remain in this work the responsibility is entirely my own.

CONTENTS

ACRONYMS

AIIM	Association for Information and Image Management
API	Application Programming Interface
BPEL	Business Process Execution Language
BPM	Business Process Management
BPMM	Business Process Maturity Model
BPMN	Business Process Modeling Notation
BPR	Business Process Reengineering
BOM	bill of materials
EPC	event-driven process chain
FMESP	Framework for the Modeling and Evaluation of Software Processes
GQM	Goal Question Metric
ICT	Information and Communication Technology
LOC	lines of code
PBWD	Product-Based Workflow Design
PCA	Principal Component Analysis
PDM	Product Data Model
WfM	Workflow Management
WfMC	Workflow Management Coalition
WfMS	Workflow Management System
YAWL	Yet Another Workflow Language

1 INTRODUCTION

1.1 MOTIVATION

Today, most companies—especially in the service industry—produce their products or services by carrying out a set of core business processes (e. g., a process for a loan application in a bank or a process for an insurance claim in an insurance company). These "processes generate most of the costs of any business" and "also strongly influence the *quality* of the product and the satisfaction of the customer" [121, p. 64].

Caused by the high competition in globalized markets which demands fast innovation cycles, companies have realized in recent years that looking at their processes and optimizing them is an important key for economic success in these days.

As a result, Business Process Management (BPM) as a process-oriented management discipline and Workflow Management (WfM) as its IT-based technological basis have emerged in the previous decades. BPM deals with the standardization of processes into process models, the subsequent implementation and execution using IT systems as well as the permanent analysis and—when indicated—optimization of these process models.

Prerequisite for this analysis and optimization is the measurement of interesting external process model attributes like, for example, costs, duration or error probability. It would be desirable to have a prediction possibility for these values before a process model has been implemented and executed.

Process measurement—a rather young research discipline—deals with this problem. It was strongly motivated and inspired by the area of software measurement (see, for example, [43] for an overview). There, external attributes like, for example, costs, implementation duration and number of errors are predicted for a piece of source code based on the values of different software metrics measuring (structural) internal attributes of the source code.

1.2 OBJECTIVE AND CONTRIBUTION

Looking at the process measurement literature, numerous proposed process model metrics can be found (see [136] and [99, pp. 114–117] for literature reviews). Most of them are inspired and adapted from software metrics. The authors claim that these metrics measure process complexity, quality and/or performance. At the same time, missing definitions of process model complexity and quality can be noticed.

Thus, this thesis first examines possible definitions and differentiations of these terms. It can be shown that there does not exist a single formal definition of complexity. Instead, numerous aspects of complexity were identified and are analyzed in different research communities. Consequently, it is problematic to say that a process model metric measures the complexity of a process model.

One main contribution of this thesis is a theoretical framework for process measurement in which the existing work can be integrated and which can help to identify open research questions. Some of them are further examined in this thesis.

The framework adapts well-established concepts from software measurement. The result is a prediction system measurement approach, which is based on measurement and prediction systems [43, p. 104]. The measurement approach consists of process model metrics measuring (structural) internal attributes and process model quality measures measuring external process model attributes. Through this, a concrete definition of process model complexity can be avoided.

Prerequisite for the application of a prediction system is its proper validation. For that, the reliability and validity [5, pp. 141, 143] of the involved metrics and measures have to be shown. Yet, both constructs have not received the necessary attention in process measurement literature so far.

A proper validation would require controlled experiments, which are very time and cost intensive. This fact—together with the possibility of a negative outcome of the experimental validation—could explain the small number of existing validated prediction systems.

The thesis suggests an approach which shall help to ease this problem by reducing the experimental effort for unsuccessful validations or validations of useless prediction systems. In order to reach this goal, the approach adds an additional analysis step before the prediction system which shall be validated is selected. In this preceding step, the behavior as well as important properties of process model metrics which are part of the potential prediction systems which shall be validated are analyzed first.

Through this, unfavorable properties of process model metrics (e. g., insufficient dispersion of metric values or strong correlations with other process model metrics) can be identified before high effort for an experimental validation of the corresponding prediction system occurs.

The approach is successfully applied to 33 process model metrics and 515 process models.

As most humans are visually thinking beings—preferring pictures to large tables of numbers—, a visualization of large process model collections based on process model metric values would be interesting. Yet, the resulting process model metric data would be very high-dimensional making visualization problematic. A second interesting question is whether there are clusters of (structurally) similar process models among a process model collection.

This thesis proposes an approach for these two goals. It comprises heatmaps [177], a compact visualization technique for high-dimensional data originally used in genetics, and scatter plots for dimensionally reduced data using Principal Component Analysis (PCA) [66] for visualizing the process model metric data. Additionally, clustering [12] [169, pp. 586–588] is used for analyzing the correlations between different process model metrics and finding (structurally) similar process models among a process model collection.

The approach is again applied to the same 33 process model metrics and 515 process models in order to make the findings comparable.

Next, the thesis deals with the measurement of structural process model understandability as an example of a very important external process model attribute. Understandability influences other quality aspects of process models like error-proneness and maintainability. Even though the importance of understandability is undoubted, Mendling *et al.* state that "we know surprisingly little about the act of modeling and which factors contribute to a 'good' process model in terms of human understandability" [101, p. 48].

Some published studies try to examine the dependencies between some influencing factors and process model understandability: In [138, 139], Sarshar *et al.* compare the understandability of different process modeling languages. Recker and Dreiling examine whether somebody's experience with one process modeling language can be helpful for understanding process models based on another modeling language he/she is not familiar with [125]. Mendling *et al.* search for dependencies between personal and process model specific (structural) properties and process model understandability [101]. In [102], Mendling and Strembeck also examine the influence of content related factors on process model understandability. Reijers and Mendling test the effect of process model modularization on process model understandability [126].

For examining structural process model understandability and validating appropriate prediction systems (as done in the above mentioned studies), one first has to quantify structural process model understandability. Thus, a proper measure for structural process model understandability which fulfills the reliability and validity requirements has to be found. Looking at the few proposed measures for structural process model understandability [101, 102, 126], serious doubts about this necessary validation arise.

In this thesis, concrete and detailed definitions for measuring structural process model understandability are given which exceed those in existing publications. Using these definitions, hypotheses about effects of measuring structural process model understandability are formulated which have to be considered in the measuring process.

Finally, an experimental evaluation comprising two experiments is conducted to examine these hypotheses. The first experiment involves 18 students, the second 178 students. The results show the applicability and the behavior of the proposed measures. Furthermore, they support most of the postulated hypotheses.

The last part of the thesis deals with the experimental evaluation of a postulated—yet not validated—prediction system.

During the design phase of a process model, choosing the adequate size of process activities (process model granularity) is a well-known problem. Vanderfeesten *et al.* have proposed a heuristic for this problem which is inspired by the concepts of *coupling* and *cohesion* in software engineering [129, 168].

In this field of study, the influence of coupling and cohesion on structural software complexity has been examined for some decades (see, for example, [34, pp. 984–985] for a short literature review). The first ideas about coupling and cohesion for the procedural programming paradigm were published in the 1970s under the name "structured design" [149, 179]. Basic coupling and cohesion metrics for the object-oriented paradigm can be found, for example, in the classic Chidamber and Kemerer metrics suite [26]. Empirical evaluations showed the influence of coupling and cohesion metrics on structural software complexity (e. g., [34, 151]).

Motivated by these results from software engineering, Vanderfeesten *et al.* introduced a process model granularity metric. This metric measures the ratio between process model coupling and cohesion. Based on this metric, they suggested a heuristic for selecting between different process model alternatives. It prefers models with high cohesion and low coupling. Vanderfeesten *et al.* also postulated the hypothesis that those process models are less error-prone during process instance execution. As they do not give an empirical validation of their heuristic and hypothesis, it is still no valid prediction system.

The thesis presents an experimentation system for analyzing the hypothesis and the results of a conducted experiment with 165 students using this experimentation system. The results do not support the heuristic of Vanderfeesten *et al.* Instead, an alternative error probability model is suggested which can explain the results of the experiment.

1.3 OUTLINE

The outline of the thesis is as follows:

Chapter 2 explains the basics of BPM, which are necessary for the remainder of the thesis.

An overview of the existing process measurement literature, a subsequent discussion as well as the introduction of the theoretical framework for process measurement which is used in this thesis are presented in Chapter 3.

In Chapter 4, an approach for reducing the experimental effort for validating prediction systems is suggested. It is based on analyzing the behavior as well as important properties of process model metrics.

An approach for visualizing and clustering large process model collections based on their process model metric values is proposed in Chapter 5.

Chapter 6 deals with measuring structural process model understandability as an example of an external process model attribute.

A process model granularity heuristic as an example of a prediction system is experimentally examined in Chapter 7.

Finally, some appendices explain necessary basics which are used within the thesis: measurement fundamentals (Chapter A), basics of empirical research (Chapter B) and measuring correlations (Chapter C).

The thesis structure is visually depicted in Figure 1.1.

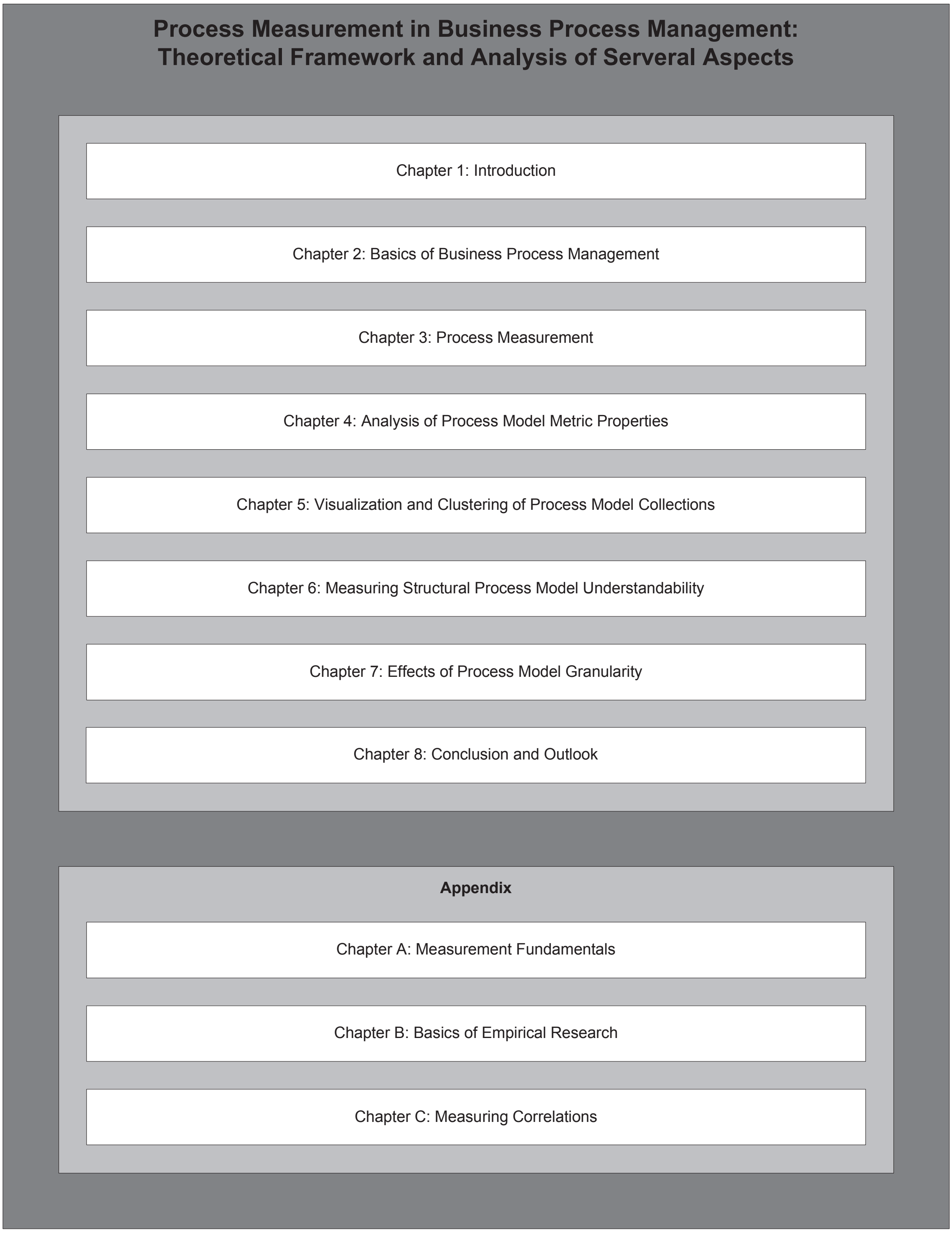


Figure 1.1: Thesis structure.

1.4 PREVIOUS PUBLICATIONS

This thesis summarizes the results of several years of research. During this time, several papers were published which present parts of the results. Some of the chapters of this thesis partially base upon these previously published papers.

The main ideas of Chapter 3 (process measurement) were already published in [89].

The visualization and clustering approach for large process model collections based on their process model metric values (Chapter 5) was already presented in [92, 93].

The own measures for structural process model understandability, the postulated hypotheses concerning measurement effects and the results of the first experiment of Chapter 6 were already described in [90, 91]. The second and larger experiment, which was conducted as a cooperation with Jan Mendling (*Humboldt-Universität zu Berlin*) and Hajo A. Reijers (*Technische Universiteit Eindhoven*), was already presented in [87, 88].

The experimental evaluation of the process model granularity heuristic (Chapter 7) was already published in [94, 95].

2 BASICS OF BUSINESS PROCESS MANAGEMENT

Caused by the high competition in globalized markets which demands fast innovation cycles, companies have realized in recent years that looking at their processes and optimizing them is an important key for economic success in these days. Powell *et al.* summarize: "[...] business processes generate most of the costs of any business, so improving efficiency generally requires improving processes. Business processes also strongly influence the *quality* of the product and the satisfaction of the customer, both of which are of fundamental importance in the marketplace." [121, p. 64] As a result, BPM as a process-oriented management discipline and WfM as its IT-based technological basis have emerged in the previous decades.

This chapter presents the necessary basics of this development. In Section 2.1, BPM as a management discipline is introduced. The IT-based technological support of this approach—known as WfM—is explained in Section 2.2. Finally, Section 2.3 gives an overview of different process modeling languages.

2.1 BUSINESS PROCESS MANAGEMENT

2.1.1 *Business Processes*

Definition

Weske states [175, p. 4]: "Business process management is based on the observation that each product that a company provides to the market is the outcome of a number of activities performed. Business processes are the key instrument to organizing these activities and to improving the understanding of their interrelationships."

Thus, one has to clarify what one means with "business process" before one can think about BPM in detail.

Ko gives an overview of different definitions of this term [72, p. 12]. According to him, most existing definitions can be traced back to a definition given by Hammer and Champy in their seminal book about Business Process Reengineering (BPR) [56, p. 35].

> "We define a business process as a collection of activities that takes one or more kinds of input and creates an output that is of value to the customer."

The next definition which is cited by Ko was proposed by Davenport in his seminal book about using information technology for process innovation [35, p. 5]. In this definition, an additional emphasis on the performed activities' structure and *how* work is done can be found.

> "[...] a process is simply a structured, measured set of activities designed to produce a specified output for a particular customer or market. It implies a strong emphasis on *how* work is done within an organization, in contrast to a product focus's emphasis on *what*."

Finally, Ko quotes a definition by Ould [116, p. 26] which adds two important elements still missing in the other definitions—the actors which perform the activities and the collaboration between them.

> "[...] key features of the thing that we call 'process':
>
> - it contains purposeful activity (ie things are done for a reason)
> - it is carried out collaboratively by a group (ie we are concerned with more than the work of the individual)
> - it often crosses functional boundaries (ie the organisation is not the process)
> - it is invariably driven by the outside world (ie our processes generally have 'customers' in some shape or form)"

Summarizing these proposed definitions, the following definition is used in this thesis.

Definition 2.1 (Business process) *A business process consists of a structured set of activities, which are performed by (potentially several) actors (humans, computers and/or machines) in an organization in order to collaboratively achieve a common business goal—the provision of a service or the production of a product for an internal or external customer.*

Classification

In the literature, one can find several classification schemes for business processes.

VAN DER AALST AND VAN HEE Van der Aalst and van Hee [162, pp. 9–10] use the role of a process within a company for classification. They distinguish production, support and managerial processes. *Production processes* produce a company's products or services and, thus, generate income for the company. *Support processes* support the production processes. This class comprises maintenance processes for the production systems as well as personnel management processes such as recruitment and selection, training and payment. The last class, the *managerial processes*, directs and coordinates the production and support processes. Here, the objectives and preconditions for the managers of the other processes are formulated, required resources are allocated and contact is held with financiers and other stakeholders.

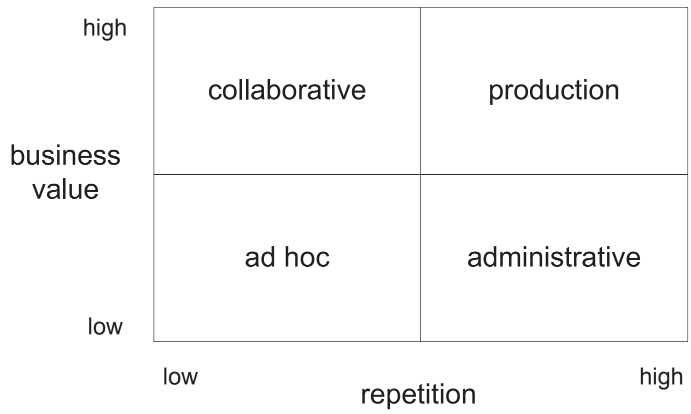


Figure 2.1: Business process classification scheme by Leymann and Roller [77, p. 10].

LEYMANN AND ROLLER The classification scheme by Leymann and Roller [77, pp. 10–12][1] classifies processes based on the dimensions business value and repetition. *Business value* defines the importance of a process for a company. A business process with high business value is a core competence of its company such as a loan granting process for a bank or a car manufacturing process for a car manufacturer. *Repetition* measures how often a process is executed in the same manner. A process with high repetition is an ideal candidate for modeling and execution with IT support. Both dimensions are divided into two value ranges—low and high—resulting in four process types: production, administrative, collaborative and ad hoc processes (see Figure 2.1).

Production processes have a high business value and repetition. They constitute a company's core business such as the loan granting process of a bank. The efficient execution of these processes is a competitive advantage. *Administrative processes* are also highly repetitive, yet they have a lower business value. Typical examples are travel expense reports and purchase approvals. *Collaborative processes* are characterized by a high business value but a low repetition. They comprise processes such as writing technical documentation or creating software. The underlying process is rather complex and specifically created for the particular task—often by customizing a more general project plan. Changes to the initial process plan are also quite common. Finally, *ad hoc processes* have both low business value and low repetition. Often, they have no predefined structure but are constructed individually each time a series of actions shall be performed. Examples are for-your-information routing as well as review and approval processes.

As a third dimension, Leymann and Roller propose the degree of automation. It measures the independence of a process execution from human actions.

WESKE In [175, pp. 17–21], Weske proposes a further classification scheme for business processes. It consists of the three dimensions degree of automation, degree of repetition and degree of structuring. The *degree of automation* measures the rate of human interaction within a process. The *degree of repetition*—as in the classification scheme by Leymann and Roller—declares how often a process is

1 The classification scheme was originally proposed by GIGA Information Group.

executed in the same manner. Finally, the *degree of structuring* indicates whether a process with all its execution rules (e. g., decision rules for a loan granting process) can be fully described.

CONCLUSION Independent from the chosen classification scheme, one can state that the higher a process's repetition rate, degree of structuring and business value, the more applicable and useful is the modeling and execution with IT support of this process.

2.1.2 *What Is Business Process Management?*

Motivation, Development and Definition of Business Process Management

Weske summarizes the problems and challenges which companies have to face today as follows:

"[. . .] in today's dynamic markets, companies are constantly forced to provide better and more specific products to their customers. Products that are successful today might not be successful tomorrow. If a competitor provides a cheaper, better designed, or more conveniently usable product, the market share of the first product will most likely diminish." [175, p. 4]

"Internet-based communication facilities spread news of new products at lightning speed, so traditional product cycles are not suitable for coping with today's dynamic markets. The abilities to create a new product and to bring it to the market rapidly, and to adapt an existing product at low cost have become competitive advantages of successful companies." [175, p. 4]

As a consequence, the optimization and flexibility of business processes got into the companies' focus and finally led into the development of BPM.

This trend towards process orientation can be traced back into the 1990s. In the seminal book *Reengineering the Corporation* by Hammer and Champy [56], the authors introduced the central idea of BPR—the radical redesign of a company's business processes. [175, p. 4] Davenport published a book [35] which describes how this process innovation can be supported by information technology.

Although radical changes in a company's processes—as suggested by BPR—can be useful in special cases, one soon recognized also the flaws of this approach such as a lack of commitment of the involved employees and operational risks caused by these drastic changes.

Thus, the wish for a more evolutionary approach arose which changes business processes in—possibly several iterative and incremental—smaller steps [72, p. 14]. This idea is realized in the BPM approach.

Definition 2.2 (Business Process Management) *Business Process Management (BPM) is a process-oriented management discipline (as quoted in [72, p. 14]). It includes methods, techniques and tools to support the design, enactment, management and*

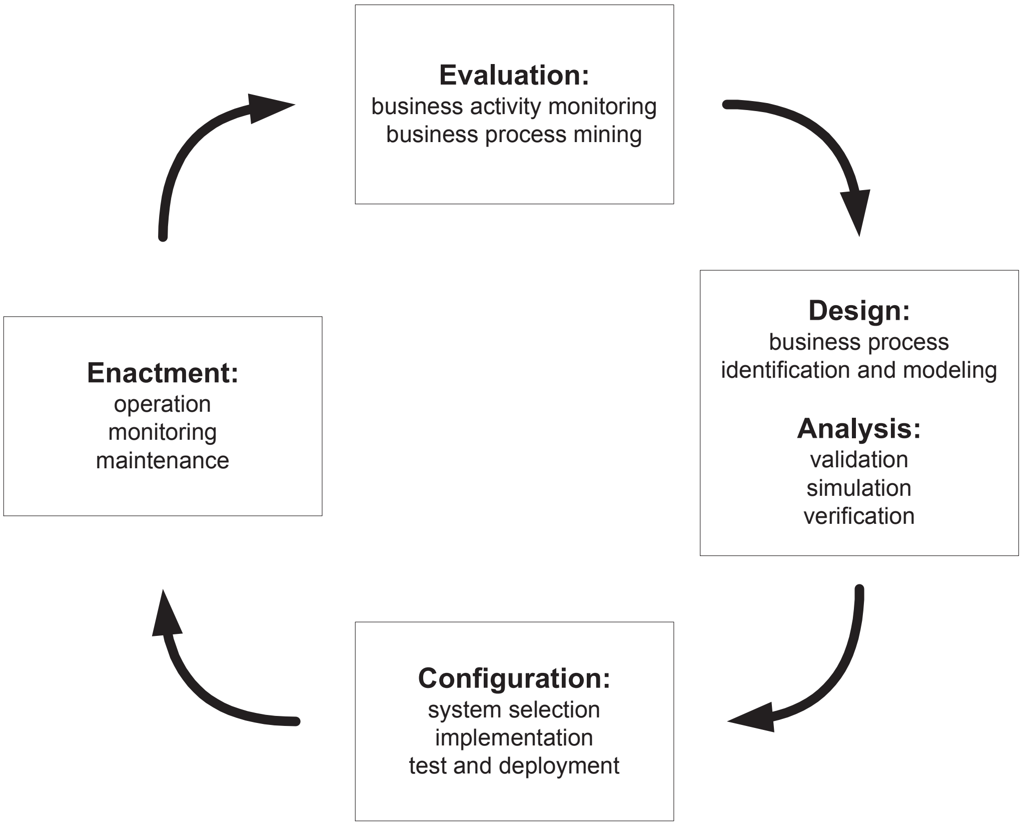


Figure 2.2: BPM lifecycle [175, p. 12].

analysis of operational[2] business processes involving humans, organizations, applications, documents and other sources of information [163, pp. 1, 4].

Business Process Management Lifecycle

To understand the exact meaning of the BPM approach, one has to look at the BPM lifecycle. According to Ko [72, p. 14], there are several different views on this lifecycle—one of the most relevant by van der Aalst *et al.* [163, p. 5]. In this thesis, the BPM lifecycle view by Weske [175, pp. 11–17] is presented. It is a slightly modified and expanded version of the view by van der Aalst *et al.*

The BPM lifecycle—as seen by Weske—consists of the four phases design & analysis, configuration, enactment and evaluation, which are arranged in a cyclic structure (see Figure 2.2). This structure represents the evolutionary and incremental BPM approach.

DESIGN AND ANALYSIS The first phase of the BPM lifecycle consists of two steps—design and analysis.

There are two possible initial situations: (1) The business process is totally new in the company and has not yet been executed or (2) it is already performed manually without explicit BPM support. Depending on the situation, there either

2 This definition restricts BPM to operational processes, i. e., processes at the strategic level or processes that cannot be made explicit are excluded [163, pp. 4–5].

exists only a vague expectation of the process—possibly in text form—or it even only exists in the personal know how of the employees involved in the process so far. In both cases, the goal of the design step is an explicit business process model (see Definition 2.4) as a formalized representation of the business process of interest. This model is represented in a process modeling language (see Section 2.3).

In the subsequent analysis step, the process model has to be validated. For that purpose, all stakeholders check whether it contains the execution sequences of all valid process instances (see Definition 2.7). Furthermore, the model can be simulated in order to collect information about the number of required resources. The analysis step can also comprise an automatic verification which checks whether the model is free of deadlocks.

CONFIGURATION In the configuration phase, the process model which was created during the first phase has to be implemented. This can be done in two different ways: (1) It can be implemented without any software support by a set of policies and procedures which the employees need to comply with or (2) it can be realized using a dedicated software system.

In the latter case, an implementation platform is chosen and the process model is enhanced with technical information in order to facilitate the enactment of the process by the BPM system. This often comprises the integration of existing software systems (legacy software systems) with the BPM system.

The phase finishes with a test of the implementation. For that purpose, the normal test approaches from software engineering are used.

ENACTMENT In the enactment phase—which is normally the longest phase of the BPM lifecycle—, the different process instances (Definition 2.7) of the business process are executed.

In case of the application of a BPM system, it controls this execution according to the constraints and business rules defined in the process model.

The BPM system's monitoring component can provide information about the execution status of a process instance.

During the whole enactment phase, execution data is collected and stored in some form of a log file. This data is the basis for the evaluation in the next phase of the lifecylce.

EVALUATION In the evaluation phase, the available log data is used to evaluate and improve the process model and its implementation. For these purposes, business activity monitoring and business process mining techniques are used.

Business activity monitoring can help to identify, for example, bottlenecks of process model implementations caused by a shortage of required resources.

Business process mining—a rather recently developing field of research [160]—can be used as a starting point for the development of process models from log files created by traditional information systems (instead of dedicated BPM systems). These traditional systems support the execution of processes even without the prior explicit definition of a process model. Furthermore, business process mining is helpful for discovering control, data, organizational and social structures of the process execution [160, p. 713].

Goals of Business Process Management

Companies hope to reach several goals when using the BPM approach. Weske [175, pp. 21–22] as well as Becker and Kahn [10, pp. 5–9] give an overview of such goals, which BPM is able to fulfill.

- **Better understanding**

  The explicit representation of business processes can help a company to get a better understanding of the operations it performs and the dependencies (possible "side effects") between the different processes [175, p. 21] [10, p. 7].

- **Standardization of business process execution**

  The explicit representation and the IT supported process execution help to narrow the gap between how a process is planned to be executed in theory and the way it is actually executed in practice [175, pp. 21–22]. Thus, a more standardized process execution is reached.

- **Improved communication**

  The explicit representation of business processes as well as using BPM terminology can improve the communication between the different stakeholders by making it more efficient and effective. Through this, also the collaborative analysis and identification of potential improvements becomes easier. [175, p. 21]

- **More flexibility**

  BPM can improve the flexibility of business processes for a faster adaptation to changing market situations and customer requirements. The explicit modeling of the process and their IT support are important factors to reach this goal. [175, p. 21] [10, p. 9]

- **Continuous process improvement**

  The evolutionary approach of BPM allows a continuous improvement of the business processes. The explicit modeling and IT-based execution enable the analysis and identification of potential for improvements. [175, p. 21]

- **Repository of business processes**

  A company can construct a repository with all its (modeled) business processes. This is an important asset as it captures the company's knowledge about what and how it performs its operations. [175, p. 21]

Table 2.1: BPM market size (in million U.S. $) [114, p. 10].

	2005	2006	2011 (projected)
Gartner	1,000 (estimate)	1,700 (estimate)	5,100
Forrester	-	1,600 (actual)	6,300
IDC	495 (actual)	890 (actual)	5,500

- **Benchmarking**

  The explicit representation and the IT support enable to collect performance measures at specific points of a process. That way, internal and external benchmarking becomes possible. [10, p. 6]

- **Enabling cooperation/outsourcing**

  When a process is modeled with all its performed operations and dependencies, it also becomes easier to virtually span processes over the borders of companies (cooperation) or to outsource parts of or whole processes which are not the core competency of that company [10, p. 8].

2.1.3 *Business Process Management in Practice*

After the definition of BPM and the presentation of the BPM lifecycle earlier in this section, this subsection gives an overview of the BPM market with information about its size and offered software tools as well as the actual usage of the BPM approach in companies.

Business Process Management Market

Table 2.1 shows the actual and estimated BPM (software) market size figures published in studies by Gartner, Forrester and IDC. The different figures—even for the present—are caused by different definitions of the BPM market used by the analysts.

All three market research companies predict a rapid market growth in the next years. Nevertheless, the BPM market is just a relatively small niche of the overall global software market as can be seen when comparing the figures with the 2009 revenue figures of, for example, Microsoft (58,437 mil. U.S. $ [103, p. 40]) and SAP (10,672 mil. € [137, p. 154]).

The BPM market is on the move. Recently, large vendors such as Oracle and IBM have entered the market by acquisition of smaller BPM vendors. Furthermore, the market is consolidating. In 2007, Gartner predicted that from more than 150 BPM vendors in 2006 only the leading 25 will be evident at the end of 2008. [114, p. 10]

The study [146] gives an overview of the BPM software market (with a German perspective). It also presents the results of a systematic evaluation of the offered software products by means of a list of criteria. Furthermore, references to other—also more global—studies are given.

Usage of Business Process Management

In this paragraph, information is given about the current usage of BPM in practice as well as the customers' opinions and beliefs.

AIIM SURVEY A survey including 812 participants (with focus on the U. S. market) conducted by the Association for Information and Image Management (AIIM) in 2007 [1] revealed data on the actual usage of BPM and the expected payback time for the BPM investment.

According to this survey, 54% have BPM experience in their company—yet, 32% of all participants only on department level. 25% plan to introduce BPM within the next six months, and the remaining 20% have no experience at all. [1, p. 8]

The question concerning the expected payback time for the investment was answered as follows: 15% expect less than one year of payback time. Half of the participants estimate an amortization time of one to three years. Another 14% expect their investment to pay off within four to five years. The remaining 21% have no calculation. [1, p. 14]

ORACLE SURVEY In 2007, an Oracle survey of more than 200 Oracle Business Process Management Suite customers was conducted worldwide. It revealed the following areas where customers expect the greatest return on investment from their investment in BPM (multiple responses possible) [114, p. 8]:

- automating or accelerating highly manual processes (26%)
- increasing visibility into processes (21%)
- improving operational excellence (20%)
- improving control over processes (20%)
- simplifying cumbersome processes (18%)
- promoting better business and IT alignment (14%)
- improving delivery of new products or services (13%)
- establishing greater governance and compliance (13%)
- improving predictability of processes (10%)
- improving customer intimacy or service (10%)
- improving support for mergers and acquisitions (7%)

BPM-ALLIANZ SURVEY A 2007 survey by the *BPM-Allianz* among 769 potential BPM users in German companies showed a quite heterogeneous interest in BPM: While in the logistics, chemistry, mechanical engineering, automotive and financial industry only between 10% and 20% of the respondents have no interest in BPM, this rate increases between 40% and 50% in the health, energy, Information and Communication Technology (ICT) and food sector. [140]

PENTADOC AND TROVARIT SURVEY Another German survey with 157 participants conducted in 2010 by Pentadoc and Trovarit [28, 119] revealed the following results:

Only 23% of the asked companies use special BPM software [119, p. 5]. Another 23% plan its introduction—yet, around half of them not within the next two years [119, p. 8].

Both groups have similar goals for a BPM usage—more process control and transparency, reduction of workload, process speedup, increase of competitiveness and reduction of costs [119, pp. 4, 9]. Respondents using BPM could reduce lead times by 37% and costs by 31% on average [119, pp. 4–5].

Participants which do not use BPM name as reasons: lack of available competent resources (almost 50%), unclear or too small benefit (around 45%), expectation of too much effort (43%) and expectation of too much costs (30%) [119, pp. 5–6].

GARTNER: AMORTIZATION According to Gartner, the usage of BPM can reduce costs up to 20%. Thus, its introduction can pay off within one year. [29, 30]

IDS SCHEER: BUSINESS INTELLIGENCE During the current financial crisis, IDS Scheer sees a trend towards Business Intelligence. That way, business process performance measures (e. g., lead time and costs) can be permanently collected during execution. They can be used as early indicators by the management in order to take actions long before the final financial figures are available. [29, 30]

2.2 WORKFLOW MANAGEMENT

Besides BPM, the term "Workflow Management (WfM)" can be found quite often in literature. WfM as a concept is older than BPM (see [52] for a—meanwhile "historic"—overview of WfM). While BPM is a process-oriented management discipline (see Definition 2.2), WfM is a technological approach for the automation of business processes.

According to van der Aalst *et al.* [163, p. 5], WfM can be seen as today's technological basis of BPM and can be integrated into the BPM lifecycle—comprising all phases except the evaluation phase.

2.2.1 *Workflow Management and Terminology*

In this subsection, the important terminology of WfM is introduced. Thereby, the definitions of the Workflow Management Coalition (WfMC) [176] are used.

Definition 2.3 (Workflow Management) *Workflow Management deals with the automation of business processes, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules [176, p. 8].*

Today, WfM constitutes the technological basis of BPM.

Definition 2.4 (Process model) *A process model is a representation of a business process in a form which supports automated manipulation or enactment by a software system [176, p. 11]*[3]

It consists of a network of activities, their relationships (possible activity execution sequences) and a set of rules that determine which alternative activity execution sequence has to be chosen at a branching.

Normally, it is possible to find several alternative process models for one business process which differ—amongst others—in their activity network structure. This fact can be compared with the following analogy: For the problem how to sort a collection of numbers (e. g., in increasing order), one can find different sorting algorithms. They differ in the details of their basic operations and their order—possibly resulting in different runtime and/or storage consumption. Nevertheless, the same input data produces the same output.

Definition 2.5 (Activity/task) *An activity (also called "task") is a description of a piece of work which forms one logical step within a process (model) [176, p. 13].*

There are manual activities, which require the execution by humans, and automated activities, which can be done fully automatically by machines and/or computers.

Each activity needs a resource (human or machine/computer) with specific "skills" to be executed. The requirements for that resource are summarized within a role.

Definition 2.6 (Role) *A role comprises attributes as skills, location and authority within an organizational structure which are either required by a resource to execute a special activity or which a resource provides [176, pp. 54–55].*

A business process (e. g., an insurance claim process) can be initiated and executed indefinitely often—normally with different data each time.

Definition 2.7 (Process instance) *A process instance is the representation of a single enactment of a business process [176, p. 16].*

It is executed according to a process model of the business process. Each process instance has its own data and can be executed independently from other process instances of the same process.

3 In [176, p. 11], the term "process definition" is used instead of "process model". Yet, "process model" is used in newer publications about BPM as, for example, [99, 163].

2.2.2 *Workflow Management Systems*

The main goal of WfM is the software-supported execution of processes. For that purpose, Workflow Management Systems (WfMSs) are used.

Definition 2.8 (Workflow Management System) *A Workflow Management System is a software system for defining process models as well as for creating and managing the execution of the corresponding process instances. It runs on one or more workflow engines which are able to interpret the process model, interact with workflow participants and—where required—invoke the use of IT tools and applications. [176, p. 9]*

A workflow engine—the most important part of a WfMS—is defined by the WfMC as follows [176, p. 57].

Definition 2.9 (Workflow engine) *A workflow engine is a software service or "engine" which provides the run time execution environment for process instances. For that purpose, it provides the following features:*

- *interpretation of the corresponding process model*
- *creation of process instances and management of their execution*
- *"navigation" and "routing" between the activities (Which activity is executed next and which branch is taken at branchings?)*
- *allocation of activities to resources according to required/offered role*

The WfMC has created the Workflow Reference Model (see Figure 2.3) as a generic architectural representation of a WfMS including its generic components and most important system interfaces (using a workflow Application Programming Interface (API) and standardized interchange formats) [58, pp. 20–27] [176, p. 23]

- import and export of process models,
- interaction with client applications,
- invocation of software tools or applications,
- interoperability between different WfMSs and
- administration and monitoring functions.

2.3 PROCESS MODELING LANGUAGES

In the area of BPM, numerous standards have emerged which are used to create process models of business processes and to interchange and/or execute these models.

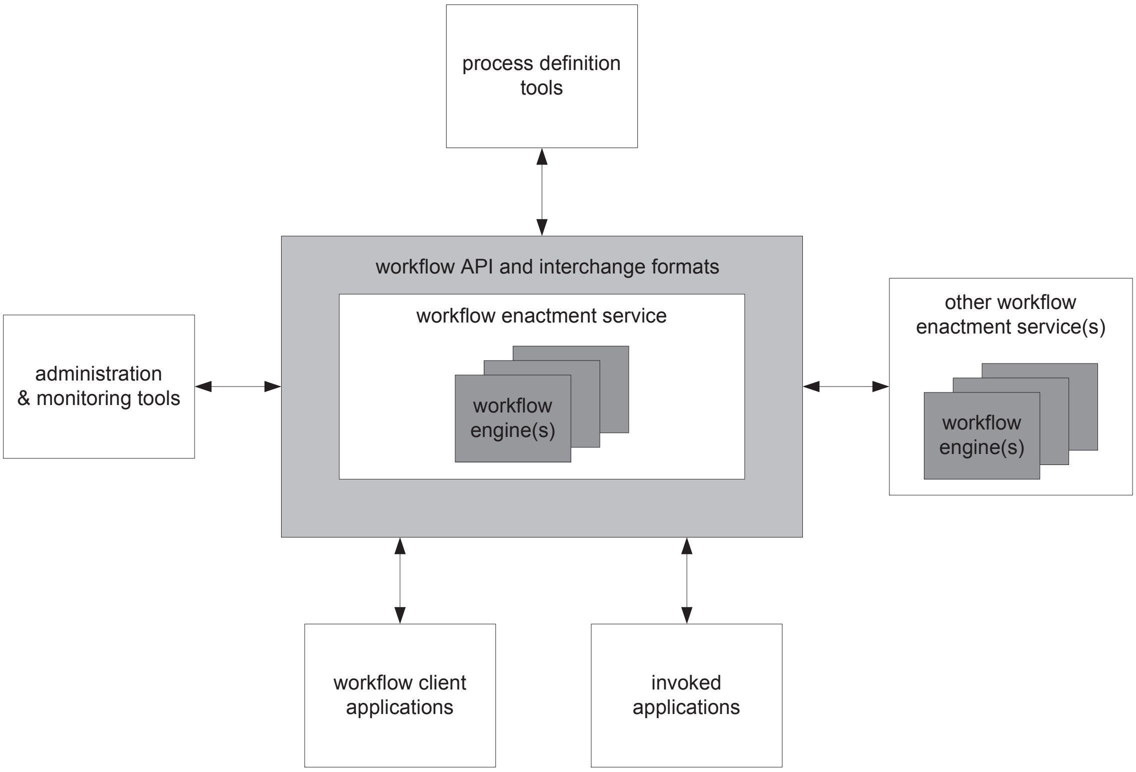


Figure 2.3: Workflow Reference Model [58, p. 20].

Ko *et al.* give an overview of these standards [73]. Furthermore, they propose a classification scheme which classifies the standards into graphical, execution, interchange and diagnosis standards [73, pp. 751–754].

In [175, pp. 125–226], Weske gives another good overview of process modeling languages with detailed explanations of the languages

- Business Process Modeling Notation (BPMN) [175, pp. 205–225],
- event-driven process chains (EPCs) [175, pp. 158–169],
- Petri nets [175, pp. 149–158],
- workflow nets [175, pp. 169–182] and
- Yet Another Workflow Language (YAWL) [175, pp. 182–200].

All these listed process modeling languages are activity-based, i. e., process models modeled with these languages consist of the activities to be executed and a network of arcs between them modeling the possible sequences of activity execution. Independently from the chosen language, several so-called control flow patterns [175, pp. 126–149] can be expressed in these languages—(linear) sequence, parallel (AND) and alternative (XOR or OR) execution being the most frequently used.

In the following two subsections, two process modeling languages which are used in this thesis are presented in more detail—EPCs (Chapter 4 and 5) and Product Data Models (Chapter 7).

2.3.1 *Event-Driven Process Chains*

Event-driven process chains (EPCs) are a development by the *Institut für Wirtschaftsinformatik* (Institute for Information Systems) at *Universität Saarbrücken*. In the 1990s, there was a project together with SAP to define a suitable business process modeling language to document the processes of the SAP R/3 enterprise resource planning system. This project produced two major results: the definition of EPCs [70] and the documentation of the SAP system in the SAP Reference Model [33, 71]. [99, p. 17]

The following formal definition is based on [175, p. 162]. Yet, it partially uses some terminology from [99, pp. 22–23] instead.

Definition 2.10 (Event-driven process chain) *An event-driven process chain is a 5-tuple* (E, F, C, m, A) *for which holds:*

- E *is a nonempty set of* events*.*
- F *is a nonempty set of* functions*.*
- C *is a set of* connectors*.*
- $m : C \mapsto \{AND, OR, XOR\}$ *is a mapping which assigns to each connector a connector type, representing* AND*,* OR *or* XOR *(exclusive or) semantics.*
- *Let* $N := E \cup F \cup C$ *be the set of* nodes*.* $A \subseteq N \times N$ *is a set of* arcs *connecting events, functions and connectors such that the following conditions hold:*
  - $G := (N, A)$ *is a connected graph.*
  - *Each function has exactly one incoming and exactly one outgoing arc.*
  - *There is at least one* start event *and at least one* end event*. Each start event has exactly one outgoing and no incoming arc. Each end event has exactly one incoming and no outgoing arc. All other events have exactly one incoming and one outgoing arc (*intermediate event*).*
  - *Each event can only be followed—possibly via connectors—by functions, and each function can only be followed—possibly via connectors—by events.*
  - *There is no cycle in an EPC which consists of connectors only.*
  - *No event is followed by a decision node, i. e., an* OR *split connector or an* XOR *split connector.*

Functions (see Figure 2.4b for the graphical notation) represent the activities of the modeled process—events (Figure 2.4a) represent their pre- and post-conditions. Connectors are the third node type of an EPC. They are used for modeling non-sequential control flows. Connectors can be divided into split (one incoming and several outgoing arcs) and join (several incoming and one outgoing arcs) connectors. Besides, each connector node has one of the three types AND (Figure 2.4c), OR (Figure 2.4d) or XOR (Figure 2.4e). At AND split connectors, all

(a) Event. (b) Function. (c) AND connector. (d) OR connector. (e) XOR connector.

Figure 2.4: Graphical notation of the different EPC components.

subsequent branches are executed in parallel. At XOR split connectors, exactly one branch is taken. OR split connectors are in between—here, at least one (possibly several ones or even all) of the subsequent branches is executed.

The execution of an EPC starts when a start event (from possibly several ones) occurs. Afterwards, the arcs between the nodes together with "decisions" at split connectors define the path of the control flow through the process model. The execution finishes when an end event (from possibly several ones) is reached.

Figure 2.5 shows an example of a small and simple process model modeled as an EPC. The process starts as soon as start event "event A" occurs. Then, "function B" is executed first. Based on the outcome of this execution, the following XOR split connector "decides" which of the two subsequent branches is taken. Depending on this "decision", the corresponding branch is executed. The control flow passes the XOR join connector as soon as this branch has been completed. When the final end event "event G" is reached, the process execution is finished.

2.3.2 *Product Data Models*

Product data models are the process modeling language used in Product-Based Workflow Design (PBWD)—a modeling methodology proposed by van der Aalst, Reijers and Limam in [128, 159]. It is based on former work by van der Aalst [157, 158] and is further examined by Reijers in [127].

The motivation of the PBWD methodology is the area of manufacturing where the interaction between the design of a product and the process to manufacture this product is studied in detail [128, p. 229] [18, pp. 469–471].

There, a so-called bill of materials (BOM) [115] [18, pp. 138–146] is used to define the design of the product. A BOM has a tree structure with the final end product as its root and raw materials and purchased products as leaves. The nodes correspond to products (end products, raw materials, subassemblies). The edges represent an *is-part-of* relation and have a cardinality to indicate the number of products needed. The simplified BOM of a car is depicted in Figure 2.6. According to that, a car is composed of an engine and a subassembly. The subassembly consists of four wheels and one chassis. [128, p. 234]

In [157, pp. 397–398], van der Aalst *et al.* emphasize that there is also this kind of dualism between product and process model in the information-intensive service industry—even though, it is seldom made explicit. They give the example of processing an insurance claim. There, the product is basically a decision: either the claim is accepted (followed by a payment) or it is rejected. Different data elements

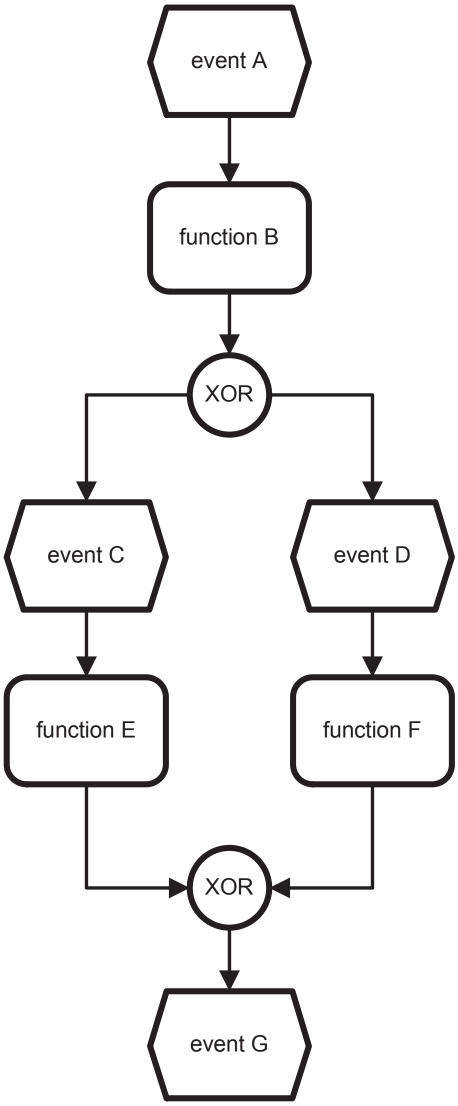


Figure 2.5: An example of an EPC.

may play a role in making this decision. One can think of these data elements as raw materials or subassemblies. The process model should manufacture the decision.

Reijers *et al.* criticize that "[i]n contrast to manufacturing, the product and the process have often diverged in actual workflows. Workflows found in banks and insurance companies for products like credit, savings and mortgages, damage and life insurance, and so on may well exist for decades. Since their first release, those processes have undergone an *organic evolution*. [...] Aside from the evolutionary changes of the processes, the state of technology of some decades ago has considerably influenced the structure of these workflows permanently. [...] So, the structure of an actual workflow may not be related to the product characteristics any more." [128, p. 230]

The PBWD methodology tries to reverse this divergence of product structure and process model for information-intensive processes of the service industry (e. g., bank, insurance and telecommunications companies). Instead of defining a

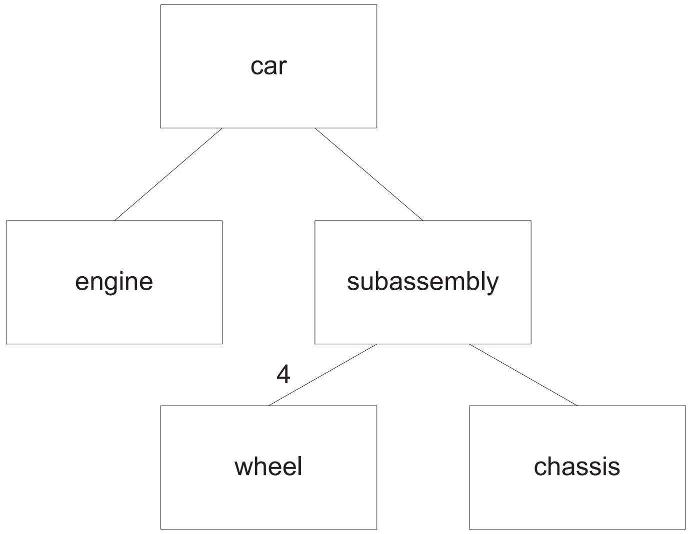


Figure 2.6: The simplified BOM of a car [128, p. 235].

process model at once, it is derived from the product structure. So, the PBWD methodology provides the following steps [128, pp. 231–232]:

1. definition of the product structure using a formalization similar to the BOM in manufacturing,

2. construction of one or several alternative process models derived from this product structure and

3. selection of the most promising process model alternative according to their estimated performance measures.

For the first step, a formalization for the product structure is necessary. Compared to manufacturing, information-intensive processes have differences which make the BOM not very useful [128, p. 234]:

- Making copies of information in electronic form is easy and cheap. Therefore, cardinalities make no sense.

- The same piece of information may be used to manufacture various kinds of new information. Therefore, also non-tree structures are possible.

- Typically, multiple ways (variants) to derive a piece of information exist.

As a consequence of these differences, Reijers *et al.* introduce a modified formalization—the Product Data Model (PDM) [128, pp. 234–236][4].

Definition 2.11 (Product Data Model) *A Product Data Model is a tuple* $(\mathtt{D}, \mathtt{C}, \mathtt{pre}, \mathtt{constr}, \mathtt{cst}, \mathtt{flow})$ *for which holds:*

- $\mathtt{D}$ *is a set of* data elements *with a special* top element $\mathtt{top} \in \mathtt{D}$.

4 The function $\mathtt{prob}$ of the original definition is omitted in this thesis as it is irrelevant here.

- C *is a set of* constraints *which can be any boolean function (including the boolean value* $\mathtt{true}$*).*
- $\mathtt{pre} : D \mapsto \mathcal{P}(\mathcal{P}(D))$ *is a function which gives for each data element the various ways of determining a value for it based on the values of different sets of other data elements so that holds:*
  - $R := \{(p, c) \in D \times D | c \in \bigcup_{es \in \mathtt{pre}(p)} es\}$ *is connected and acyclic, i. e., there are no "dangling" data elements and a value of a data element does not depend on itself.*
  - *The top element cannot be used for determining the value of any other data element:*

$$\forall (p, c) \in R : c \neq \mathtt{top}$$

  - *If there is a data element which does not require any other data element, one denotes for ease of analysis the set of required data elements as the empty set:*

$$\forall d \in D : \varnothing \in \mathtt{pre}(d) \Rightarrow \mathtt{pre}(d) = \{\varnothing\}$$

- $F := \{(p, cs) \in D \times \mathcal{P}(D) | cs \in \mathtt{pre}(p)\}$ *is a set of* production rules*, based on the definition of* $\mathtt{pre}$*.* F *consists of all ordered pairs of data elements between which a dependency may exist.*
- $\mathtt{constr} : F \mapsto C$ *is a function which associates a constraint to each production rule so that there are no constraining conditions on the production of data element values which does not require the values of other data elements:*

$$\forall d \in D : \mathtt{pre}(d) = \{\varnothing\} \Rightarrow \mathtt{constr}((d, \{\varnothing\})) = \mathtt{true}$$

- $\mathtt{cst} : F \mapsto \mathbb{N}$ *is a function which gives the cost of using a production rule.*
- $\mathtt{flow} : F \mapsto \mathbb{N}$ *is a function which gives the time it takes to use a production rule.*

A PDM is an acyclic graph. Its nodes are data elements which all have a value. There is one special data element, the $\mathtt{top}$ data element, which represents the final decision of the corresponding process. The edges of the graph represent the relations between the data elements which are given by the $\mathtt{pre}$ function. It yields for each data element d zero or more variants to determine a value for d. If one supposes for the data elements $d, e, f \in D$ that $\{e, f\} \in \mathtt{pre}(d)$, then the value of d can be determined using the values of e and f—one says that $(d, \{e, f\})$ is a production rule of d. Each production rule has a corresponding constraint (a boolean function) and can only be executed if its constraint evaluates to $\mathtt{true}$. Furthermore, the execution of each production rule causes a specific amount of costs and needs a specific time. There are special data elements whose values do not depend on those of others. They are called *leaves*.

An example of a PDM is depicted in Figure 2.7. The corresponding process checks whether a person is suitable as a helicopter pilot. On the one hand, there

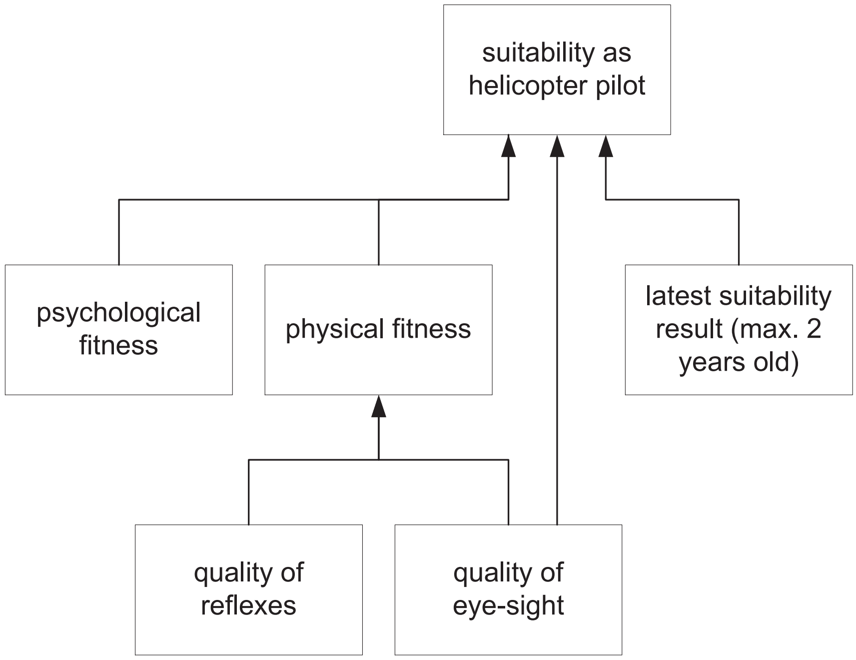


Figure 2.7: Example of a PDM for a process checking the suitability as a helicopter pilot [159, p. 399].

are production rules (e. g., from quality of reflexes and quality of eye-sight to physical fitness) which need more than one data element to compute the value of another one. On the other hand, there are other production rules (e. g., latest suitability result to suitability as helicopter pilot) which only need exactly one data element to compute the value of another one. Furthermore, there exist three variants how to compute the final decision (suitability as helicopter pilot). Based on the evaluation of their constraints, one of them can be selected: If the quality of eye-sight is too bad, the final decision can immediately be made: the examinee is not suitable. If the latest suitability result is not older than two years, its result is adopted. Only otherwise, a more complicated check has to be performed. The quality of eye-sight is an example of a data element whose value can be used for several production rules.

In the next step of the PBWD methodology, one or several alternative process models have to be derived from the PDM. For that, each production rule is transformed into an adequate activity which performs the corresponding production rule. The control flow between these activities has to preserve the relation induced by the `pre` function of the PDM. So, the values of the data elements can be computed in a correct sequence.

For the last step of the PBWD methodology, Reijers *et al.* propose some heuristics how to select the optimal alternative process model based on the PDM's `cst` or `flow` function. As the details are unimportant for this thesis, the interested reader is referred to [128, pp. 240–244].

3 PROCESS MEASUREMENT

3.1 INTRODUCTION

As already mentioned in Chapter 2, companies have identified their processes and the optimization of them as critical success factors for their businesses in recent years. Therefore, the measurement of interesting process model attributes as, for example, costs, duration and error probability is a prerequisite for the assessment and—where applicable—improvement of process models. A possibility for predicting these attributes before the actual implementation and execution of the process models would be desirable. Process measurement—a rather young research discipline—is concerned with these questions. This thesis deals with some aspects of process measurement which so far have only been examined inadequately.

In this chapter, an overview of the literature in the field of process measurement is given and it is tried to arrange it into an overall context in order to identify open research questions.

This overview of the published literature shows that many process model metrics adapted from software metrics were suggested. For many of them, the authors claim that they measure process model complexity, quality and/or performance. At the same time, missing definitions of process model complexity and quality can be noticed.

Thus, possible definitions and differentiations of these terms are examined resulting in the process measurement approach used in this thesis—a prediction system measurement approach adapted from software measurement which avoids concrete definitions of process model complexity and quality. Based on internal process model attributes (e. g., structural process model metrics), it tries to predict external process model attributes (e. g., duration and error probability). The approach is completed by a discussion of a proper validation of the prediction system.

Furthermore, the Goal Question Metric approach for selecting process model metrics and process model quality measures as well as different purposes of process measurement (understand, control and improve) are discussed and the existing process measurement work is assessed according to the measurement framework introduced in this chapter. As a result, a lack of necessary validation work (compared to the number of proposed metrics) can be noticed.

The remainder of this chapter is organized as follows: In Section 3.2, related work on process measurement is presented. It is followed by a discussion of how to define and/or measure process complexity, quality and performance in Section 3.3. In Section 3.4, measurement and prediction systems are explained, the process measurement approach used in this thesis is introduced and the necessary validation steps are presented. The different application possibilities of metrics are shown in Section 3.5. Afterwards, the existing process measurement work is assessed according to the measurement framework introduced in this chapter (Section 3.6). Finally, Section 3.7 closes the chapter with a conclusion and an outlook on identified open research questions which are dealt with in the remainder of this thesis.

3.2 RELATED WORK

In this section, an overview of the published literature about process measurement is given. In addition to own literature research, it is based on other literature reviews by Sánchez González *et al.* [136] and Mendling [99, pp. 114–117].

Looking at the literature, one can separate two types of measurement purposes for process models [136, pp. 119–120]: measuring the process model design (a "static" property) and measuring results of a process model execution (a "dynamic" property).

The latter purpose has been studied in depth in other sciences for years [136, p. 116]—often under the name process performance. In [63, 64], Jansen-Vullers *et al.* give an overview of different performance measurement systems. Based on them, they suggest their own framework for process performance measures with the dimensions time, cost, quality and flexibility. For each of these dimensions, they propose a set of measures.

The really new aspect of the process measurement discipline developing during the previous years are metrics for the other measurement purpose—for measuring the process model design. Thus, the remainder of this section only presents publications about this measurement purpose.

In [75], Lee and Yoon introduce 15 complexity metrics for Petri nets. They distinguish between *structural* (e. g., number of places and transitions, cyclomatic number) and *dynamic* (including number of markings and tokens as well as degree of parallel firing) complexity metrics. Furthermore, they analyze the correlations between these metrics for 75 practical Petri nets which were randomly selected from the literature.

Nissen presents the knowledge-based system KOPeR, which helps during the reengineering of a process model [112]. For that purpose, KOPeR uses a set of metrics measuring the process model and reengineering heuristics.

Morasca deals with measuring internal attributes of Petri nets for concurrent software specifications [108]. He identifies size, length, complexity and coupling as interesting attributes. For each of them, he defines a set of axiomatic properties which corresponding metrics have to fulfill. Afterwards, he suggests several metrics for these four attributes and validates them against the properties.

Latva-Koivisto's research report [74] is probably the first publication dealing with measuring the complexity of business process models. He makes some interesting remarks on how to define complexity (which are presented in more detail in Subsection 3.3.1). Then, he introduces several metrics for structural complexity based on graph-theory.

Inspired by McCabe's cyclomatic number for control-flow graphs of software [86], Cardoso recommends the control-flow complexity metric (CFC) for process models [21]. The metric sums up the number of states a process model can reach after a split connector depending on its type. In [22], Cardoso tests the metric for correlation with received complexity. In [24], he offers a specialization of the CFC metric for Business Process Execution Language (BPEL) process models.

In [20], Cardoso discusses data-flow complexity metrics for web processes in BPEL. He differentiates between data, interface and interface integration complexity. Yet, only for interface complexity, he advices a concrete metric.

In [23], Cardoso proposes another metric for measuring the log-based complexity (LBC) of process models in BPMN. This metric is based on the number of different log traces which can be generated from the execution of a process model. The metric is iteratively defined with basic numbers for the different workflow patterns [161].

Jung presents a metric which measures the entropy of process models [67]. Thereby, the entropy of a process model is defined as the uncertainty or variability of the control flow caused by XOR and OR splits as well as loops.

Gruhn and Laue suggest complexity metrics for business process models analogous to software complexity metrics [54]. In [53], they adapt the cognitive weights metric from software engineering to business process models.

Rolón *et al.* recommend several metrics for business process models in BPMN [132]. Their metrics are an adaptation and extension of the Framework for the Modeling and Evaluation of Software Processes (FMESP) [49].

In their survey paper [19], Cardoso *et al.* propose new metrics analogous to existing metrics for software as lines of code (LOC), the Halstead complexity metrics [55] and the information flow metric by Henry and Kafura [57]. Additionally, they present already published metrics like CFC [21] and the metrics proposed by Latva-Koivisto in [74].

In [129, 168], Vanderfeesten *et al.* introduce a heuristic for the proper size of individual activities in process models (process model granularity). Activities can consist of (several) basic operations. The operations of one activity should "belong" together (highly cohesive)—while different activities should be independent from each other (loosely coupled). For that purpose, they introduce a process model cohesion and a process model coupling metric, a coupling/cohesion ratio and a design heuristic based on this ratio.

Vanderfeesten *et al.* suggest a weighted coupling metric for process models with different weights for the different connector types [166].

Analyzing the SAP Reference Model process models with an automatic verification tool, Mendling *et al.* detect faulty EPC process models [96]. In a second step, they try to find possible predictors (based on 15 metrics) for these errors using logistic regression. In [97], Mendling proposes a density metric and repeats the regression test. Mendling and Neumann suggest and test additional six metrics as error predictors in [100]. In his PhD thesis [98], which was also published in book version [99], Mendling gives 28 metrics for EPC process models (some of them taken from [96] and [100], but also new ones). Once again, he uses logistic regression in order to identify possible predictors for faulty process models.

In [101], Mendling *et al.* present an experiment for identifying influencing personal (theoretical knowledge and practical experience) and structural factors on process model understandability. They use the SCORE measure—the sum of correct answers to eight closed and one open question on a process model—to measure understandability.

In [167], Vanderfeesten *et al.* introduce the cross-connectivity metric. It measures the average strength of connection between all pairs of process model nodes. They empirically evaluate the metric using data of [101].

Mendling and Strembeck present a second experiment for identifying influencing factors on process model understandability [102]. This time, also content related factors (task labels) are analyzed.

In [135], Sanchez *et al.* emphasize the importance of process measurement for the determination of the maturity level according to the Business Process Maturity Model (BPMM) [113].

3.3 DISCUSSION OF RELATED WORK

The literature review of process measurement in the previous section is followed by a discussion about details of this related work in this section.

The first thing one can notice is the fact that authors define their proposed metrics using different process modeling languages (e. g., BPEL, BPMN, EPCs and Petri nets). Often, these definitions could also be adapted to other modeling languages—but in some cases, this is not possible.

The second notable fact is that the authors state that their metrics measure different concepts: So, several publications exist which try to measure process model complexity using complexity metrics (e. g., [21, 74]). Yet, one also finds articles dealing with process model quality and quality metrics (e. g., [165]) as well as process model performance (e. g., [63, 64]). Nevertheless, proper definitions of these terms are missing.

Also Sánchez González *et al.* criticize in their literature review that "the authors describe measures according to what they believe their measures quantify, and the majority of them do not follow any standard, or have previously performed a theoretical validation of the measures, which may lead to confusion" [136, p. 121]. Furthermore, they observe complexity as the most measured concept according to the corresponding authors' classification [136, pp. 120–121].

Thus, the remainder of this section deals with finding proper definitions for the terms "process model complexity" (Subsection 3.3.1) as well as "process model quality" and "process model performance" (Subsection 3.3.2).

3.3.1 *Process Model Complexity*

Today, the term "complexity" is used in many domains—not only in process measurement. Yet, only for very particular fields (e. g., computational complexity theory and Kolmogorov complexity, which are presented in the remainder of this subsection), mathematical definitions are available. Generally, only "philosophical definitions" exist. The Merriam-Webster's Collegiate Dictionary, for example, defines the adjective "complex" as "hard to separate, analyze, or solve" [106, p. 235] (as quoted in [74, p. 4]).

The remainder of this subsection is structured as follows: First, statements on process model complexity found during the literature review are reproduced. Afterwards, several aspects of complexity which are research objects in different research disciplines are presented. Next, some thoughts on the meaningfulness of measuring complexity are given. Finally, consequences regarding measuring process model complexity are drawn.

Complexity of Process Models

To the author's knowledge, Latva-Koivisto published the first paper [74] which deals with finding a complexity measure especially for process models. He cites [74, pp. 4–5] some interesting ideas about complexity by Edmonds:

"This means that it [complexity] is a highly abstract construct relative to the language of representation and the type of difficulty that concerns one." [41, p. 379] "The relevant type of 'difficulty' depends somewhat upon your goals in modelling. Different kinds of difficulty will result in different measures of complexity [...]." [41, p. 381]

Latva-Koivisto states that a measure of complexity is related to [74, p. 5]:

- the use of the measure,
- the kind of difficulty associated with the use,
- the objective of the analysis and
- the language of representation of the problem.

Cardoso defines process model complexity as "the degree to which a process[1] is difficult to analyze, understand or explain. It may be characterized by the number and intricacy of activity interfaces, transitions, conditional and parallel branches, the existence of loops, roles, activity categories, the types of data structures, and other process characteristics." [21, p. 202]

In [24, p. 36], he writes about the relation of complexity to other attributes according to his opinion: "A process[2] can be measured according to different attributes. The attribute that we will target and study is the complexity associated with BPEL processes[3]. Attributes such as time, cost, and reliability have already received some attention from researchers [...]."

1 "Process model" in the nomenclature of this thesis.
2 "Process model" in the nomenclature of this thesis.
3 "Process models" in the nomenclature of this thesis.

Aspects of Complexity

In several fields of research, different aspects of complexity are examined. For some of them, formal definitions exist—yet, for most of them, only informal textual descriptions of the concept are available.

Here, a short overview of some important aspects of complexity is given.

COMPUTATIONAL COMPLEXITY THEORY Computational complexity theory deals with the computability of problems and the runtime and/or space requirements of algorithms for solving computable problems.

Computability The definition of computability is based on the theoretical concept of a Turing machine which was introduced by Alan M. Turing in [155].

A *Turing machine* consists of an infinite tape used as memory, a read-write head for reading or writing symbols from/to the tape, a finite set of states (including a start state), a subset of accepting states and a so-called transition function. Depending on the symbol read from the tape and the current state, the transition function tells the "next step" of the Turing machine, i. e., the subsequent state, the symbol to write to the tape (by replacing the symbol read before) and whether the read-write head should move one symbol to the left or right on the tape after writing. When the Turing machine gets into an accepting state, it stops processing. [59, pp. 318–319]

For the definition of computability, one also needs the concept of computable functions. A function f is a *computable function* if some Turing machine computes the function, i. e., on every input x on its tape, the Turing machine halts with $f(x)$ on its tape [143, p. 210].

According to the *Church-Turing thesis*, the intuitive notion of algorithms and the Turing machine algorithms are equivalent, i. e., everything computable is computable by a Turing machine [143, p. 157].

Time and Space Complexity If one knows that a problem is computable, the next interesting step is to look at its runtime and space (memory) requirements. In doing so, one has to notice for some computable problems that they are intractable in practice.

Before time and space complexity are defined, three asymptotic notations [32, pp. 41–50] are introduced which are used later.

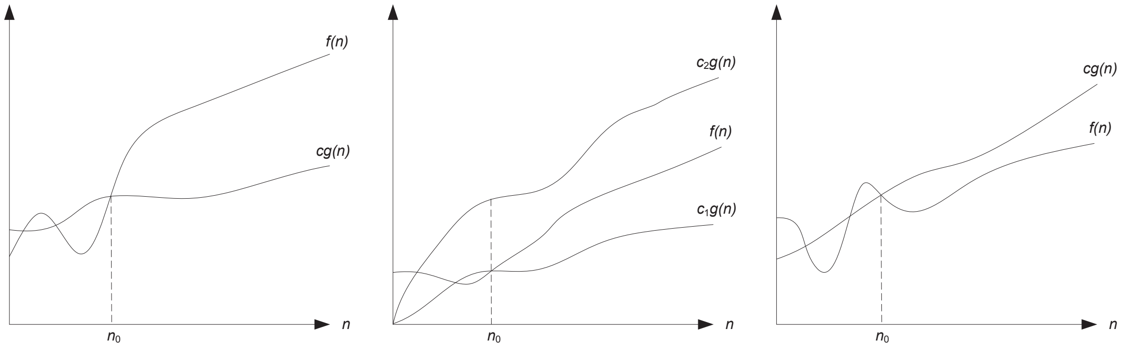
Definition 3.1 (Ω-notation) *For a given function $g(n)$, one denotes by $\Omega(g(n))$ the set of functions*

$$\Omega(g(n)) := \{f(n) | \exists c, n_0 > 0 \, \forall n \geqslant n_0 : 0 \leqslant c g(n) \leqslant f(n)\} \quad .$$

The Ω-notation is used to indicate an asymptotic lower bound of a function.

Definition 3.2 (Θ-notation) *For a given function $g(n)$, one denotes by $\Theta(g(n))$ the set of functions*

$$\Theta(g(n)) := \{f(n) | \exists c_1, c_2, n_0 > 0 \, \forall n \geqslant n_0 : 0 \leqslant c_1 g(n) \leqslant f(n) \leqslant c_2 g(n)\} \quad .$$



(a) $f(n) \in \Omega(g(n))$ (b) $f(n) \in \Theta(g(n))$ (c) $f(n) \in \mathcal{O}(g(n))$

Figure 3.1: Graphical examples of the (a) Ω-, (b) Θ- and (c) $\mathcal{O}$-notations. In each subfigure, the shown value of n_0 is the minimum possible value. [32, p. 43]

The Θ-notation states that a function lies between an asymptotic lower and upper bound.

Definition 3.3 ($\mathcal{O}$-notation) *For a given function $g(n)$, one denotes by $\mathcal{O}(g(n))$ the set of functions*

$$\mathcal{O}(g(n)) := \{f(n) | \exists c, n_0 > 0 \, \forall n \geqslant n_0 : 0 \leqslant f(n) \leqslant cg(n)\} \quad .$$

The $\mathcal{O}$-notation is used to indicate an asymptotic upper bound of a function.

Graphical examples of these three notations are depicted in Figure 3.1.

The *time complexity* of an algorithm is the number of computation steps $f(n)$ depending on the input size n. Often, a so-called *worst-case analysis* is performed, i. e., the longest possible runtime for each input size n is considered. As one is normally interested in the *asymptotic* runtime of an algorithm for large inputs, one uses one of the three asymptotic notations above to indicate lower or upper bounds for the runtime. [143, pp. 252–253]

Space complexity is correspondingly defined as the required memory size $f(n)$ of an algorithm depending on the input size n. Here again, one normally uses one of the asymptotic notations as one is mainly interested in the behavior for large inputs. [143, pp. 307–308]

The runtime depending on input size n for several "algorithms" which need different numbers of computation steps are listed in Table 3.1 under the assumption that 1,000,000 computation steps can be executed per second.

As one can see, the required runtime increases with very differing speed for the different time complexity classes. The last case with time complexity 2^n stands out extremely, as the runtime already reaches astronomic values for small input sizes. It grows exponentially—making such problems intractable in practice.

Thus, it is desirable to "find" and use algorithms with at least polynomial runtime. In computational complexity theory, the class **P** of problems with deterministic algorithm with polynomial runtime is defined for this purpose [143, pp. 260–267].

Unfortunately, many real world problems are not element of this class. They have a higher, for example, exponential runtime.

Table 3.1: Runtime depending on input size n for several "algorithms" with different number of computation steps (under assumption: 1,000,000 computation steps per second).

	runtime for following number of computation steps				
input size n	$\log_2 n$	n	$n \log_2 n$	n^2	2^n
10^1	0.000003 s	0.00001 s	0.00003 s	0.0001 s	0.001 s
10^2	0.000007 s	0.0001 s	0.0007 s	0.01 s	$4 \cdot 10^{16}$ years
10^3	0.000010 s	0.001 s	0.010 s	1 s	astronomic
10^4	0.000013 s	0.01 s	0.13 s	1.7 min	astronomic
10^5	0.000017 s	0.1 s	1.66 s	2.8 h	astronomic
10^6	0.000020 s	1 s	19.9 s	11.6 days	astronomic

For many of these problems, the class **NP** of problems with nondeterministic algorithm with polynomial runtime is defined [143, pp. 268–274]. The theoretical background for a nondeterministic algorithm is a nondeterministic Turing machine which is a Turing machine with a nondeterministic transition function. Such a nondeterministic Turing machine can be emulated by a "normal" deterministic one. Yet, it needs exponential instead of polynomial runtime.

An important subset of **NP** is the set of **NP**-complete problems. A problem is **NP**-complete if it is in **NP** and every problem in **NP** is polynomial time reducible to it. [143, pp. 275–287]

This class of problems contains many problems which are important in the real world. A large collection of such **NP**-complete problems can be found in [50].

KOLMOGOROV COMPLEXITY Kolomogorov complexity is a concept from algorithmic information theory. Informally, the Kolmogorov complexity of an object is the length of the shortest description of this object using a universal description language. An object whose shortest description is longer than that of another one is seen as more "complex" according to this concept. [78, pp. 1–3].

More details and possible applications of Kolmogorov complexity can be found, for example, in [78].

PRODUCT COMPLEXITY Product complexity is a more structural property of companies' products/services or product/service portfolios. It originates not only in the structure of one single product/service, but also in the number of different types (variants) of one product/service and their possible interdependencies.

In the automotive industry, for example, there is a huge number of possible variants of one single car model caused by the combinatoric explosion between the variants of car components as engine, color, seat covers, entertainment/navigation system, etc.

The communication industry—as a second example—has to deal with many different tariffs—either due to different customer groups (e. g., private and business customers) or due to old contracts, which continue to exist with a tariff which is no longer offered to new customers.

Beside product complexity in the narrow sense as explained above, the concept can also be transfered to similar problems as, for example, the number and types of

- different processes being executed in a company[4],
- supply chains for production and
- different running IT systems in a company.

Product complexity cannot be reduced without looking at the company's business environment as product complexity is linked with many different individual customer needs and wishes in a globalized market with high competitive pressure. Marti gives a good explanation of this dilemma [82, p. xxv]:

> "In the field of complexity management, the two dimensions of external and internal complexity receive special attention from theorists and practitioners alike. The two complexity dimensions pose a major challenge to enterprises because they require different and often conflicting treatment. External complexity (customer requirements, competitive forces, technological changes, etc.) pushes companies to broaden their product portfolios and introduce product variety, which in turn increases the enterprise-internal complexity (such as product complexity, organizational complexity, production complexity, etc.). Efforts to reduce internal complexity and slash the corresponding complexity costs typically require compromising the customization of products. This in turn complicates the task of differentiating oneself from competitors."

More details about the definition of the problem itself and possible approaches to deal with it can be found, for example, in [79] (structural complexity management, product complexity) and [13] (complexity management in supply chains).

NETWORKS

Introduction Networks consist of a set of nodes which are (partially) linked with each other. They can be used to model connected objects in the real world. Examples are social networks (people knowing and/or communicating with each other), infrastructure networks (e. g., water, electricity and telecommunication networks) and biological networks (e. g., biochemical, neural and ecological networks).

4 This has to be distinguished from the process complexity of these processes.

Besides modeling, important research questions in this area are network resilience (against random or intentional failures of parts of a network), epidemics on networks (spread of diseases or computer viruses) and dynamical systems on networks.

A good introduction into networks is the textbook by Newman [111].

Even though no real complexity measure is defined in this area, networks—and especially their dynamics—are an important aspect of complexity in the real world.

Network Models Over the years, several (theoretical) network models were proposed and analyzed in order to get better insights into the creation process and the properties of real world networks.

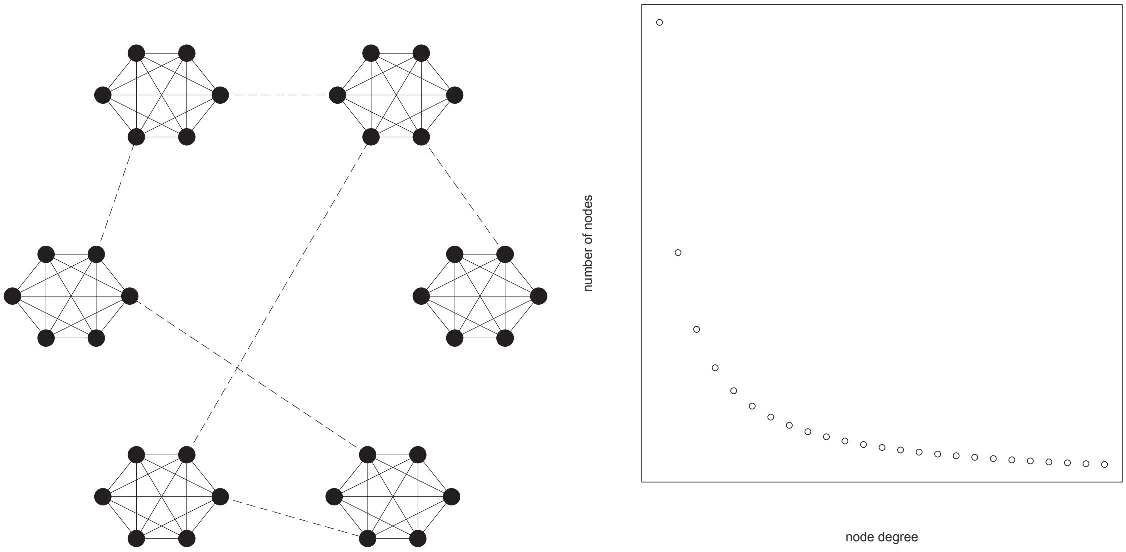
Random graphs (see [65] for details) are probably the oldest of these models. There are two main "construction methods" for getting such a graph with n nodes: (1) randomly choosing one of all graphs with n nodes and (2) starting with n nodes and no edges and than adding each possible edge between the nodes with a predefined probability. In recent years, it was discovered that random graphs do not show and explain some important characteristics which one can find in real world networks. Thus, other network models were suggested.

The development of the next network model, *small world networks*, was triggered by a meanwhile famous experiment conducted by Milgram [104, 154]. He asked 296 randomly chosen persons in the USA to send a letter to a target person in Massachusetts who they did not know. Yet, they were not allowed to send it directly to the target person but were told to send the letters only to a person they personally knew and ask him/her to proceed accordingly with the final goal to reach the target person. 64 letters (21.6%) finally reached the target person with an average of 5.2 intermediaries.

The reason for this small world phenomenon is a special network structure with many local clusters (almost everybody in a cluster knows each other as, e. g., relatives, friends or colleagues) and some so-called weak links between these clusters (one member of one cluster knows somebody of another one). This structure is depicted in Figure 3.2a.

Random graphs do not show—and thus—explain these properties. In [172], Watts and Strogatz propose an algorithm which produces networks with these small world properties.

Yet, also small world networks do not perfectly represent the entire characteristics of real work networks. Most node degree distributions of real networks follow a so-called power law, i. e., these networks have many nodes with few connected "neighbors" and only few nodes (so-called "hubs") with a large number of "neighbors". One can take, for example, airports and the flight connections between them as an example. There are many small airports which only offer flights to a few larger airports. On the other hand, there is a relative small number of large airports (hubs as, e. g., London Heathrow) which are connected by flights to many other hubs worldwide. This situation is depicted in Figure 3.2b. The resulting networks are called *scale-free networks*.



(a) Local clusters and weak links (dashed lines) in a small world network.

(b) Node degree distribution of a scale-free network.

Figure 3.2: Illustrations of properties of (a) small world and (b) scale-free networks.

In [7, 8], Barabási *et al.* suggest and analyze an algorithm which is able to produce such free-scale networks. The main principle of this algorithm is that a new edge is inserted between a new node and and an already existing node by a probability proportional to the current node degree of the existing node.

DYNAMICAL SYSTEMS A dynamical system is a system whose state changes over time. The system's state can be seen as a point in the so-called *phase space*—the space of possible states of the system.

Deterministic Chaos An important and interesting property of some dynamical systems is the so-called deterministic chaos. This effect was first observed by Lorenz [80] when he conducted computer experiments in the area of weather forecasting. The effect has become known as the "butterfly effect". It means that small changes of the initial conditions of a dynamical system (e. g., the flap of a butterfly) may cause large differences in its long-term behavior.

The effect is further explained with another example: population dynamics [83] [118, pp. 467–604].

Here, the Verhulst equation

$$x_{n+1} = a(1 - x_n)x_n \quad , n \in \mathbb{N}_0, x_0 \in [0, 1], a \in [0, 4] \tag{3.1}$$

models the size of a population over time. This population size is a number between 0 and 1. Parameter a is a growth factor.

For small values of parameter a (e. g., $a = 2$), the population size converges to one value independently from the start value x_0 (see Figure 3.3a). For larger values of a (e. g., $a = 3.3$), the long-term population size alternates between two values (see Figure 3.3c). If one further increases a, instead of alternation between two values, an alternation between four, then eight, etc. different values occurs. For $a = 4$, finally, a non-periodic behavior of the population size can be observed. Furthermore, the time series becomes totally different in the long term even for very small differences of the start value x_0 (see Figure 3.3b and 3.3d).

This effect is the reason for the term "deterministic chaos". Even though the underlying equation of the iteration is known and all iteration steps can be deterministically computed if one knows the start value, the long-term behavior of the iteration is "chaotic" as one cannot determine the start value without any small error in practice.

The long-term behavior of the Verhulst equation for the different values of a can be visualized using the *Feigenbaum diagram* (see Figure 3.4). Here, only the long-term values of the iteration are displayed depending on the value of a. One can observe the behavior described above: First one fix point, then two, four, eight, etc. alternating values and finally the chaotic behavior for large values of a.

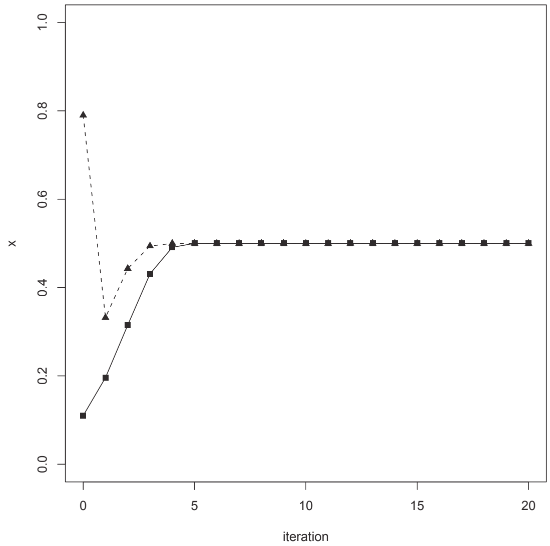
Control Theory Control theory deals with the question whether it is possible to bring a dynamical system into a certain state. If it is possible, the identification of the necessary control inputs is the second issue. An overview of control theory can be found in [14]. An analysis of the computational complexity of different control problems is given in [145].

Consequences for economic processes Also economic processes can be seen—and consequently analyzed—as dynamical systems. Examples are nonlinear dynamics in production [124], the "bullwhip effect" in supply chains [76] and the beer distribution game [147].

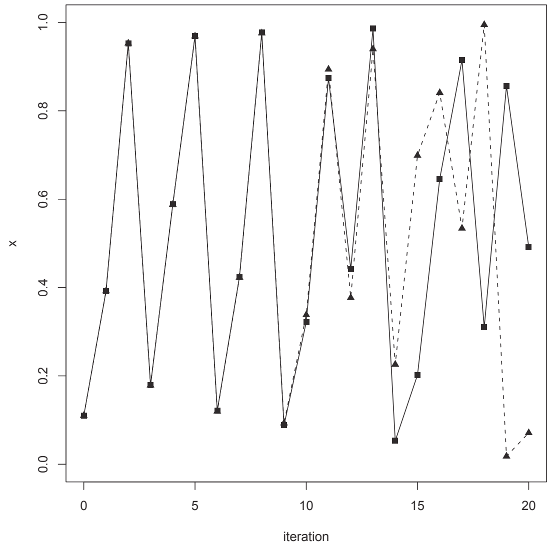
COMPLEX PROBLEM SOLVING This field of research examines the behavior of humans while solving complex problems. Funke lists several properties of complex problems [47, pp. 186–187]:

- intransparency: Only some variables are directly observable. The state of the underlaying system has to be inferred only from this observations.
- multiple goals: Multiple—partially contradictory—goals exist.
- connectivity of variables: Changing one variable has a big influence on the states of other variables.
- dynamic development: The state of the underlying system changes over time—even without active interventions.
- time-delayed effects: The effects of interventions occur only with a time delay.

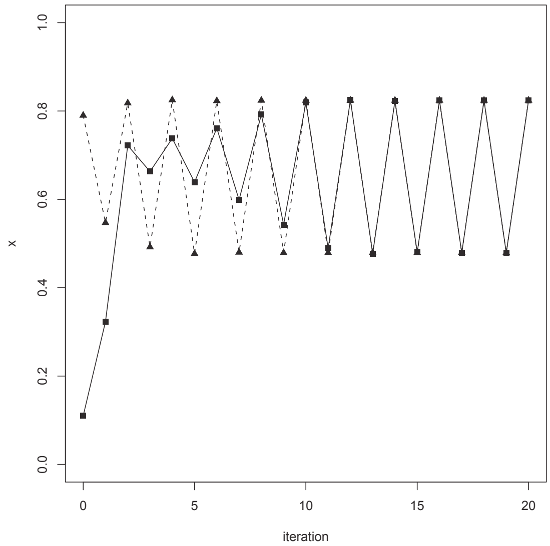
Using different experimentation systems (see, for example, [47, pp. 188–204] for an overview), several influencing factors (e. g., [45, pp. 20–21] [47, pp. 204–206]



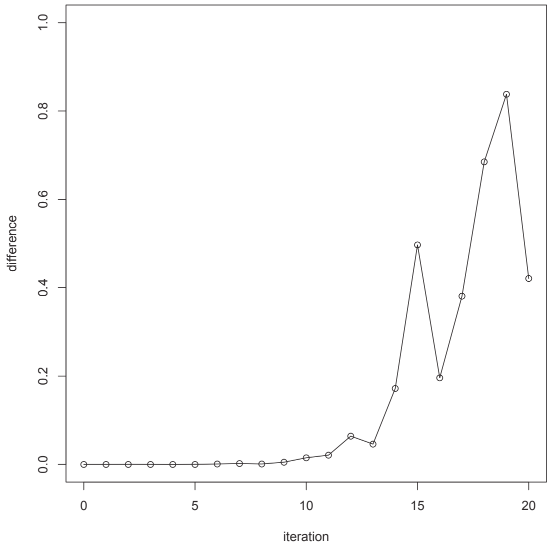
(a) Time series for $a = 2$ and $x_0 = 0.11$ (squares) or $x_0 = 0.79$ (triangles) respectively.



(b) Time series for $a = 4$ and $x_0 = 0.11$ (squares) or $x_0 = 0.11001$ (triangles) respectively.



(c) Time series for $a = 3.3$ and $x_0 = 0.11$ (squares) or $x_0 = 0.79$ (triangles) respectively.



(d) Absolute value of difference between iteration steps of the two start values x_0 of Subfigure (b).

Figure 3.3: Different time series of the Verhulst equation.

[48, pp. 250–251]) were examined in studies (see [48, pp. 251–260] for a review of studies according to the examined factor).

There are two main types of conducted studies: studies using computer-simulated mircoworlds [15] as well as studies using finite state automata or linear equation systems [17, pp. 42–53]. For the first type, a system (e. g., a city) is modeled in the computer. The test subject (e. g., acting as the city mayor) can influence some variables of the simulated system (e. g., the tax rate) in order to achieve a certain goal.

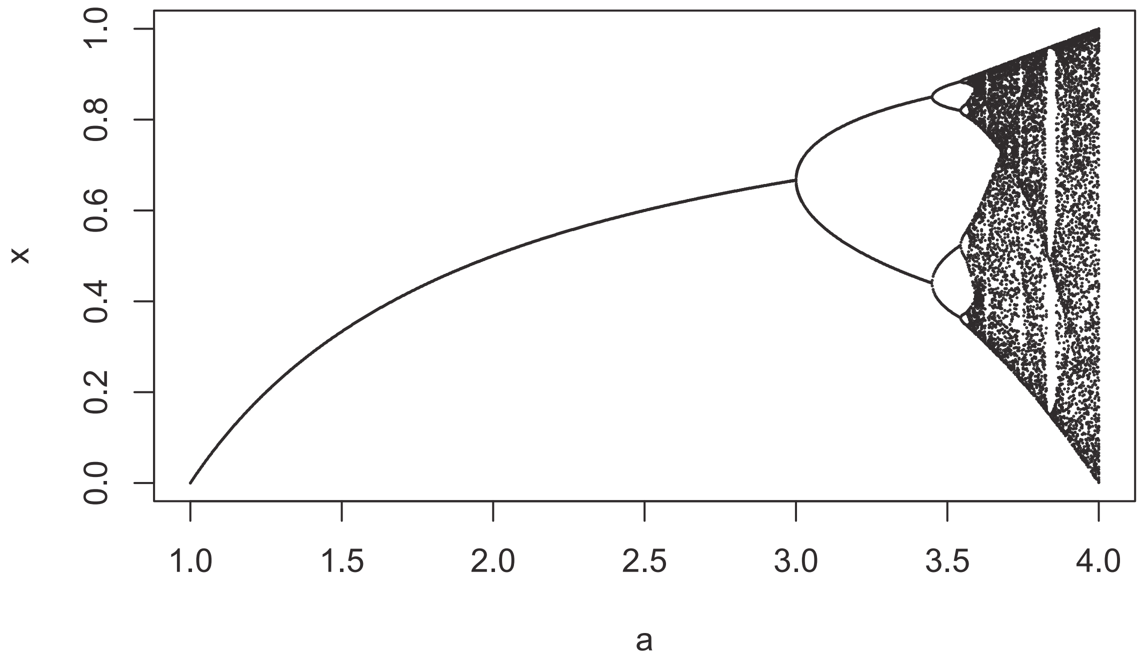


Figure 3.4: Feigenbaum diagram of the Verhulst equation.

Dörner and Wearing propose a systematic approach for problem solving [40, pp. 67–70]: First, a concrete and measurable goal has to be formulated. Then, a mental representation of the system has to be gained. Based on a prediction of the system's future development, own actions have to be planned. After these actions are conducted, their effects have to be assessed in relation to the planned goal. Depending on the outcome of this assessment, some of the previous steps have to be redone in a modified manner.

Some important empirical results for real behavior of humans are presented by Dörner and Wearing in [40, pp. 72–83]. It can be noticed that humans in general have big problems when solving complex problems and that they make some typical mistakes.

Dörner gives a rather informal overview of this field of research in [39].

Is Measuring Complexity Meaningful?

The variety of different aspects of complexity shown above suggests that it is at best possible to measure one of these complexity aspects and not complexity in total. Subsequently, the question concerning the informative value of such a metric arises. Which information does an absolute value of such a metric give us without any further context information?

Ashby states regarding this question [4, p. 1]:

> "The word 'complex', as it may be applied to systems, has many possible meanings, and I must first make my use of it clear. There is no obvious or pre-eminent meaning, for although all would agree that the brain is complex and a bicycle simple, one has also to remember that to a butcher the brain of a sheep is simple while a bicycle, if

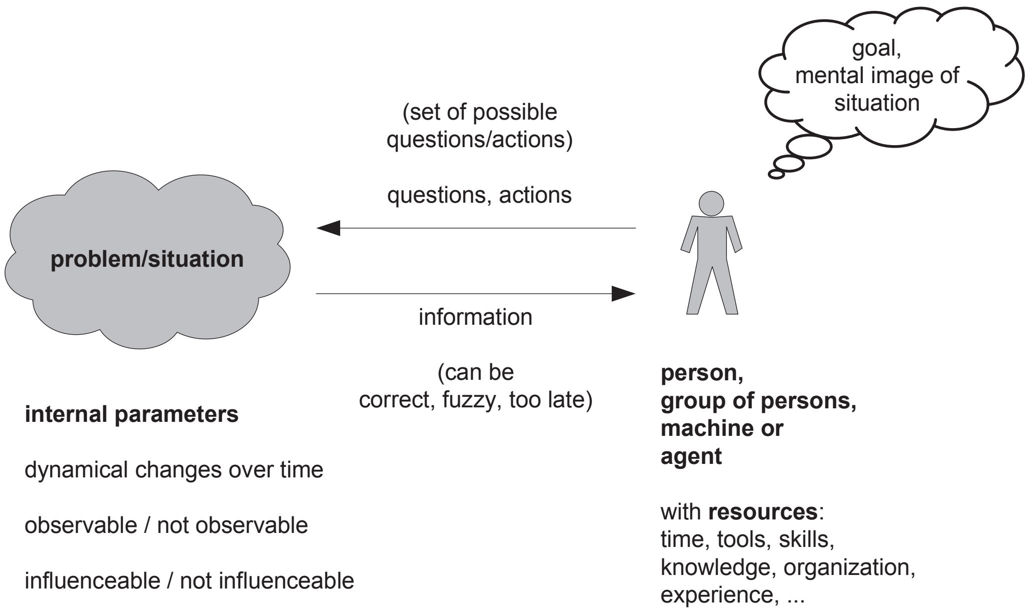


Figure 3.5: Complexity between problem/situation and agent [141, p. 30]. "Complexity is in the eye of the beholder."

> studied exhaustively (as the only clue to a crime) may present a very great quantity of significant detail.
>
> Without further justification, I shall follow, in this paper, an interpretation of 'complexity' that I have used and found suitable for about ten years. I shall measure the degree of *'Complexity' by the quantity of information required to describe the vital system*. To the neurophysiologist the brain, as a feltwork of fibers and a soup of enzymes, is certainly complex; and equally the transmission of a detailed description of it would require much time. To a butcher the brain is simple, for he has to distinguish it from only about thirty other 'meats', so not more than $\log_2 30$, i. e. about 5 bits, are involved. This method admittedly makes a system's complexity purely *relative to a given observer;* it rejects tha [*sic*] attempt to measure an absolute, or intrinsic, complexity; but this acceptance of complexity as something in the eye of the beholder is, in my opinion, the only workable way of measuring complexity."

Thus, he emphasizes that complexity "originates" through the connection between a problem and its observer—instead of being an absolute or intrinsic property.

Seese concretizes this statement in a more detailed model [141, pp. 11, 29-30] (see Figure 3.5). Also here, complexity "originates" between a problem or a situation on the one hand and an observer (e. g., a person, a group of persons, a machine or a (software) agent) on the other hand.

The problem/situation has a set of—initially unknown—internal parameters which can dynamically change over time. These parameters can be observable by the observer or not as well as influenceable or not.

The observer on the other hand has a set of resources (e. g., time, tools, skills, knowledge, organization, experience) and of possible actions. The observer can gather information about the problem/situation. This information can be correct or incorrect, fuzzy or simply coming too late.

Based on the observer's developing mental image of the situation, he/she can try to reach his/her goal by conducting adequate actions.

According to this model, complexity is a somewhat "subjective" property of a problem or situation—mainly influenced by the observer's resources.

In his blog entry *Complexity is in the Eye of the Beholder* [173], Weber explicates this thought that complexity depends on the observer's resources:

> "The question is, where is the boundary between complex and simple? I think that this border only exists in the human mind. In fact the border is movable and depends on the individual mind.
>
> Nothing is complicated to Nature (the universe). Everything just is. Confusion and complication only arise in the human mind while trying to understand the universe.
>
> Complexity is a function of our capacity to assimilate, store and process data. If we were suddenly twice as smart, some complicated problems would seemingly become simple problems. The problems themselves did not change, only we changed."

Consequences

Already some authors of process measurement literature stated that there are different aspects of process model complexity. This statement was supported by the presentation of different aspects of general complexity. These complexity aspects can also be analyzed for process models:

- **Computational complexity theory**

  time complexity of algorithms for process models (e. g., for execution, analysis, optimization)

- **Product complexity**

  process portfolio of a company, interacting IT infrastructure

- **Networks**

  process models seen as networks, communication network of persons involved in a process

- **Dynamical systems**

  changing state during process model execution

- **Complex problem solving**

  problems concerning human interaction with process models

This fact suggests that it is at best possible to measure one of these complexity aspects and not process model complexity in total.

Furthermore, it was shown that complexity can be seen as something which "originates" through the connection between the problem and its observer instead of an intrinsic property of the problem itself. Because of this, it is even hardly possible to find a complexity metric for a special complexity aspect. The reason that this works so well in computational complexity theory is the fact that here, the "observer" with its "resources" can be mathematically defined by using Turing machines. For most of the other presented complexity aspects—especially those with human interaction involved—, such a stringent formal definition is impossible—resulting in a rather vague complexity concept.

3.3.2 *Process Model Quality and Performance*

Process Model Quality

According to Kan, quality can be defined as "conformance to requirements" or "fitness to use" [69, pp. 2]. He mentions two views of quality: the costumer's view on quality and the company's view on quality. For a costumer, quality is the "perceived value of the product he or she purchased, based on a variety of variables such as price, performance, reliability, and satisfaction" [69, pp. 3]. For a company, quality means that the customer's requirements on the product quality are fulfilled *and* that its own production costs are lower than the price for selling the product.

Adapted to process model quality, one can give the following definition.

Definition 3.4 (Process model quality) *For a customer, process model quality means that the process model's outcome (a product, a piece of information or a decision) is correct, arrives within adequate time and to an adequate price.*

For a company, all these factors also belong to quality—but additionally, the price for the process model execution must be lower than the price which the customer is willing to pay and, furthermore, the process model should be easily adaptable to changed circumstances.

Process Model Performance

Besides process model complexity and quality, process model performance is a third concept which can be found in process measurement literature. In [63, 64], Jansen-Vullers *et al.* suggest a performance measurement framework consisting of the four dimensions time, cost, quality and flexibility. Quality is separated into internal and external quality in their framework. As their quality concept is practically equivalent to Kan's quality concept, and Kan's quality concept can include time, cost and flexibility, it is proposed here to only use the term "process model quality"—but therefore, in the extensive meaning by Kan described above.

3.4 PROCESS MEASUREMENT APPROACH

In this section, the measurement approach used throughout this thesis is introduced.

The foundation of this approach is the observation made in the previous section that it is impossible to find a stringent definition of process model complexity. The reason is that this term has so many different aspects as has been shown in Subsection 3.3.1.

Instead, a prediction system measurement approach which avoids concrete definitions of process model complexity and quality is used. It is adapted from software measurement where a similar problem (definition of software complexity) exists.

Before this measurement approach can be presented, measurement and prediction systems have to be explained in Subsection 3.4.1. Afterwards, the actual measurement approach is introduced (Subsection 3.4.2). Finally, the necessary validation steps for such a prediction system are shown in Subsection 3.4.3.

3.4.1 *Measurement and Prediction Systems*

According to Fenton and Pfleeger, the usual meaning of measurement is "that we wish to assess some entity that already exists. This measurement for assessment is very helpful in understanding what exists now or what has happened in the past." [43, p. 42]

Based on this statement, they define measurement systems as followed [43, p. 104]:

Definition 3.5 (Measurement system) *A measurement system is used to assess an* existing *entity by numerically characterizing one or more of its attributes.*

"However, in many circumstances, we would like to predict an attribute of some entity that does not yet exist." [43, p. 42] For example, Balasubramanian and Gupta mention that interesting process model performance measures "like process cost, cycle time, process throughput and process reliability [...] can be calculated only after process execution and are of limited use in predicting future process performance[5]" [6, p. 680]. Consequently, they note the importance of indicators for process model performance at the pre-implementation stage [6, pp. 680–681]. Cardoso emphasizes the importance "to develop methods and measurements to automatically identify complex processes[6] and complex areas of processes[7]" [21, p. 202].

For that second purpose of measurement, Fenton and Pfleeger define prediction systems [43, p. 104]:

5 "Process model performance" in the nomenclature of this thesis.
6 "Process models" in the nomenclature of this thesis.
7 "Process models" in the nomenclature of this thesis.

Definition 3.6 (Prediction system) *A prediction system is used to predict some attribute of a* future *entity, involving a mathematical model with associated prediction procedures.*

Besides the use for *future* entities, as stated in the definition of Fenton and Pfleeger, prediction systems can also be used to predict some attribute of an *existing* entity which is measurable only in a very laborious manner.

3.4.2 *Process Measurement Approach—An Adaptation from Software Measurement*

As seen in Section 3.3, it is a major problem of the existing process measurement literature to propose numerous metrics and measures for which it is claimed that they measure process model complexity, quality and performance without giving a proper definition of these terms . Thus, it is hardly possible to use such metrics and measures in practice if one even cannot tell exactly what one is measuring and which consequences theses values have.

In this subsection, the process measurement approach used throughout this thesis is introduced. It tries to avoid the problems of the existing process measurement literature by providing a theoretical framework in which the existing works can be integrated.

All presented suggestions for a definition of process model complexity (Subsection 3.3.1) show that complexity is no such property like length or mass which can be measured directly using meters or kilograms respectively. So, a more "philosophical" discussion (cf. Cardoso's definition in Subsection 3.3.1) starts which does not solve the underlying problem.

All authors who propose a metric claiming it would measure process model complexity—yet, owing a proper definition of the term—apply what Weinberg and Weinberg call the "Humpty Dumpty Method" [174, p. 313]. The name is based on the quote

> "When *I* use a word," Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean—neither more nor less."
>
> "The question is," said Alice, "whether you *can* make words mean different things."
>
> "The question is," said Humpty Dumpty, "which is to be master—that's all."

from Lewis Carroll's novel *Through the Looking-glass and What Alice Found There* [25, p. 114]. According to Weinberg and Weinberg, these authors simply define "complexity in such a way that it does exactly what you want it to do" [174, p. 313].

Therefore, an alternative measurement approach is suggested here. It is inspired from software measurement where a similar dilemma concerning the definition of the term "software complexity" exists. There, a prediction system

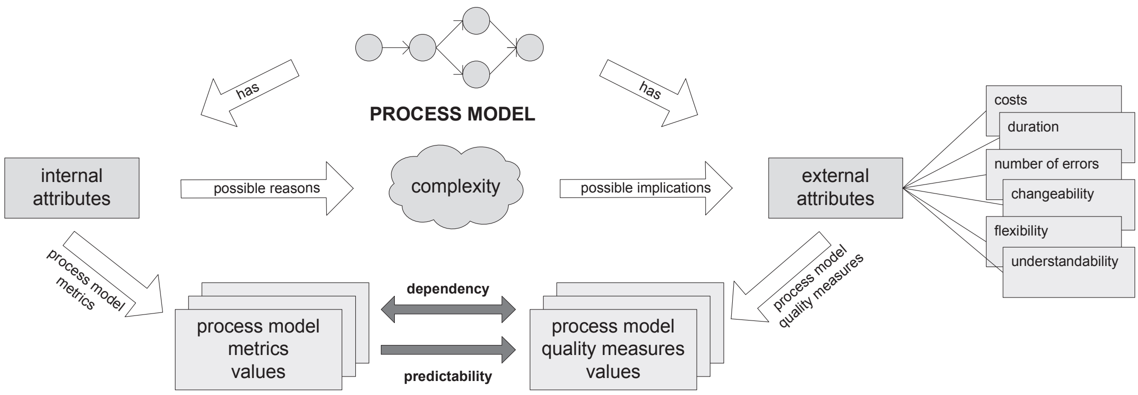


Figure 3.6: Prediction system measurement approach for process measurement.

measurement approach helped to overcome this problem (see, e. g., [43, pp. 74–75]). The measurement approach can be adapted to process measurement as described in the remainder of this subsection.

What is more important than process model complexity itself—especially for economic reasons—are the *implications* of this complexity like costs, time, duration, number of errors, changeability, flexibility, understandability, etc. (aspects of process model quality according to Subsection 3.3.2). All these quantities have the advantage to be quantifiable and measurable.[8] The disadvantage is that they can only be measured after the process model has been implemented and executed.

To overcome this problem, the adaptation of the prediction system measurement approach from software measurement comes into play. It is sketchily depicted in Figure 3.6.

A process model has *internal* and *external* attributes.

Internal attributes can be measured purely in terms of the process model, separately from its behavior [43, p. 74]. These attributes (e. g., structural properties like the number of activities/tasks) *could* contribute to the process model complexity. Numerous internal attributes are imaginable and appropriate metrics have already been proposed (especially for structural properties) or can be defined. Using these metrics, one gets corresponding metric values of the process model.

External attributes can be measured only with respect to how the process model relates to its environment [43, p. 74]. The external attributes like costs, time, number of errors, etc. are possibly affected by the process model complexity and are measurable. External attributes are aspects of process model quality (and performance respectively).

The last step of the proposed approach is also the most important one: One has to show a dependency between the metric values and the measure values of the external attribute. If such a dependency exists, the metric can be used as a predictor for the external attribute at a much earlier time. Thus, a prediction system is formed.

8 For, e. g., costs, time, duration, number of errors, this is trivial to see. But also attributes like changeability, flexibility and understandability are measurable if one looks at the costs, time, number of errors, etc. it takes to change or understand a process model. Fenton and Pfleeger give some ideas for measuring maintainability in [43, pp. 354–355].

Because of the proposed measurement approach, the term "process model metric" is recommended for the metrics measuring internal attributes instead of "complexity metric" or "complexity measure" as complexity itself is not measured. For the measures measuring external attributes, the term "process model quality measure" is recommended. Also consider the discussion about the terms "measure" and "metric" at the end of Section A.1 in this context.

As told above, the existing metrics and measures can be integrated into the suggested measurement approach: Most metrics from Section 3.2—including those which are told to measure process model complexity—measure structural (i. e., internal) attributes. Thus, they are process model metrics in the nomenclature of this thesis. Measures which were developed for measuring process performance in other sciences years ago are process model quality measures in the nomenclature of this thesis.[9]

At this point, it is appropriate to take a look back to the ideas of Edmonds and Latva-Koivisto (see Subsection 3.3.1). As they suggest, there is not one single complexity metric which can measure every aspect of the process model complexity in the proposed approach. Instead, several different pairs of metrics for internal and measures for external attributes can exist forming a prediction system and so representing *one* of the existing links between *reasons* and *implications* of process model complexity.

In this context, Cardoso is believed here to be subject to a misconception when he puts complexity at the same "level" as attributes such as time, cost and reliability (see end of paragraph *Complexity of Process Models* in Subsection 3.3.1). Instead, complexity can be the reason for these attributes.

3.4.3 *Validation*

Before the presented measurement approach can be applied in practice, the corresponding two measurement systems and the prediction system have to be validated. The respective steps are explained in this subsection.

Validation of Measurement Systems

The statements in this subsection apply both to process model metrics and process model quality measures, unless otherwise stated.

OBJECTIVE/SUBJECTIVE MEASURES For the process model quality measures measuring external process model attributes, there are two kinds of measures: objective and subjective measures. Objective measures are performance-based and measure, e. g., time, costs and number of errors. Subjective measures are perception-based and measure, e. g., how difficult a subject rates a process model. See Section A.3 for more details about the measurement of such non-physical properties.

9 Note that some metrics from Section 3.2 which are called "process quality metric" by their authors (e. g., in [23]) are process model metrics in the nomenclature of this thesis.

REQUIREMENTS OF METRICS/MEASURES The following requirements reliability and validity are relevant for the measurement of all (non-physical) properties and are not process measurement specific. For more details see Subsection A.3.2.

- *reliability/consistency*: Metric/measure values obtained by different observers of the same process model have to be consistent [74, p. 3] [21, p. 202]. For mathematically defined process model metrics, this is automatically fulfilled. But for process model quality measures measuring external process model attributes like understandability, the exact measurement conditions are important to fulfill this requirement. Kan gives a good example [69, pp. 70–71]: If one wants to measure the height of a person, the measurements should be taken at a special time of day (e. g., always in the morning) and always barefooted. Otherwise, the measure values of the same person could vary a lot.
- *validity*: According to Kan [69, pp. 71–72], validity can be classified into *construct validity* and *content validity*. The first checks whether the metric/measure really represents the theoretical concept to be measured (e. g., is church attendance a good measure for religiousness?). The second checks whether the metric/measure covers the range of meanings included in the concept (e. g., a test of mathematical ability for elementary pupils cannot be limited to addition but should also include subtraction, multiplication, division and so forth).
- *computability/ease of implementation/automation*: A computer program can calculate the value of the process model metric in finite time—and preferably quickly. The difficulty of the implementation of the method which computes the process model metric is within reasonable limits. [74, p. 4] [21, p. 202]

  This requirement, which was found in the process measurement literature, only applies to process model metrics (measuring internal process model attributes) which are mathematically defined and can be computed automatically.

These requirements are important as "good predictive theories follow only when we have rigorous measures of specific, well-understood attributes" [43, p. 108].

Validation of Prediction Systems

According to the adapted measuring approach (see Subsection 3.4.2), a proposed process model metric has to be validated against a concrete external attribute (process model quality measure). The goal of such a validation is to show a dependency between the process model metric values and the corresponding external attribute in question. As Fenton and Pfleeger state, "[r]ather than being a mathematical proof, validation involves confirming or refuting a hypothesis" [43, p. 104].

The validation can be done either by using existing data (e. g., from log files) or by conducting experiments (to get new data). Fenton and Pfleeger emphasize the advantages of experiments as the level of control and the level of replication are much higher [43, p. 120]. Basics about empirical investigations (e. g., experimental design among other things) can be found in [43, pp. 117–152] and Section B.3.

As there can be different kinds of dependencies (e. g., positive linear, negative linear and many forms of non-linearity) [69, pp. 77–80] (see Figure 3.7 for examples), scatter plots are a good method to visually search for any form of dependency (also non-linear). The next step is to use a measure of correlation like Spearman's rank correlation coefficient (see Section C.2) or Pearson's product-moment correlation coefficient (see Section C.1). If a dependency is found, one can also try to find an equation which mathematically describes the dependency (e. g., using linear regression, multivariate regression, non-linear regression). [43, pp. 199–200]

In the field of software measurement, IEEE Standard 1061 (IEEE Standard for a Software Quality Metrics Methodology) gives a method for validating prediction systems [60, pp. 10–13] which checks among correlation also additional properties as tracking, consistency, predictability, discriminative power and reliability[10].

MEASUREMENT DIMENSIONS Prediction systems are only valid for very special conditions. According to Fenton and Pfleeger, "validation must take into account the measurement's purpose; a measure may be valid for some uses but not for others" [43, p. 107].

Consequently, the conditions during validation and the later use of the prediction system must be consistent. The following four "measurement dimensions" are generally important conditions. For special cases, additional conditions may exist.

- **Process model metric (internal process model attribute)**
  The process model metric defines the "measurement rule" for quantifying the chosen internal process model attribute.

- **Process model quality measure (external process model attribute)**
  The external process model attribute (probably affected by process model complexity) whose value correlates with the process model metric value.

- **Subjects**
  Which persons are involved in the measurement? Possible persons are, e. g., process designers, process analysts, programmers and end-users (i. e., the employees working in the process). As these persons have different skills and different views of the process, the values of the same external process model attribute (e. g., time, costs and number of errors) can differ a lot depending on the involved persons (subjects).

10 In [60], "reliability" has another meaning than the homonymous requirement for valid measurement systems presented earlier in this subsection.

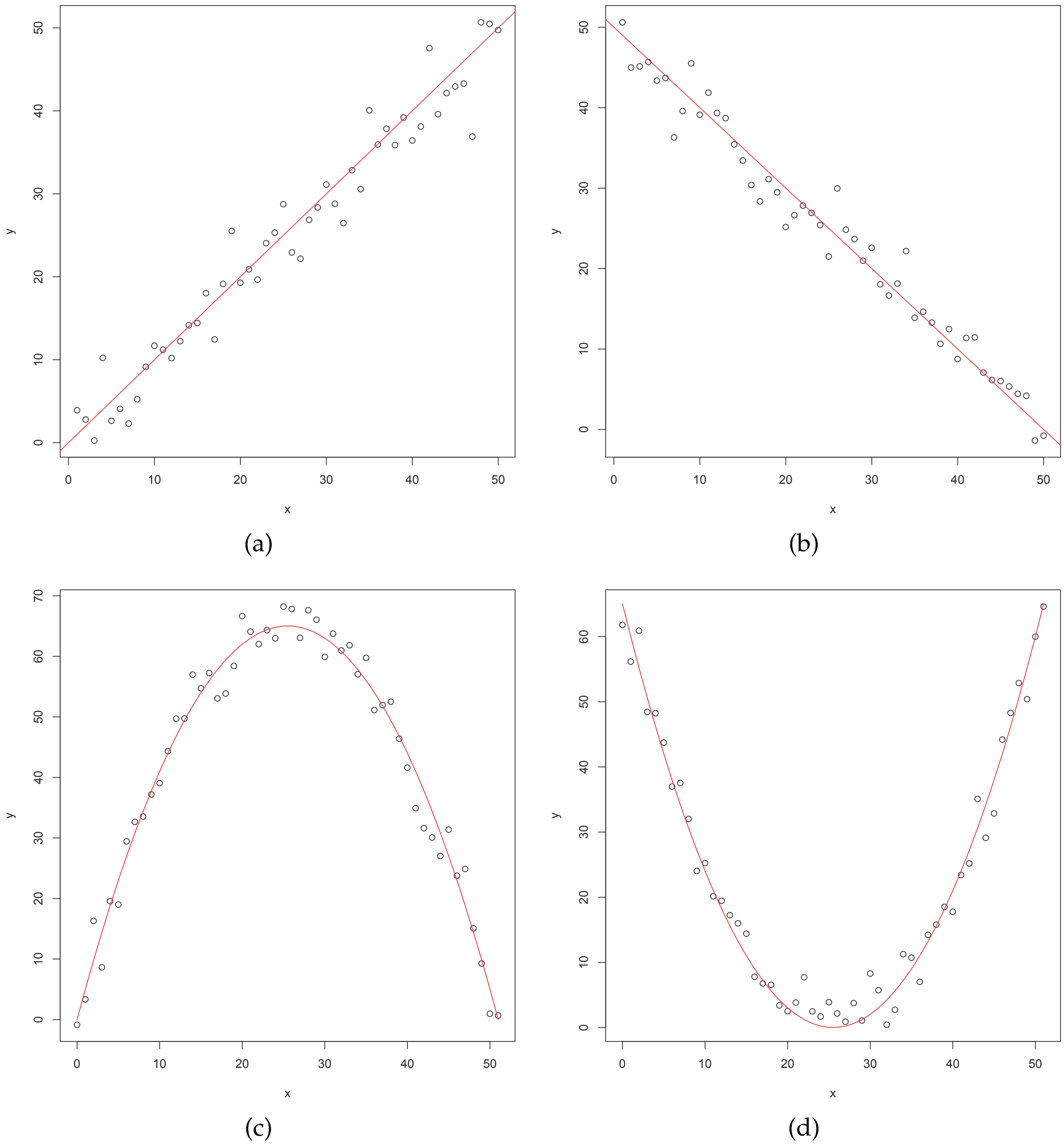


Figure 3.7: Four possible types of dependencies between two variables [69, p. 79].

- **Process phase**
  As in software engineering, a process life cycle consists of several process phases: modeling, analysis, implementation, deployment, execution, maintenance and modification of the process model (cf. the BPM lifecycle in Subsection 2.1.2).

  In contrast to software engineering, process model execution is an additional phase in which problems can occur. After a software program is implemented, no new errors are introduced by executing the program. But as process models are executed (at least partially) by humans, additional errors can occur while executing a process model.

INTERPRETATION OF PROCESS MODEL METRIC VALUES After having validated a prediction system, one has to identify the range or threshold between "good" and "problematic" metric values of the process model metric contained in the prediction system. Only with this knowledge, one can detect problematic process models and take countermeasures.

3.5 APPLICATION OF PROCESS MEASUREMENT

Having established valid measurement and prediction systems for process models, the question arises what to do with these metrics.

In this section, several possible applications of process measurement are presented. It can be used both for process models that are newly implemented and for finding and dealing with "those existing processes[11] that are good candidates for improvement and simplification, or even complete reengineering" [74, p. 3].

3.5.1 *Selection of Metrics and Measures*

As there exist numerous metrics for process models, first, one has to select proper metrics for the considered "problem". Using all available or accidentally selected metrics would just generate numerous numerical values without any purpose for the considered "problem".

Basili *et al.* propose an approach for the selection of metrics for software measurement—the *Goal Question Metric (GQM) approach* [9]. This approach is also applicable for process measurement and can be used both for selecting process model metrics and process model quality measures.

The approach has three levels: conceptual level (goal), operational level (question) and quantitative level (metric). At the first level, a precise goal is defined. A set of questions for assessing and achieving the goal is established at the second level. At the third level, a set of metrics is assigned to each question in order to quantitatively answer the questions. The resulting GQM model has a hierarchical structure with possibly several goals, multiple questions per goal and several metrics per question. A metric can be assigned to multiple questions. Figure 3.8 shows an example for such a hierarchical GQM model structure.

Using this top-down approach, only useful metrics, measures and possibly prediction systems for the current "problem" are selected and no unnecessary metric/measure values are collected.

3.5.2 *Different Measurement Purposes*

For the field of software measurement, Fenton and Pfleeger mention three different measurement purposes [43, pp. 13–14] which can also be adopted to process measurement.

11 "Process models" in the nomenclature of this thesis.

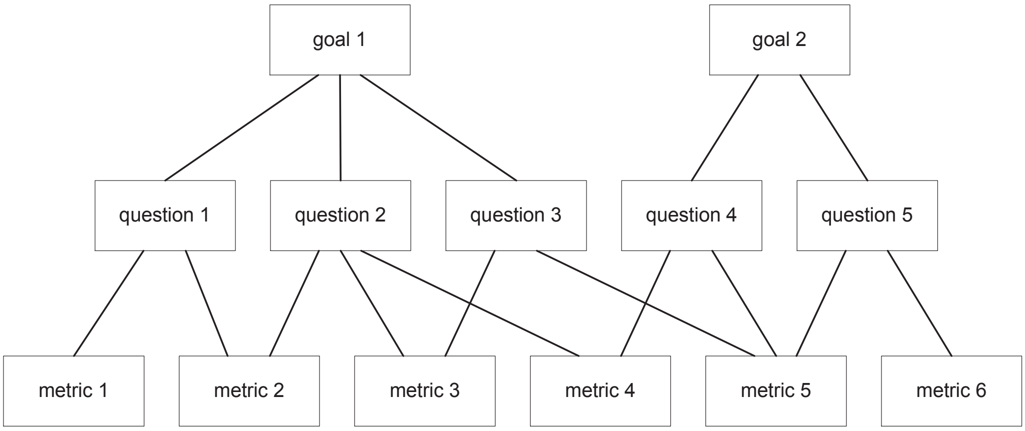


Figure 3.8: Hierarchical GQM model structure [9, p. 529].

Understand

For this first purpose, a process model is only measured using different selected metrics and/or measures to get a better understanding about what happens within this process model. Afterwards, no changes or concrete actions are conducted. Through this, the process model can be compared while being modified over time (modifications not caused by process measurement!) or it can be compared with other process models within the same company.

For this purpose, only valid measurement systems are necessary.

Control

Here, a process model is also measured, but not with the general goal to change it. Instead, the findings of the measurement are used to manage and control the assignment of employees in testing and finding errors. Consequently, the limited manpower can be deployed more intensively in "problematic" process models or process model parts. Possible actions can comprise, for example, test cases or inspections as in software engineering. A process model is only changed in order to fix a found error.

For this purpose, valid prediction systems are necessary.

Improve

For the third purpose, a process model is measured. If a bad quality is measured (for existing process models) or predicted (for new process models), the process model will be changed in order to improve it. So, the goal is to reduce *unnecessary* complexity within the process model. As complexity itself is not measured in the process measurement approach used in this thesis (see Subsection 3.4.2), reducing complexity means changing the process model's internal attribute(s) in such a way that the external attribute(s)—the process model quality—increases according to the prediction system.

One has to consider that the complexity of a process model cannot be reduced arbitrarily [19, p. 117]. Here, one must distinguish between the *intrinsic* complexity of a process[12] and the complexity of a process model. The chosen process model is not independent of the overall problem. So, it has a "natural" minimal complexity. This fact was already referred to by Fenton and Pfleeger for software measurement [43, p. 267].

One can compare this with an example of an analogous problem—runtime complexity of algorithms: The general problem of sorting has a (mathematically proven) minimal complexity of $\Omega(n \log n)$ [32, pp. 165–167]. The Heapsort sorting algorithm, for example, has complexity $\mathcal{O}(n \log n)$ [32, pp. 127–137]. But nevertheless, more inefficient sorting algorithms exist (e. g., Insertion Sort with $\mathcal{O}(n^2)$, see[13]).

But even if a reduction of complexity is possible and would probably cause higher quality, one should first compare the costs for the process model change with the expected increase of incomes with this process in order to decide whether to actually implement the changes.

As the quality for the changed process model is predicted within this purpose, valid prediction systems are necessary.

3.6 ASSESSMENT OF EXISTING WORK

In this section, the existing process measurement work is assessed according to the theoretical framework introduced in this chapter.

Most of the proposed process model metrics (measuring internal attributes) are adapted from software metrics. As they all have a mathematical definition, they fulfill the reliability/consistency and computability requirements of Subsection 3.4.3. So, they form valid measurement systems.

Only four works dealing with validating prediction systems could be found in the literature:

- **Cardoso: Validation of control-flow complexity metric (CFC) [22]**

  Cardoso conducted a laboratory experiment and computed Spearman's rank correlation coefficient between the CFC values of process models and the subjective complexity values stated by the experiment's subjects. He could show a statistically significant correlation. But it is not clear how this subjective complexity is connected to any external process model attribute (process model quality). So, it is no practically relevant prediction system.

- **Mendling *et al.*: Using process model metrics for predicting faulty EPCs [96–100]**

  604 EPC process models of the SAP Reference Model were analyzed using the verification tool WofYAWL. Through this, 34 faulty process models were

12 The use of the term "intrinsic complexity" may be confusing here as in Subsection 3.3.1, it was stated that complexity is no intrinsic property of a problem—yet, it "originates" between the problem and its observer. Here, "intrinsic complexity" is meant as the "natural" minimal complexity in this "subjective" meaning.

13 According to [32, pp. 23–27], Insertion Sort has worst-case running time of $\theta(n^2)$. Using Theorem 3.1 in [32, p. 46], running time $\mathcal{O}(n^2)$ follows.

identified. Multivariate logistic regression was used to predict faulty process models. As all metrics fulfill the requirements and the correlation could statistically be shown, it is a valid prediction system.

- **Cardoso, Mendling, Reijers, Strembeck, van der Aalst, Vanderfeesten: Influencing factors on process model understandability [101, 102, 167]**

  In [101], Mendling *et al.* conducted a laboratory experiment and assessed Pearson's product-moment correlation coefficients between several process model metrics and a measure called SCORE intended to measure understandability (process model quality) as well as a linear regression between the process model metrics and SCORE. The SCORE measure is computed as the sum of correct answers to just eight closed and one open question on a process model.

  Vanderfeesten *et al.* introduced the cross-connectivity metric (CC) [167]. It was added to the process model metrics and into the data collected in [101]. No significant correlation between CC and SCORE could be found. But CC is part of a better linear regression model between process model metrics and SCORE.

  In [102], Mendling and Strembeck did another experiment examining influencing factors on process model understandability. Besides dependencies between personal and structural (process model metrics) factors, also content related factors (task labels) were analyzed. Here, understandability is measured with six yes/no questions on the process models. Again, a linear regression model was found.

  Because of its simple definition, the content validity and reliability of the SCORE measure are questionable. It is not clear whether all aspects of process model understandability are covered. The small number of asked questions and the non-systematic selection of these questions could cause that only especially easy or difficult process model parts are examined by the questions. Consequently, SCORE is no valid measurement system. But this makes the whole prediction system for process model understandability invalid.

  These points of criticism together with an experimental evaluation of the hypotheses are explained in detail in Chapter 6.

- **Reijers, van der Aalst, Vanderfeesten: Process model granularity heuristic [129, 168]**

  Vanderfeesten *et al.* introduce a heuristic for the proper size of individual activities in process models (process model granularity) inspired by software engineering. Activities can consist of (several) basic operations. The operations of one activity should "belong" together (highly cohesive)—while different activities should be independent from each other (loosely coupled). For that purpose, they introduce a process model cohesion and a process model coupling metric as well as a coupling/cohesion ratio (process model granularity metric). Based on this metric, Vanderfeesten *et al.* have suggested a heuristic for selecting between different process model alternatives. It prefers models with high cohesion and low coupling. They have also postulated the hypotheses that those process models are less error-prone during process instance execution and better maintainable because they are easier to understand.

They only give a motivation why this heuristic could be correct and present an example project where this heuristic was applied in practice. Yet, as a formal verification is missing, it is still no valid prediction system.

An experiment for evaluating the validity of the proposed heuristic is presented in Chapter 7.

Thus, the assessment had the same result as Sánchez González *et al.* already noticed [136, p. 124]: In the published literature, one can find a strong tendency to create and propose new metrics and measures without any validation. In future research, more attention should be paid on the empirical validation of existing proposals instead of defining new ones.

3.7 conclusion

In this chapter, the state of the art in process measurement was presented and the theoretical framework for process measurement used in this thesis was introduced.

For this purpose, an overview of publications on process measurement was given. Many proposed process model metrics are adapted from software metrics and are claimed to measure process model complexity, quality and/or performance. It could be observed that there are no concrete definitions of process model complexity and process model quality in the literature. Often, both terms are even used as synonyms.

Thus, a discussion of these terms followed. It could be shown that there does not exist a single formal definition of complexity. Instead, numerous aspects of complexity were identified and are analyzed in different research communities. Consequently, it is problematic to say that a process model metric measures the complexity of a process model.

The main contribution of this chapter is a theoretical framework for process measurement in which the existing work can be integrated and which can help to identify open research questions leading to new research directions in process measurement.

For this, the more well-established concepts from software measurement were adopted for process measurement: The result was a prediction system measurement approach, which is based on measurement and prediction systems. The measurement approach consists of process model metrics measuring (structural) internal attributes and process model quality measures measuring external process model attributes. Through this, a concrete definition of process model complexity can be avoided. Nevertheless, process model complexity, quality and performance fit into this measurement approach.

Furthermore, the necessity for a proper validation of measurement and prediction systems was emphasized. Reliability and validity were identified as important requirements for metrics and measures. Yet, both constructs have not received the necessary attention in process measurement literature so far.

The Goal Question Metric approach for the selection of process model metrics and process model quality measures was recommended and different purposes of process measurement (understand, control and improve) were presented.

A concluding assessment of the existing process measurement work showed that there is still a lack of deeper comprehension of the behavior of the proposed process model metrics as well as a missing proper validation of prediction systems using the numerous proposed metrics. Also the creation of process model quality measures for measuring external process model attributes which fulfill the reliability and validity requirements is important.

Some of these points are addressed in the remainder of this thesis: In Chapter 4, some important properties concerning the behavior of proposed process model metrics are analyzed. A visualization technique for and a clustering approach based on the process model metric values of process model collections are suggested in Chapter 5. In Chapter 6, hypotheses concerning problems with existing measures for measuring structural process model understandability (an example of a measuring system) are presented, better measures are proposed and both is empirically examined by conducting an experiment. Finally, the recommended process model granularity heuristic by Vanderfeesten *et al.*—as an example of an unvalidated prediction system—is empirically analyzed by an experiment in Chapter 7.

4

ANALYSIS OF PROCESS MODEL METRIC PROPERTIES

4.1 INTRODUCTION

Chapter 3 showed that numerous process model metrics have been proposed in literature in the previous years. Nevertheless, it had to be noticed that only a small minority of them are part of a validated prediction system.

A proper validation would require controlled experiments (see Appendix B), which are very time and cost intensive. This fact—together with the possibility of a negative outcome of the experimental validation—could explain the small number of existing validated prediction systems.

In this chapter, an approach is proposed which shall help to ease this problem by reducing the experimental effort for

- unsuccessful validations or
- validations of useless prediction systems.

The validation attempt for a prediction system is considered unsuccessful if the assumed dependency between the prediction system's process model metric and the process model quality measure cannot be shown.

A prediction system A is considered useless if there already exists another *validated* prediction system B which predicts the same process model quality measure at least as good as A based on a process model metric which is highly correlated with the process model metric belonging to A. Thus, this prediction system would have no additional value compared to the already existing one.

In order to reach the goal, the approach adds an additional analysis step before the prediction system which shall be validated is selected (see Figure 4.1 for a visual location within the measurement approach of Subsection 3.4.2). In this preceding step, the behavior as well as important properties of process model metrics which are part of the potential prediction systems which shall be validated are first analyzed. Through this, unfavorable properties of process model metrics (e. g., insufficient dispersion of metric values or strong correlation with other process model metrics) can be identified before the high effort for an experimental validation of the corresponding prediction system occurs.

The remainder of this chapter is organized as follows: In Section 4.2, the approach for analyzing process model metric properties is introduced. The examined properties are divided into general properties which only depend on a process model metric's definition (Subsection 4.2.1) and those properties which are specific for a selected and examined process model collection (Subsection 4.2.2). Afterwards, the presented approach is applied to a set of process model metrics and a collection of process models (Section 4.3). The chapter closes with a conclusion (Section 4.4).

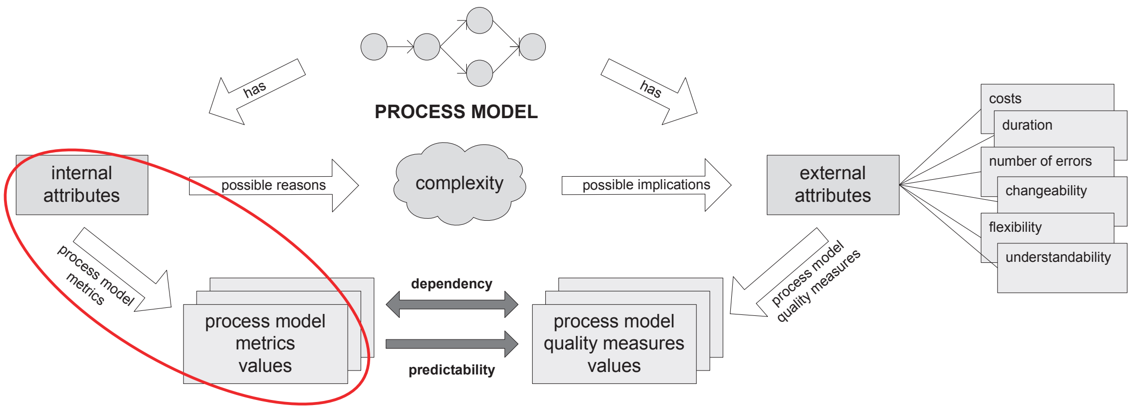


Figure 4.1: Analysis step visually located within the measurement approach of Subsection 3.4.2.

4.2 APPROACH FOR THE ANALYSIS OF PROCESS MODEL METRIC PROPERTIES

In this section, the general approach for analyzing process model metric properties is introduced independently from specific process model metrics and process model collections.

In doing so, one has to distinguish between general and process model collection specific properties. *General properties* (Subsection 4.2.1) hold independently from the considered process models just because of the metric's definition. *Process model collection specific properties* (Subsection 4.2.2) are true for the examined process models—but not implicitly generalizable to other process model collections.

4.2.1 *General Properties*

The approach comprises the analysis of the following general properties.

Theoretical Value Ranges

In the first step, the theoretical value ranges of the process model metrics are examined. It is also checked to which set of numbers (natural numbers $\mathbb{N}$[1], rational numbers $\mathbb{Q}$ or real numbers $\mathbb{R}$) the metric values belong.

If a process model metric has only values in the set of natural numbers and a very small value range, this is equivalent to different categories. In this case, one could check whether these few categories are really useful for characterizing different process models.

Scale Types

Next, the scale type (see Section A.2) of each process model metric is identified.

1 In this thesis, $\mathbb{N}$ is defined as $\mathbb{N} := \{0, 1, 2, 3, \ldots\}$.

This information is important as some statistical operations are only meaningful for specific scale types. An overview of different scale types and their meaningful statistical operations can be found in Table A.2[2].

Behavior for Specific Process Models

The next property which is examined is the behavior of the process model metrics depending on process model structure.

As one cannot test the metrics for every possible process model, one uses at least some generic standard process models which often appear in real life or form an important subprocess of real process models. Amongst these standard process models are sequential and parallel process models.

Depending on the analyzed process model metric, other forms of behavioral examinations are also imaginable.

Correlations Based on Process Model Metric Definitions

As the last general property, it is examined whether general correlations between some of the process model metrics exist just because of their definitions and independent of specific process models.

4.2.2 *Process Model Collection Specific Properties*

The following process model collection specific properties are examined by the approach.

Descriptive Statistics

In this step, the value distributions of the different process model metrics are to be analyzed.

In order to get a compact quantitative overview, the following descriptive statistics are computed for each process model metric:

- minimum, 25% quantile [117, pp. 25–26], median [117, pp. 19–21], 75% quantile [117, pp. 25–26] and maximum (for metrics on ordinal scale) as well as mean [117, pp. 16–17] (for metrics on interval scale) as *measures of location*,
- range (R) [117, p. 22], interquartile range (IQR)[3] [156, p. 55], median absolute deviation (MAD)[4] (for metrics on ordinal scale), standard deviation (sd) [117, p. 22, 35–36] (for metrics on interval scale) as well as coefficient of variation (CV) [117, pp. 33–34] (for metrics on ratio scale) as *measures of dispersion*.

2 In Table A.2, only those statistical operations which are used in the approach for analyzing process model metric properties are listed.

3 difference between 75% and 25% quantiles

4 median of the set of absolute values of the differences between the values and the median of these values

As can be seen in the brackets, some statistics require that the metric has a special scale type. Otherwise, the resulting number would be meaningless. See Section A.2 for details.

If one is interested in more than just some compact quantitative numbers, the value distributions of the process model metrics can be visualized using histograms. Through this, one gets more detailed information.

Correlations

In the next step, possible correlations (see Appendix C) between process model metrics are examined.

For that purpose, Spearman's rank correlation coefficient (Section C.2) can be computed. It is a measure of correlation between two variables assessing how well an *arbitrary monotonic* function could describe the relationship. The two variables must be measured at least on an ordinal scale. Their random distribution is unimportant.

The second statistic, empirical Pearson's product-moment correlation coefficient (see Section C.1), measures the *linear* relationship between two variables. These variables must be measured at least on an interval scale and must be nearly normally distributed.

In order to identify also non-linear dependencies between process model metrics, the scatter plot matrix can be analyzed. Yet, the resulting $n \times n$ matrix for a larger number n of analyzed process model metrics soon becomes unhandily large.

Principal Component Analysis

The last analysis step is the conduction of a PCA.

According to Jolliffe, "[t]he central idea of principal component analysis is to reduce the dimensionality of a data set in which there are a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This reduction is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first *few* retain most of the variation present in *all* of the original variables." [66, p. ix]

So, a PCA searches for a linear transformation (change of basis) so that the new basis vectors—the principal components—are ordered decreasingly according to their proportion of total variance. If the first few components comprise a high proportion of total variance, one can omit the remaining components (dimension reduction) without loosing much of the original information of the data set. The resulting dimensionally reduced data is normally easier to analyze and visualize.

For more technical details, the reader is referred to the textbook by Jolliffe [66].

In the context of this approach, a PCA can be helpful in order to identify the possible redundancy between the different process model metrics.

4.3 EXPERIMENTAL APPLICATION

In this section, the abstract approach presented in the previous Section 4.2 is applied to a set of selected process model metrics and a process model collection.

4.3.1 *Selected Process Model Metrics and Process Model Collection*

In a first step, a set of process model metrics and a process model collection have to be selected.

As every process model metric is only applicable to some specific process modeling languages (see Section 2.3), the following two requirements have to be fulfilled by this selection:

1. In order to compare the properties of the selected process model metrics, they all have to be applicable to the same process modeling language(s).
2. To be able to analyze the selected process model metrics according to the process models of the selected process model collection, these process models have to be modeled using a process modeling language to which the selected process model metrics are applicable.

So, one has to choose a process modeling language for which both numerous process model metrics and a large process model collection exist. EPCs (see Subsection 2.3.1) are such a language.

Selected Process Model Metrics

The choice of EPCs as process modeling language has the big advantage that there are many process model metrics which have either been proposed directly for EPCs or which can easily be adjusted to be applicable to EPCs.

So, 33 EPC process model metrics could be selected for the subsequent analysis. They are listed in Table 4.1.

For each metric, a reference[5] and a short definition are given. For some metrics, it is only referred to the original reference as the definition is too long for the table.

For the metric definitions, the terminology and symbols of Definition 2.10 are used. Furthermore, S denotes the set of split connectors and J the set of join connectors ($C = S \cup F$). Adding one of the labels AND, XOR or OR as an index, both symbols can be used to denote a special subset (e. g., S_{AND} stands for the set of all AND split connectors). In addition, each connector $c \in C$ has an in-degree $d_{in}(c) = |\{(n_1, n_2) \in A | n_2 = c\}|$, an out-degree $d_{out}(c) = |\{(n_1, n_2) \in A | n_1 = c\}|$ and a degree $d(c) = d_{in}(c) + d_{out}(c)$.

5 The references in [99] can be found on the pages 117–130.

Table 4.1: Selected process model metrics for EPCs (Part 1 of 2).

name	symbol	reference	definition
number start events	S_{E_S}	[96, 99]	
number internal events	$S_{E_{Int}}$	[96, 99]	
number end events	S_{E_E}	[96, 99]	
number events	S_E	[99]	$S_E(G) := \|E\| = S_{E_S}(G) + S_{E_{Int}}(G) + S_{E_E}(G)$
number functions	S_F	[96, 99]	$S_F(G) := \|F\|$
number AND splits	$S_{S_{AND}}$	[96, 99]	
number AND joins	$S_{J_{AND}}$	[96, 99]	
number XOR splits	$S_{S_{XOR}}$	[96, 99]	
number XOR joins	$S_{J_{XOR}}$	[96, 99]	
number OR splits	$S_{S_{OR}}$	[96, 99]	
number OR joins	$S_{J_{OR}}$	[96, 99]	
number connectors	S_C	[99]	$S_C(G) := \|C\| = S_{S_{AND}}(G) + S_{J_{AND}}(G) + S_{S_{XOR}}(G) + S_{J_{XOR}}(G) + S_{S_{OR}}(G) + S_{J_{OR}}(G)$
number nodes	S_N	[99, 100]	$S_N(G) := \|N\| = S_E(G) + S_F(G) + S_C(G)$
number arcs	S_A	[96, 99]	$S_A(G) := \|A\|$
diameter	$\mathtt{diam}$	[99]	length of the *longest* path (= number of arcs on this path) from a start event to an end event
density (1)	Δ	[99]	$\Delta(G) := \frac{\|A\|}{\|N\| \cdot (\|N\|-1)}$: number of arcs divided by the maximum number of arcs for the same number of nodes
density (2)	D	[97]	D is a second and more complicated density metric with respect to the EPC syntax constraints. See [97, pp. 3–4].
coefficient of connectivity	CNC	[74, 99]	$CNC(G) := \frac{\|A\|}{\|N\|}$
coefficient of network complexity	CNC_K	[74]	$CNC_K(G) := \frac{\|A\|^2}{\|N\|}$
cyclomatic number	CN	[74]	$CN := \|A\| - \|N\| + 1$: number of linearly independent cycles (arc directions are ignored)

Selected Process Model Collection

The SAP Reference Model [33, 71], which was part of SAP R/3 until version 4.6, was selected as process model collection. This collection of EPC process models has already been used for several experiments found in the literature [96–99].

The collection's process models were validated according to the requirements of Definition 2.10. Out of the 604 non-trivial EPCs of the SAP Reference Model, 89 had to be removed because of invalidity[6].

Finally, 515 EPC process models remained for the experimental application of the approach.

4.3.2 *Results Concerning General Properties*

In this subsection, the results concerning the general properties are presented.

6 no start event, no end event, a function with not exactly one predecessor and one successor node, an event with more than one predecessor or successor node or several graph components

Table 4.1: Selected process model metrics for EPCs (Part 2 of 2).

name	symbol	reference	definition
avg. connector degree	$\overline{d_C}$	[99]	$\overline{d_C}(G) := \frac{1}{\lvert C\rvert}\sum_{c\in C} d(c)$ (see [a])
max. connector degree	$\widehat{d_C}$	[99]	$\widehat{d_C}(G) := \max\{d(c) \mid c \in C\}$ (see [a])
separability	Π	[99, 100]	$\Pi(G) := \frac{\lvert\{n\in N \mid n \text{ is cut-vertex}\}\rvert}{\lvert N\rvert-2}$: A cut-vertex is a node whose deletion separates the process model into multiple components.
sequentiality	Ξ	[99, 100]	$\Xi(G) := \frac{\lvert A\cap((E\cup F)\times(E\cup F))\rvert}{\lvert A\rvert}$: number of arcs between non-connector nodes divided by the number of arcs
depth	Λ	[99]	Depth relates to the maximum nesting of structured blocks in a process model. See [99, pp. 124–125].
mismatch	MM	[99]	$MM(G) := \sum_{l\in\{AND,XOR,OR\}}\left(\left\lvert\sum_{c\in S_l} d_{out}(c) - \sum_{c\in J_l} d_{in}(c)\right\rvert\right)$ (see [b]): sum of mismatches for each connector type
heterogeneity	CH	[99]	$CH(G) := -\sum_{l\in\{AND,XOR,OR\}} p(l)\cdot\log_3 p(l)$ (see [a]): entropy over the different connector types
cyclicity	CYC	[99, 100]	$CYC_N(G) := \frac{\lvert N_C\rvert}{\lvert N\rvert}$: number of nodes N_C on a cycle (arc directions are *not* ignored) divided by the number of nodes
token splits	TS	[99]	$TS(G) := \sum_{c\in S_{AND}\cup S_{OR}}(d_{out}(c)-1)$: number of newly introduced tokens by split connectors
control flow complexity	CFC	[21, 96, 99]	$CFC(G) := \sum_{c\in S_{AND}} 1 + \sum_{c\in S_{XOR}} d_{out}(c) + \sum_{c\in S_{OR}}\left(2^{d_{out}(c)}-1\right)$: sum over all split connectors weighted by their number of possible states after the split
join complexity	JC	[96]	$JC(G) := \sum_{c\in J_{AND}} 1 + \sum_{c\in J_{XOR}} d_{in}(c) + \sum_{c\in J_{OR}}\left(2^{d_{in}(c)}-1\right)$: sum over all join connectors weighted by their number of possible states before the join
weighted coupling	CP	[166]	CP measures the average coupling between all pairs of connected events and functions—weighted according to the type of connection. See [166, p. 42].
cross-connectivity	CC	[167]	CC measures the average strength of connection between all pairs of nodes. See [167, pp. 483–484].

a Metric value is 0 for $\lvert C\rvert = 0$ (source: personal communication with Jan Mendling).
b The original definition printed in [99, p. 125] is faulty (source: personal communication with Jan Mendling).

Theoretical Value Ranges

The theoretical value ranges of the selected process model metrics and the set of numbers to which their values belong are listed in the second column of Table 4.2.

Proof. The theoretical value ranges and the sets of numbers have to be proved. As the determination of the set of numbers to which the process model metrics' values belong is trivial when one looks at their mathematical definitions, no detailed proof for this aspect is given here. Proofs are only presented for the value ranges. In doing so, some results presented later in this subsection (paragraph *Behavior for Specific Process Models* and *Correlations Based on Process Model Metric Definitions*) are already used here.

Regarding "size metrics": All "size metrics" (metric S_{E_S} till S_A in Table 4.1) measure the number of objects of a specific type in an EPC. So, the values cannot be negative. There is no maximal limit for the metrics. Only some metrics have a

Table 4.2: Theoretic value ranges and scale types of the selected process model metrics.

metric	theoretic value range	scale type
S_{E_S}	$[1, \infty) \cap \mathbb{N}$	ratio
$S_{E_{Int}}$	$[0, \infty) \cap \mathbb{N}$	ratio
S_{E_E}	$[1, \infty) \cap \mathbb{N}$	ratio
S_E	$[2, \infty) \cap \mathbb{N}$	ratio
S_F	$[1, \infty) \cap \mathbb{N}$	ratio
$S_{S_{AND}}$	$[0, \infty) \cap \mathbb{N}$	ratio
$S_{J_{AND}}$	$[0, \infty) \cap \mathbb{N}$	ratio
$S_{S_{XOR}}$	$[0, \infty) \cap \mathbb{N}$	ratio
$S_{J_{XOR}}$	$[0, \infty) \cap \mathbb{N}$	ratio
$S_{S_{OR}}$	$[0, \infty) \cap \mathbb{N}$	ratio
$S_{J_{OR}}$	$[0, \infty) \cap \mathbb{N}$	ratio
S_C	$[0, \infty) \cap \mathbb{N}$	ratio
S_N	$[3, \infty) \cap \mathbb{N}$	ratio
S_A	$[2, \infty) \cap \mathbb{N}$	ratio

metric	theoretic value range	scale type
$diam$	$[2, \infty) \cap \mathbb{N}$	ratio
Δ	$(0, \frac{1}{3}] \cap \mathbb{Q}$	ratio
D	$[0, 1) \cap \mathbb{Q}$	ordinal
CNC	$[\frac{2}{3}, 2) \cap \mathbb{Q}$	ratio
CNC_K	$[1\frac{1}{3}, \infty) \cap \mathbb{Q}$	ratio
CN	$[0, \infty) \cap \mathbb{N}$	ratio
$\overline{d_C}$	$[0, \infty) \cap \mathbb{Q}$	ratio
$\widehat{d_C}$	$[0, \infty) \cap \mathbb{N}$	ratio
Π	$(0, 1] \cap \mathbb{Q}$	ordinal
Ξ	$[0, 1] \cap \mathbb{Q}$	ratio
Λ	$[0, \infty) \cap \mathbb{N}$	ratio
MM	$[0, \infty) \cap \mathbb{N}$	ratio
CH	$[0, 1] \cap \mathbb{R}$	ordinal
CYC	$[0, 1) \cap \mathbb{Q}$	ratio
TS	$[0, \infty) \cap \mathbb{N}$	ratio
CFC	$[0, \infty) \cap \mathbb{N}$	ratio
JC	$[0, \infty) \cap \mathbb{N}$	ratio
CP	$(0, \frac{21}{52} \approx 0.404] \cap \mathbb{Q}$	ordinal
CC	$(0, 1) \cap \mathbb{Q}$	ordinal

higher minimal value than 0 because of some requirements of the definition of EPCs (see Definition 2.10): An EPC must have at least one start event ($S_{E_S}(G) \geqslant 1$), one end event ($S_{E_E}(G) \geqslant 1$) and one function ($S_F(G) \geqslant 1$). Consequently, $S_E(G) \geqslant 2$ and $S_N(G) \geqslant 3$ hold. According to (4.17), $S_A(G) \geqslant S_N(G) - 1$ and so $S_A(G) \geqslant 2$ hold ($S_A(G) = 2$ for sequential process models of type SEQ-1).

Regarding metric $diam$:

- lower bound: The smallest process model SEQ-1 has the minimal longest path from a start to an end event with $diam$ value 2.
- upper bound: According to Table 4.3, the $diam$ value becomes arbitrarily large for SEQ-n process models. So, there is no maximal value.

Regarding metric Δ:

- lower bound: Metric Δ measures a ratio between existing arcs and possible arcs. As an EPC has at least two arcs, this ratio cannot be negative and cannot be 0. According to Table 4.3, the Δ values for the SEQ-n converge to 0. So, 0 (excluded) is the minimal possible value.

- upper bound: The Δ value of a SEQ-1 (with three nodes) is $\frac{1}{3}$. An EPC with four nodes is impossible. According to (4.25), $\Delta(G) \leqslant \frac{2-\frac{6}{S_N(G)}}{S_N(G)-1} \leqslant \frac{2-\frac{6}{5}}{5-1} = 0.2$ for $S_N(G) \geqslant 5$. So, $\frac{1}{3}$ is the maximal value.

Regarding metric D**:**

- lower bound: According to its definition, metric D has value 0 for EPCs G with $S_C(G) \leqslant 1$ (this includes the sequential process models SEQ-n). For the remaining EPCs, the original definition in [97] can be transformed into

$$D(G) := \frac{S_A(G) - S_N(G) + 1}{c_{max} + S_N(G) - 2S_C(G) + 1} \quad . \tag{4.1}$$

The nominator equals metric CN with $CN(G) \geqslant 0$ (see equation (4.12)). For the denominator (shortly written here as $D'(G)$),

$$\begin{aligned}
D'(G) &= \begin{cases} \left(\frac{S_C(G)}{2}+1\right)^2 + S_N(G) - 2S_C(G) + 1 & \text{for } S_N(C) \text{ even} \\ \left(\frac{S_C(G)-1}{2}+1\right)^2 + \frac{S_C(G)-1}{2} + 1 & \\ \qquad +S_N(G) - 2S_C(G) + 1 & \text{for } S_N(C) \text{ odd} \end{cases} \\
&= \begin{cases} \frac{1}{4}(S_C(G))^2 - S_C(G) + S_N(G) + 2 & \text{for } S_N(C) \text{ even} \\ \frac{1}{4}(S_C(G))^2 - S_C(G) + S_N(G) + 1.75 & \text{for } S_N(C) \text{ odd} \end{cases} \qquad (4.2) \\
&\overset{(4.20)}{\geqslant} \begin{cases} \frac{1}{4}(S_C(G))^2 - (\frac{2}{3}S_N(G) - 2) + S_N(G) + 2 & \text{for } S_N(C) \text{ even} \\ \frac{1}{4}(S_C(G))^2 - (\frac{2}{3}S_N(G) - 2) + S_N(G) + 1.75 & \text{for } S_N(C) \text{ odd} \end{cases} \\
&= \begin{cases} \frac{1}{4}(S_C(G))^2 + \frac{1}{3}S_N(G) + 4 & \text{for } S_N(C) \text{ even} \\ \frac{1}{4}(S_C(G))^2 + \frac{1}{3}S_N(G) + 3.75 & \text{for } S_N(C) \text{ odd} \end{cases} \\
&> 0
\end{aligned}$$

holds. So, the D values cannot be negative and value 0 is the minimal one.

- upper bound: For EPCs G with $S_C(G) \leqslant 1$, the (maximal) D value is 0. The following considerations are true for the remaining EPCs ($S_C(G) \geqslant 2$). Replacing the denominator of (4.1) by (4.2) results in

$$
\begin{aligned}
D(G) &= \begin{cases} \frac{S_A(G)-S_N(G)+1}{\frac{1}{4}(S_C(G))^2-S_C(G)+S_N(G)+2} & \text{for } S_N(C) \text{ even} \\ \frac{S_A(G)-S_N(G)+1}{\frac{1}{4}(S_C(G))^2-S_C(G)+S_N(G)+1.75} & \text{for } S_N(C) \text{ odd} \end{cases} \\
&\overset{(4.18)}{\leqslant} \begin{cases} \frac{(2S_N(G)-6)-S_N(G)+1}{\frac{1}{4}(S_C(G))^2-S_C(G)+S_N(G)+2} & \text{for } S_N(C) \text{ even} \\ \frac{(2S_N(G)-6)-S_N(G)+1}{\frac{1}{4}(S_C(G))^2-S_C(G)+S_N(G)+1.75} & \text{for } S_N(C) \text{ odd} \end{cases} \\
&= \begin{cases} \frac{S_N(G)-5}{\underbrace{\frac{1}{4}(S_C(G))^2 - S_C(G)}_{\geqslant -1 \text{ for } S_C(G)\geqslant 2}+S_N(G)+2} & \text{for } S_N(C) \text{ even} \\ \frac{S_N(G)-5}{\underbrace{\frac{1}{4}(S_C(G))^2 - S_C(G)}_{\geqslant -0.75 \text{ for } S_C(G)\geqslant 3}+S_N(G)+1.75} & \text{for } S_N(C) \text{ odd} \end{cases} \\
&\leqslant \frac{S_N(G)-5}{S_N(G)+1} = 1 - \frac{6}{S_N(G)+1} \overset{S_N(G)\to\infty}{\longrightarrow} 1 \quad .
\end{aligned}
$$

The D metric values for the process models AND-n converge to 1 (see Table 4.3). So, 1 (excluded) is the maximal value.

Regarding metric CNC**:**

- lower bound:

$$
CNC(G) = \frac{S_A(G)}{S_N(G)} \overset{(4.17)}{\geqslant} \frac{S_N(G)-1}{S_N(G)} = 1 - \frac{1}{\underbrace{S_N(G)}_{S_N(G)\geqslant 3}} \geqslant 1 - \frac{1}{3} = \frac{2}{3}
$$

  SEQ-1 has this CNC value (see Table 4.3).

- upper bound: The CNC value of the smallest process model SEQ-1 (three nodes) is $\frac{2}{3}$. An EPC with four nodes is impossible. For all remaining nodes,

$$
CNC(G) = \frac{S_A(G)}{S_N(G)} \overset{(4.18)}{\leqslant} \frac{2S_N(G)-6}{S_N(G)} = 2 - \frac{6}{S_N(G)} < 2
$$

  holds. The CNC values of the process models AND-n converge to 2 (see Table 4.3). So, 2 (excluded) is the maximal value.

Regarding metric CNC_K**:**

- lower bound: According to the proof of (4.22), $CNC_K(G) \geqslant S_N(G) - 2 + \frac{1}{S_N(G)}$ holds. The minimal value of the right side of this inequality is reached for $S_N(G) = 3$. In that case, the inequality can be further transformed into $CNC_K(G) \geqslant 3 - 2 + \frac{1}{3} = 1\frac{1}{3}$. SEQ-1 has this minimal CNC_K value.

- upper bound: According to Table 4.3, the CNC_K value becomes arbitrarily large for SEQ-n process models. So, there is no maximal value.

Regarding metric CN:

- lower bound: According to Theorem 4.1, metric CN counts the number of linearly independent cycles (see Definition 4.1) (arc directions are ignored) of an EPC. So, this value cannot be negative. For sequential process models SEQ-n, the value is 0 (see Table 4.3).
- upper bound: Following Table 4.3, the CN value becomes arbitrarily large for AND-n process models. So, there is no maximal value.

Regarding metric $\overline{d_C}$ and $\widehat{d_C}$:

- lower bound: As a connector node cannot have a negative degree, also $\overline{d_C}$ and $\widehat{d_C}$ cannot be negative. For the sequential process models SEQ-n, both metrics have a value of 0.
- upper bound: Following Table 4.3, the values of both metrics become arbitrarily large for AND-n process models. So, there is no maximal value.

Regarding metric Π:

- lower bound: The ratio cannot become negative. In every EPC, a start event must be connected with a non-event node as its only neighbor. Consequently, this node is a cut-vertex and the ratio cannot become 0. According to Table 4.3, the Π values of the AND-n process models converge to 0. So, 0 (excluded) is the minimal value.
- upper bound: As an EPC has at least one start and one end event and both cannot be a cut-vertex, the ratio cannot become larger than 1. The sequential process models SEQ-n have a value of exactly 1 (see Table 4.3).

Regarding metric Ξ:

- lower bound: As metric Ξ measures the ratio between the arcs connecting non-connector nodes and all arcs, the value cannot become negative. The AND-n process models have a value of 0 (see Table 4.3).
- upper bound: The ratio cannot become larger than 1. The sequential process models SEQ-n have a value of 1.

Regarding metric Λ and MM:

- lower bound: Because of their definitions, both metrics cannot have negative values. The sequential process models SEQ-n have value 0 for both metrics.
- upper bound: As there are no maximal nesting depth and no maximal sum of mismatches, the values of both metrics can become arbitrarily large.

Regarding metric CH**:**

- lower bound: The single summands of the CH definition are $p(l) \cdot \log_3 p(l)$ with $p(l) \in [0, 1]$. For this value range of $p(l)$, $\log_3 p(l) \leqslant 0$ and consequently $p(l) \cdot \log_3 p(l) \leqslant 0$ hold. As the final sum is multiplied by -1, the CH values cannot be negative. For process models with no connectors, the metric CH is defined as 0.

- upper bound: For the three probabilities $p_1 := p(\text{AND})$, $p_2 := p(\text{XOR})$ and $p_3 := p(\text{OR})$, the equation

  $$p_3 = 1 - p_1 - p_2 \tag{4.3}$$

  holds. Consequently, metric CH can be interpreted as a function

  $$f(p_1, p_2) := -\left[p_1 \log_3 p_1 + p_2 \log_3 p_2 + (1 - p_1 - p_2) \log_3 (1 - p_1 - p_2)\right] . \tag{4.4}$$

  In order to find extrema, one gets the two partial derivatives

  $$\frac{\partial f}{\partial p_1} = \frac{\ln(1 - p_1 - p_2) - \ln p_1}{\ln 3} \tag{4.5}$$

  $$\frac{\partial f}{\partial p_2} = \frac{\ln(1 - p_1 - p_2) - \ln p_2}{\ln 3} \quad . \tag{4.6}$$

  For extrema,

  $$\frac{\partial f}{\partial p_1} = \frac{\partial f}{\partial p_2} \stackrel{!}{=} 0 \tag{4.7}$$

  must hold. This is only true for $p_1 = p_2 = p_3 = \frac{1}{3}$. Here, the function f—and consequently metric CH—has its maximal value 1.

Regarding metric CYC**:**

- lower bound: Metric CYC measures the ratio between nodes on a cycle (arc directions are *not* ignored) and all nodes. Consequently, the value cannot be negative. SEQ-n process models have a value of 0 (see Table 4.3).

- upper bound: Imagine an EPC G with one start event followed by an XOR join. The XOR join is following alternately by n functions and n events until the XOR join is reached again (forming a cycle!). Between one of the functions and events on the cycle, there is an XOR split inserted. The second outgoing arc of this split is connected with an end event. So, there are $2n + 2$ of the total $2n + 4$ nodes on the cycle. Consequently,

  $$\text{CYC}(G) = \frac{2n+2}{2n+4} = \frac{n+1}{n+2} = 1 - \frac{1}{n+2} \stackrel{n \to \infty}{\longrightarrow} 1$$

  holds. As at least one start and one end node cannot lie on a cycle, value 1 is not reachable. The ratio cannot become larger than 1. So, 1 (excluded) is the maximal value.

Regarding metric TS**,** CFC **and** JC**:**

- lower bound: Because of their definitions, negative metric values are impossible. The sequential process models SEQ-n have value 0 for all three metrics (see Table 4.3).
- upper bound: According to Table 4.3, the values of all three metrics become arbitrarily large for OR-n process models. So, there is no maximal value.

Regarding metric CP**:**

- lower bound: As the arc weights cannot be negative, the CP metric values can also not become negative. Each EPC has at least two arcs and their weights are larger than 0. The metric values of the sequential process models SEQ-n converge to 0 (see Table 4.3), So, 0 (excluded) is the minimal value.
- upper bound: The sequential process model SEQ-1 has CP value $\frac{1}{3}$. This is the smallest possible EPC. There is no EPC with four nodes. For $S_N(G) \geqslant 5$, the following considerations are true. An EPC G has at most $2S_N(G) - 6 - S_C(G)$ pairs of non-connector nodes (maximal number of arcs of G minus number of connectors as there is one arc per connector more than pairs of non-connector nodes next to the connector) which each can have a maximal weight of 1. Consequently,

$$
\begin{aligned}
CP(G) &\leqslant \frac{2S_N(G) - 6 - S_C(G)}{(S_N(G) - S_C(G))(S_N(G) - S_C(G) - 1)} \\
&= \frac{(S_N(G) - S_C(G) - 1) + (S_N(G) - 5)}{(S_N(G) - S_C(G))(S_N(G) - S_C(G) - 1)} = \frac{1 + \frac{S_N(G)-5}{S_N(G)-S_C(G)-1}}{S_N(G) - S_C(G)} \\
&\overset{(4.20)}{\leqslant} \frac{1 + \frac{S_N(G)-5}{S_N(G)-(\frac{2}{3}S_N(G)-2)-1}}{S_N(G) - S_C(G)} = \frac{\frac{4S_N(G)-12}{S_N(G)+3}}{S_N(G) - S_C(G)} \\
&\overset{(4.20)}{\leqslant} \frac{\frac{4S_N(G)-12}{S_N(G)+3}}{S_N(G) - (\frac{2}{3}S_N(G) - 2)} = \frac{\frac{4S_N(G)-12}{S_N(G)+3}}{\frac{S_N(G)+6}{3}} \\
&= \frac{12S_N(G) - 36}{(S_N(G))^2 + 9S_N(G) + 18}
\end{aligned}
$$

holds. This fraction has its maximal value $\frac{21}{52} \approx 0.404$ for $S_N(G) = 10$.[7]

Regarding metric CC**:**

- lower bound: The metric CC is defined as average strength of connection between all pairs of nodes. The nominator (average strength of connection) is always larger than 0, the denominator (number of pairs of nodes) cannot

7 Even after an intensive search, no EPC was found with this CP value. SEQ-1 and AND-2 with both CP value $\frac{1}{3}$ are the EPCs with the largest found CP value. Nevertheless, an EPC with a CP value between $\frac{1}{3}$ and $\frac{21}{52}$ could exist.

be negative. Consequently, the CC values are larger than 0. For the process models AND-n, the metric values converge to 0 (see Table 4.3). So, 0 (excluded) is the minimal value.

- upper bound: The strength of a connection between two nodes cannot be larger than 1. So, metric CC cannot have values larger than 1. As no path from an end to a start event exists, the average strength between all node pairs cannot be exactly 1. Imagine an EPC G with one start event followed by an AND join. The AND join is following alternately by n functions and n events until the AND join is reached again (forming a cycle!). Between one of the functions and events on the cycle, there is an AND split inserted. The second outgoing arc of this split is connected with an end event. So, $S_N(G) = 2n + 4$ holds. There is a path from the start event to every other node. The value of each of this $2n + 3$ connections is 1. Furthermore, there is a path from each of the $2n + 2$ nodes on the cycle to every other node on the cycle and to the end event. The value of these $(2n + 2)(2n + 2)$ connections is also 1. For G's CC metric value

$$CC(G) = \frac{(2n+3) + (2n+2)(2n+2)}{(2n+4)(2n+3)} = 1 - \frac{4n+5}{4n^2+14n+12} \overset{n\to\infty}{\Longrightarrow} 1$$

holds. So, 1 (excluded) is the maximal CC value.

■

Scale Types

The scale types of the selected process model metrics are listed in the third column of Table 4.2.

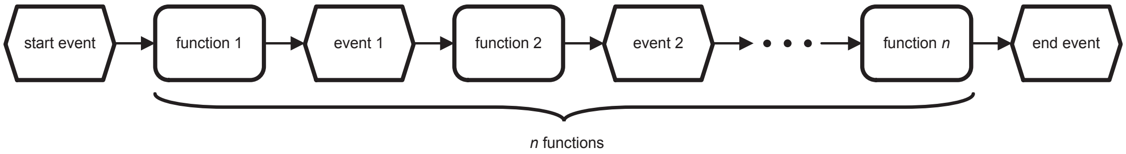
Most metrics are ratio scales. But there are also some exceptions:

- The definition of *separability* (Π) (subtraction of 2 in the denominator) only considers the case of exactly one start and one end event. But as there can be several start and end events, value differences and ratios are meaningless. So, the metric is an ordinal scale.
- For *heterogeneity* (CH), value differences and ratios are meaningless because of taking the logarithm. So, the metric is an ordinal scale.
- For *density* (D), *weighted coupling* (CP) and *cross-connectivity* (CC), things are a little unclear. To err on the side of conservatism, these metrics are listed to be ordinal scales.

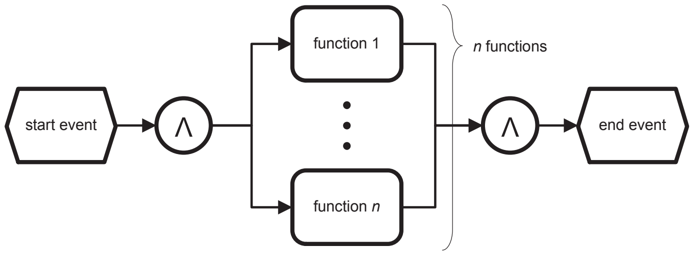
Behavior for Specific Process Models

As explained in Subsection 4.2.1, sequential and parallel process models are used for examining the selected process model metrics' behavior for them.

For the subsequent analysis, four types of generic EPC process models are defined:



(a) Sequential process model SEQ-n ($n \geqslant 1$).



(b) Parallel process model AND-n ($n \geqslant 2$).

Figure 4.2: Sequential and parallel process models.

- SEQ-n ($n \geqslant 1$): sequential EPC with n functions (see Figure 4.2a),
- AND-n ($n \geqslant 2$): EPC with n parallel (AND connectors) functions (see Figure 4.2b),
- XOR-n ($n \geqslant 2$): EPC with n alternative (XOR connectors) functions (AND-n with XOR instead of AND connectors) and
- OR-n ($n \geqslant 2$): EPC with n alternative (OR connectors) functions (AND-n with OR instead of AND connectors).

The metric values for these four process model types are listed in Table 4.3. As one can see, some process model metric values are constant, while others increase infinitely by increasing n or converge to a limit value.

Besides the behavior of the selected process model metrics for sequential and parallel process models, the special behavior of the metric *heterogeneity* (CH) is examined here.

For this metric, only the relative (not absolute) number of the three connector types $p(\text{AND}), p(\text{XOR}), p(\text{OR}) \in [0, 1]$ is important. Consequently, all process models can be represented as a point $\vec{p} := (p(\text{AND}), p(\text{XOR}), p(\text{OR}))^T$ in the 3-dimensional space forming a triangular area on a 2-dimensional hyperplane (as $p(\text{AND}) + p(\text{XOR}) + p(\text{OR}) = 1$,[8] see Figure 4.3).

For the metric value, it is unimportant how the relative number of connector types is actually mapped to the single connector types (e. g., a process model with one AND and one XOR connector has the same CH metric value as a process model with one OR and one XOR connector). This symmetry is represented by the three dashed lines (axes of symmetry) in Figure 4.3.

The metric has its minimum value 0 for process models with only one connector type (ratio $0 : 0 : 1$, minimal connector type heterogeneity) or no connectors at all

8 The sole exception where this equation does not hold are process models with no connectors at all.

Table 4.3: Behavior of the selected process model metrics for sequential and parallel process models (Part 1 of 2).

metric	SEQ-n	AND-n	XOR-n	OR-n
S_{E_S}	1	1		
$S_{E_{Int}}$	$n-1$	0		
S_{E_E}	1	1		
S_E	$n+1$	2		
S_F	n	n		
$S_{S_{AND}}$	0	1	0	0
$S_{J_{AND}}$	0	1	0	0
$S_{S_{XOR}}$	0	0	1	0
$S_{J_{XOR}}$	0	0	1	0
$S_{S_{OR}}$	0	0	0	1
$S_{J_{OR}}$	0	0	0	1
S_C	0	2		
S_N	$2n+1$	$n+4$		
S_A	$2n$	$2n+2$		

as well as its maximum 1 for process models with the same number of all three connector types (ratio $\frac{1}{3} : \frac{1}{3} : \frac{1}{3}$, maximal connector type heterogeneity).

Correlations Based on Process Model Metric Definitions

The last general property, which is examined here, are possible correlations between process model metrics because of their definitions.

First, the statement in Table 4.1 that process model metric CN counts the number of linearly independent cycles (arc directions are ignored) is proved.

If the edges in a graph G with e edges are numbered $1, 2, \ldots, e$, then a cycle is defined by a vector

$$\vec{\mu} := (\mu_1, \mu_2, \ldots, \mu_e)^T \tag{4.8}$$

with

$$\mu_i := \begin{cases} 0 & \text{if corresponding edge is not part of cycle} \\ 1 & \text{if cycle traverses corresp. edge in edge's direction} \\ -1 & \text{if cycle traverses corresp. edge against its direction} \end{cases} . \tag{4.9}$$

If the graph is undirected, a “virtual direction” has to be assigned to each edge in order to use the above notation. [11, p. 12]

Table 4.3: Behavior of the selected process model metrics for sequential and parallel process models (Part 2 of 2). In the table, $f(n) := \frac{1}{2^{n+1}-1} + \frac{2^{n+1}-2}{(2^{n+1}-1)(n+1)}$ is used.

<table>
<tr><th>metric</th><th>SEQ-n</th><th>AND-n</th><th>XOR-n</th><th>OR-n</th></tr>
<tr><td>diam</td><td>$2n$</td><td colspan="3">4</td></tr>
<tr><td>Δ</td><td>$\frac{1}{2n+1} \xrightarrow{n\to\infty} 0$</td><td colspan="3">$\frac{2(n+1)}{(n+3)(n+4)} \xrightarrow{n\to\infty} 0$</td></tr>
<tr><td>D</td><td>0</td><td colspan="3">$\frac{n-1}{n+5} \xrightarrow{n\to\infty} 1$</td></tr>
<tr><td>CNC</td><td>$\frac{2n}{2n+1} \xrightarrow{n\to\infty} 1$</td><td colspan="3">$\frac{2(n+1)}{n+4} \xrightarrow{n\to\infty} 2$</td></tr>
<tr><td>CNC_K</td><td>$\frac{4n^2}{2n+1} \xrightarrow{n\to\infty} 2n-1$</td><td colspan="3">$\frac{4(n+1)^2}{n+4} \xrightarrow{n\to\infty} 4n-8$</td></tr>
<tr><td>CN</td><td>0</td><td colspan="3">$n-1$</td></tr>
<tr><td>$\overline{d_C}$</td><td>0</td><td colspan="3">$n+1$</td></tr>
<tr><td>$\widehat{d_C}$</td><td>0</td><td colspan="3">$n+1$</td></tr>
<tr><td>Π</td><td>1</td><td colspan="3">$\frac{2}{n+2} \xrightarrow{n\to\infty} 0$</td></tr>
<tr><td>Ξ</td><td>1</td><td colspan="3">0</td></tr>
<tr><td>Λ</td><td>0</td><td colspan="3">1</td></tr>
<tr><td>MM</td><td>0</td><td colspan="3">0</td></tr>
<tr><td>CH</td><td>0</td><td colspan="3">0</td></tr>
<tr><td>CYC</td><td>0</td><td colspan="3">0</td></tr>
<tr><td>TS</td><td>0</td><td>$n-1$</td><td>0</td><td>$n-1$</td></tr>
<tr><td>CFC</td><td>0</td><td>1</td><td>n</td><td>2^n-1</td></tr>
<tr><td>JC</td><td>0</td><td>1</td><td>n</td><td>2^n-1</td></tr>
<tr><td>CP</td><td>$\frac{1}{2n+1} \xrightarrow{n\to\infty} 0$</td><td>$\frac{2n}{(n+1)(n+2)} \xrightarrow{n\to\infty} 0$</td><td>$\frac{2}{(n+1)(n+2)} \xrightarrow{n\to\infty} 0$</td><td>$\frac{2(2^n+n-2)}{(2^n-1)(n+1)(n+2)} \xrightarrow{n\to\infty} 0$</td></tr>
<tr><td>CC</td><td>0.5</td><td>$\frac{4n+6}{(n+3)(n+4)} \xrightarrow{n\to\infty} 0$</td><td>$\frac{[f(n)]^4+2[f(n)]^3+(2n+1)[f(n)]^2+(2n+2)[f(n)]}{(n+3)(n+4)} \xrightarrow{n\to\infty} 0$</td><td>$\frac{2(n+1)^4+(2n+1)(n+1)^2+2(n+1)+1}{(n+3)(n+4)(n+1)^4} \xrightarrow{n\to\infty} 0$</td></tr>
</table>

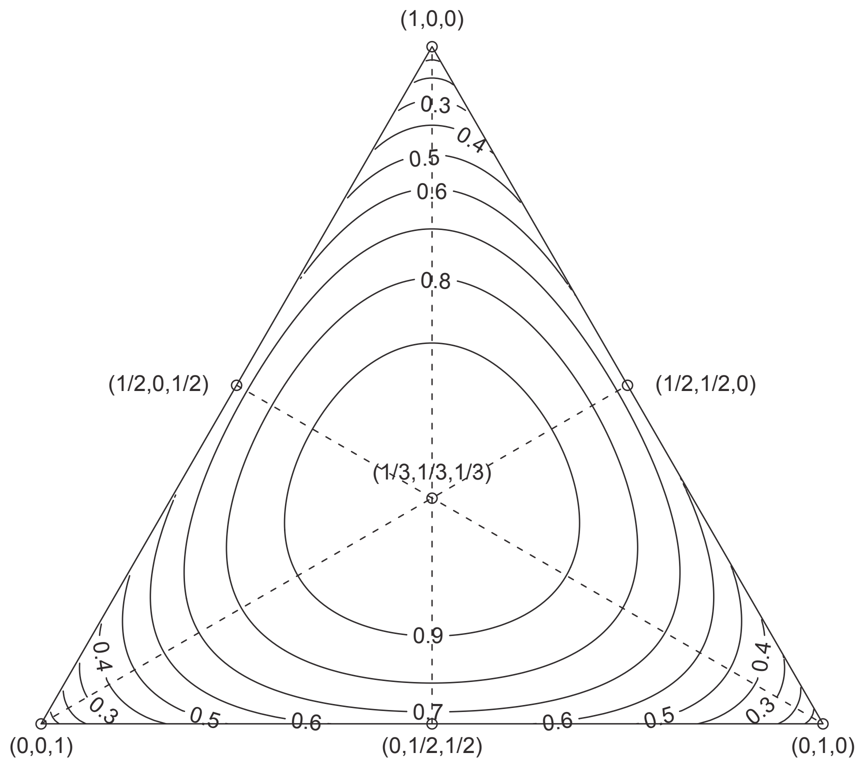


Figure 4.3: Values of the process model metric *heterogeneity* (CH) depending on the probabilities of the three connector types AND, XOR or OR (2-dimensional hyperplane in $\mathbb{R}^3$ transformed into $\mathbb{R}^2$).

It has to be noticed that a cycle can traverse edges against its direction according to this notation. For process model metric CN, this definition of a cycle is used (the arc directions are ignored). In contrast, all arcs have to be traversed in their directions in order to form a cycle for process model metric CYC (the arc directions are *not* ignored).

Using the above notation, linearly independent cycles can be defined [11, p. 15].

Definition 4.1 (Linearly independent cycles) *The cycles $\vec{\mu}_1, \vec{\mu}_2, \ldots, \vec{\mu}_k$ are said to be linearly independent if the equation*

$$a_1\vec{\mu}_1 + a_2\vec{\mu}_2 + \ldots + a_k\vec{\mu}_k = 0 \tag{4.10}$$

is only true for $a_1 = a_2 = \ldots = a_k = 0$.

The maximum number of linearly independent cycles of a graph (its *cyclomatic number*) is a property of the graph. It can be computed using the following Lemma 4.1 [152, p. 23].

Lemma 4.1 *Let G be a graph, n the number of its nodes, e the number of its edges and p the number of its components. Then,*

$$CN(G) = e - n + p \tag{4.11}$$

is the number of linearly independent cycles of the graph.

Proof. The proof works by induction over the number e of edges.

Basis Let G be a graph with n nodes and no edges ($e = 0$). Consequently, each single node is a component ($p = n$) and the graph has no cycle ($CN(G) = 0$). So, $CN(G) = e - n + p = 0 - n + n = 0$ holds.

Induction step When an additional edge is added to the graph, two possible cases can occur:

1. The new edge connects two components. As a consequence, the number of components is reduced by 1 ($p' = p - 1$) and the number of linearly independent cycles stays unchanged ($CN'(G) = CN(G)$). $CN'(G) = e' - n' + p' = (e + 1) - n + (p - 1) = e - n + p = CN(G)$ still holds.

2. The new edge connects two nodes of one component. As a consequence, the number of components stays unchanged ($p' = p$). As there was already a path between the two nodes (they are in the same graph component), adding the new edge between them results in a new cycle. As this new cycle contains the new edge, which cannot be part of any already existing cycle, the new cycle is linearly independent of all other cycles. Consequently, the number of linearly independent cycles is increased by 1 ($CN'(G) = CN(G) + 1$). Also here, $CN'(G) = e' - n' + p' = (e + 1) - n + p = CN(G) + 1$ holds.

■

With the help of Lemma 4.1, Theorem 4.1 can now be proved.

Theorem 4.1 *The process model metric* CN *defined as* $CN(G) := |A| - |N| + 1$ *for any EPC* G *counts the number of linearly independent cycles (the arc directions are ignored).*

Proof. According to Definition 2.10, G is a connected graph. So, it has only one component ($p = 1$). Consequently, equation (4.11) becomes $CN(G) = |A| - |N| + 1$ for the EPC. ■

Because of the definition of metric CN, there is a strong mathematical connection between the involved metrics S_A, S_N and CN.

Theorem 4.2 *The points* $\vec{p}(G_i) := \begin{pmatrix} S_A(G_i) \\ S_N(G_i) \\ CN(G_i) \end{pmatrix}$ *consisting of the corresponding process model metric values of a set of EPCs* G_i *lie on a 2-dimensional hyperplane of the* $\mathbb{R}^3$.

Proof. According to definition, $CN(G_i) = S_A(G_i) - S_N(G_i) + 1$ holds. Consequently, the points

$$\begin{aligned}\vec{p}(G_i) &= \begin{pmatrix} S_A(G_i) \\ S_N(G_i) \\ CN(G_i) \end{pmatrix} = \begin{pmatrix} S_A(G_i) \\ S_N(G_i) \\ S_A(G_i) - S_N(G_i) + 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + S_A(G_i)\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + S_N(G_i)\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\end{aligned}$$

form a 2-dimensional hyperplane of the $\mathbb{R}^3$ which is spanned by two linearly independent vectors. ∎

In the remainder of this paragraph, the correlations between different process model metrics are examined. To start with, no strong correlations in means of Spearman's rank correlation coefficient (see Section C.2) or Pearson's product-moment correlation coefficient (see Section C.1) just based on the metrics' definitions could be proved. Yet, some inequalities between pairs of process model metrics could be found which constrain the metric values in such a way that the existence of strong (linear) correlations is highly possible.

Theorem 4.3 *For every EPC* G*, the inequalities*

$$0 \leqslant CN(G) \tag{4.12}$$

and

$$CN(G) \leqslant S_N(G) - 5 \quad , S_N(G) \geqslant 5 \tag{4.13}$$

hold.

Proof. The two inequalities are proved in separate steps.

Regarding (4.12): According to Theorem 4.1, $CN(G)$ is the number of linearly independent cycles—when the arc directions are ignored. This number cannot be negative. For the sequential process models SEQ-n, this number is 0 (see Table 4.3). So, $CN(G) \geqslant 0$ holds.

Regarding (4.13): In order to add an additional cycle (arc directions are ignored) to an EPC, there have to be two connectors (one split and one join connector) in the EPC. Between these two connectors, an additional path to the "sequential one" must exist. The smallest example for this construct is an AND-2 (see Figure 4.2b). At least one event or function node lies on each of these paths between two connectors. So, the number of cycles of an EPC is limited by the number of nodes (at least one node per cycle). Yet, not all nodes lie on the mentioned paths between two connectors. At least five nodes (one start and one end event, one split and one join connector as well as one function) are used at other places (think of an AND-2 as in Figure 4.2b). Consequently, $CN(G) \leqslant S_N(G) - 5$ holds for $S_N(G) \geqslant 5$. ∎

Theorem 4.4 *For every EPC* G*, the inequalities*

$$0 \leqslant CN(G) \tag{4.14}$$

and

$$CN(G) \leqslant \frac{1}{2} S_A(G) - 2 \quad , S_A(G) \geqslant 4 \tag{4.15}$$

hold.

Proof. The two inequalities are proved in separate steps.

Regarding (4.14): Analogous to the proof of (4.12).

Regarding (4.15):

$$\begin{aligned} CN(G) &= S_A(G) - S_N(G) + 1 \\ \Leftrightarrow S_N(G) &= S_A(G) - CN(G) + 1 \end{aligned} \tag{4.16}$$

Let $S_A(G) \geqslant 4$. Then, $S_N(G) \geqslant 5$ (as the only process model with three nodes is the SEQ-1 with two arcs and an EPC with four nodes is impossible). According to (4.13), $CN(G) \leqslant S_N(G) - 5$ holds. This inequality can be transformed as follows:

$$\begin{aligned} CN(G) &\leqslant S_N(G) - 5 \overset{(4.16)}{=} (S_A(G) - CN(G) + 1) - 5 \\ &= S_A(G) - CN(G) - 4 && | + CN(G) \\ \Leftrightarrow \quad 2CN(G) &\leqslant S_A(G) - 4 && | \cdot \tfrac{1}{2} \\ \Leftrightarrow \quad CN(G) &\leqslant \tfrac{1}{2} S_A(G) - 2 \end{aligned}$$

■

Theorem 4.5 *For every EPC* G*, the inequalities*

$$S_N(G) - 1 \leqslant S_A(G) \tag{4.17}$$

and

$$S_A(G) \leqslant 2S_N(G) - 6 \quad , S_N(G) \geqslant 5 \tag{4.18}$$

hold.

Proof. The two inequalities are proved in separate steps.

Regarding (4.17): Analogous to the proof of Lemma 4.1, an EPC G with $S_N(G)$ nodes needs at least $S_N(G) - 1$ arcs to be connected ($S_A(G) \geqslant S_N(G) - 1$). For the sequential process models SEQ-n, $S_A(G) = S_N(G) - 1$ holds (see Table 4.3).

Regarding (4.18):

$$CN(G) = S_A(G) - S_N(G) + 1 \Leftrightarrow S_A(G) = CN(G) + S_N(G) - 1$$

Let $S_N(G) \geqslant 5$.

$$S_A(G) = CN(G) + S_N(G) - 1 \overset{(4.13)}{\leqslant} (S_N(G) - 5) + S_N(G) - 1 = 2S_N(G) - 6$$

■

Theorem 4.6 *For every EPC* G, *the inequalities*

$$0 \leqslant S_C(G) \tag{4.19}$$

and

$$S_C(G) \leqslant \frac{2}{3} S_N(G) - 2 \tag{4.20}$$

hold.

Proof. The two inequalities are proved in separate steps.

Regarding (4.19): An EPC must not have any connectors (e. g., sequential process models SEQ-n). The number of connectors cannot be negative. So, $S_C \geqslant 0$.

Regarding (4.20): For every EPC G,

$$\begin{aligned} S_N(G) &= S_C(G) + S_E(G) + S_F(G) \\ \Leftrightarrow S_C(G) &= S_N(G) - S_E(G) - S_F(G) \end{aligned} \tag{4.21}$$

holds. G has at least one start and one end event as well as one function (three nodes). For each connector, there are two possible cases: (1) There is at least one additional branch which does not re-join with the first one. On this branch, at least one start or end event has to exist. (2) There is at least one additional branch which re-joins with the first one. So, there is a second connector at the re-join place. At least one node lies on the "produced" cycle between the two connectors. So, there is at least half a non-connector node per connector in the second case. Consequently, (4.21) can be transformed into the following inequality.

$$\begin{aligned} S_C(G) &\leqslant S_N(G) - 3 - \tfrac{1}{2} S_C(G) \quad | + \tfrac{1}{2} S_C(G) \\ \Leftrightarrow \tfrac{3}{2} S_C(G) &\leqslant S_N(G) - 3 \quad | \cdot \tfrac{2}{3} \\ \Leftrightarrow S_C(G) &\leqslant \tfrac{2}{3} S_N(G) - 2 \end{aligned}$$

■

Theorem 4.7 *For every EPC* G, *the inequalities*

$$S_N(G) - 2 < CNC_K(G) \tag{4.22}$$

and

$$CNC_K(G) < 4S_N(G) - 16 \quad , S_N(G) \geqslant 5 \tag{4.23}$$

hold.

Proof. The two inequalities are proved in separate steps.

Regarding (4.22):

$$\begin{aligned} CNC_K(G) &= \frac{(S_A(G))^2}{S_N(G)} \\ &\overset{(4.17)}{\geqslant} \frac{(S_N(G)-1)^2}{S_N(G)} = \frac{(S_N(G))^2 - 2S_N(G) + 1}{S_N(G)} = S_N(G) - 2 + \frac{1}{S_N(G)} \\ &> S_N(G) - 2 \end{aligned}$$

Regarding (4.23): Let $S_N(G) \geqslant 5$.

$$\begin{aligned} CNC_K(G) &= \frac{(S_A(G))^2}{S_N(G)} \\ &\overset{(4.18)}{\leqslant} \frac{(2S_N(G)-6)^2}{S_N(G)} = \frac{4(S_N(G))^2 - 24S_N(G) + 36}{S_N(G)} \\ &= 4S_N(G) - 24 + \underbrace{\frac{36}{S_N(G)}}_{\leqslant 7.2} \\ &< 4S_N(G) - 16 \end{aligned}$$

■

Theorem 4.8 *For every EPC* G*, the inequalities*

$$\frac{1}{S_N(G)} \leqslant \Delta(G) \tag{4.24}$$

and

$$\Delta(G) \leqslant \frac{2 - \frac{6}{S_N(G)}}{S_N(G) - 1} \quad , S_N(G) \geqslant 5 \tag{4.25}$$

hold.

Proof. The two inequalities are proved in separate steps.

Regarding (4.24):

$$\Delta(G) = \frac{S_A(G)}{S_N(G)(S_N(G)-1)} \overset{(4.17)}{\geqslant} \frac{S_N(G)-1}{S_N(G)(S_N(G)-1)} = \frac{1}{S_N(G)}$$

Regarding (4.25): Let $S_N(G) \geqslant 5$.

$$\Delta(G) = \frac{S_A(G)}{S_N(G)(S_N(G)-1)} \overset{(4.18)}{\leqslant} \frac{2S_N(G)-6}{S_N(G)(S_N(G)-1)} = \frac{2 - \frac{6}{S_N(G)}}{S_N(G) - 1}$$

■

As a consequence of Theorem 4.8, one can see that between the process model metrics Δ and S_N there is a relation of the type $\Delta(G) \approx \frac{1}{S_N(G)}$.

4.3.3 *Results Concerning Process Model Collection Specific Properties*

In this subsection, the results concerning the process model collection specific properties are presented.

Descriptive Statistics

The values of the descriptive statistics for the selected process model metrics applied to the SAP Reference Model EPCs are listed in Table 4.4. Values which are not meaningful because of the process model metric's scale type are printed in italics. Nevertheless, they are not totally skipped as they often lead to fruitful results[9].

The histograms of the selected process model metrics give more detailed information than just some compact quantitative numbers. They are depicted in Figure 4.4, 4.5 and 4.6.

Most process model metrics (including all "size metrics") have a high frequency for small values and almost continuously decreasing frequencies for increasing metric values resulting in a long tail at the right side. The *number of nodes* metric (S_N) (see Figure 4.5a) is a typical example for this behavior.

The metrics *control flow complexity* (CFC) (see Figure 4.6f) and *join complexity* (JC) (see Figure 4.6g) have a similar behavior with some few extreme outliers (see quartiles in Table 4.4). These are caused by contained OR splits with high out-degree (for metric CFC) and OR joins with high in-degree (for metric JC) respectively.

The metric *coefficient of connectivity* (CNC) (see Figure 4.5f), which is defined as $\mathrm{CNC}(G) := \frac{|A|}{|N|}$, has three different areas of values: (1) $\mathrm{CNC} < 1$ for $|A| < |N| \Rightarrow |A| = |N| - 1$ (as $|A| \geqslant |N| - 1$): The smallest possible value is $\frac{2}{3}$ for process model SEQ-1 (see paragraph *Theoretical Value Ranges* in Subsection 4.3.2). The fraction $\frac{|A|}{|N|} = \frac{|N|-1}{|N|}$ converges to 1 for $|N| \to \infty$. The metric values of all sequential processes SEQ-n form the sequence $(a_n) = (\frac{2}{3}, \frac{4}{5}, \frac{6}{7}, \ldots)$ with $a_n = \frac{2n}{2n+1}$ and are situated in the interval $[\frac{2}{3}, 1)$. (2) $\mathrm{CNC} = 1$ for $|A| = |N|$ (3) $\mathrm{CNC} > 1$ for $|A| > |N|$: The upper bound of CNC is 2 (see paragraph *Theoretical Value Ranges* in Subsection 4.3.2). Within the SAP Reference Model, 325 process models (63.1%) have a metric value less than 1 (31 process models with value $\frac{2}{3}$, 40 process models with 0.8, ...), 75 process models (14.6%) have value 1 and 115 process models (22.3%) have a value greater than 1. The process model with the second greatest value $1\frac{1}{3}$ is an AND-5; the process model with the greatest value $1\frac{15}{21}$ (≈ 1.714) is an AND-17.

---

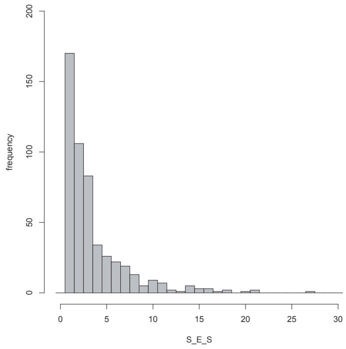
9 Stevens states regarding this problem [148, p. 679]: "In the strictest propriety the ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply a knowledge of something more than the relative rank-order of data. On the other hand, for this 'illegal' statisticizing there can be invoked a kind of pragmatic sanction: In numerous instances it leads to fruitful results. While the outlawing of this procedure would probably serve no good purpose, it is proper to point out that means and standard deviations computed on an ordinal scale are in error to the extent that the successive intervals on the scale are unequal in size."

Table 4.4: Descriptive statistics of the selected process model metrics for the EPCs of the SAP Reference Model (Part 1 of 2).

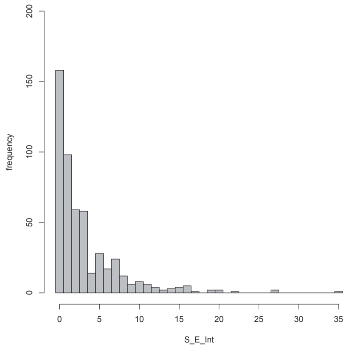
metric	min	Q25	median	Q75	max	R	IQR	MAD	mean	sd	CV
S_{E_S}	1	1	2	4	27	26	3	1.000	3.534	3.594	1.017
$S_{E_{Int}}$	0	0	2	4	35	35	4	2.000	3.134	4.400	1.404
S_{E_E}	1	1	3	5	41	40	4	2.000	4.151	4.671	1.125
S_E	2	4	7	13	76	74	9	4.000	10.819	10.297	0.952
S_F	1	2	3	5	43	42	3	2.000	3.911	3.935	1.006
$S_{S_{AND}}$	0	0	1	1	10	10	1	1.000	1.047	1.492	1.426
$S_{J_{AND}}$	0	0	0	2	10	10	2	0.000	1.045	1.589	1.521
$S_{S_{XOR}}$	0	0	0	1	14	14	1	0.000	0.909	1.483	1.632
$S_{J_{XOR}}$	0	0	0	1	8	8	1	0.000	0.998	1.496	1.499
$S_{S_{OR}}$	0	0	0	1	6	6	1	0.000	0.590	1.108	1.876
$S_{J_{OR}}$	0	0	0	1	7	7	1	0.000	0.429	0.898	2.092
S_C	0	1	3	6	43	43	5	2.000	5.017	6.187	1.233
S_N	3	8	14	24	130	127	16	7.000	19.748	18.577	0.941
S_A	2	7	13	24	138	136	17	8.000	20.056	20.842	1.039

Table 4.4: Descriptive statistics of the selected process model metrics for the EPCs of the SAP Reference Model (Part 2 of 2). Values which are not meaningful because of the process model metric's scale type are printed in italics.

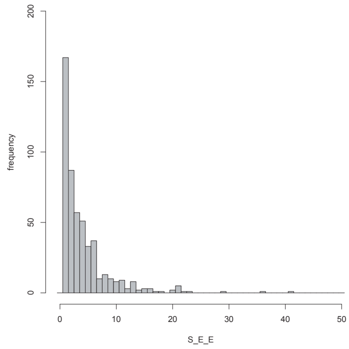
metric	min	Q25	median	Q75	max	R	IQR	MAD	mean	sd	CV
diam	2	4	8	12	38	36	8	4.000	9.291	6.527	0.703
Δ	0.008	0.043	0.077	0.125	0.333	0.325	0.082	0.035	0.099	0.079	0.802
D	0.000	0.000	0.000	0.040	0.727	0.727	0.040	0.000	*0.031*	*0.062*	*1.997*
CNC	0.667	0.889	0.947	1.000	1.714	1.048	0.111	0.058	0.946	0.124	0.131
CNC_K	1.333	6.562	13.067	25.037	166.631	165.297	18.475	7.924	20.588	23.526	1.143
CN	0	0	0	1	26	26	1	0.000	1.309	2.859	2.184
$\overline{d_C}$	0.000	3.000	3.333	3.800	18.000	18.000	0.800	0.333	3.332	1.510	0.453
$\widehat{d_C}$	0	3	4	5	19	19	2	1.000	4.332	2.731	0.630
Π	0.105	0.467	0.600	0.714	1.000	0.895	0.248	0.114	*0.599*	*0.201*	*0.336*
Ξ	0.000	0.086	0.200	0.400	1.000	1.000	0.314	0.133	0.281	0.282	1.003
Λ	0	0	0	1	4	4	1	0.000	0.569	0.722	1.270
MM	0	2	4	8	40	40	6	3.000	5.876	6.137	1.044
CH	0.000	0.000	0.579	0.793	1.000	1.000	0.793	0.367	*0.431*	*0.383*	*0.889*
CYC	0.000	0.000	0.000	0.000	0.722	0.722	0.000	0.000	0.017	0.083	4.808
TS	0	0	1	4	33	33	4	1.000	3.111	4.949	1.591
CFC	0	1	3	9	262,143	262,143	8	3.000	881.596	12,999.475	14.745
JC	0	1	3	8	32,784	32,784	7	3.000	209.058	2,502.365	11.970
CP	0.003	0.029	0.056	0.119	0.333	0.330	0.090	0.037	*0.091*	*0.088*	*0.970*
CC	0.005	0.057	0.115	0.242	0.500	0.495	0.185	0.073	*0.178*	*0.159*	*0.892*



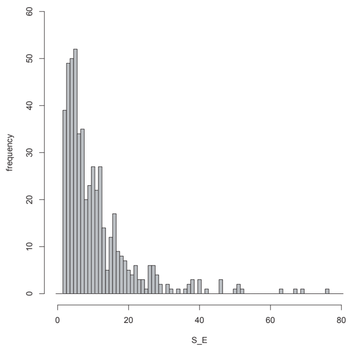
(a) S_{E_S} histogram.



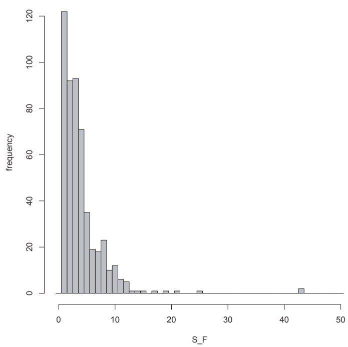
(b) $S_{E_{Int}}$ histogram.



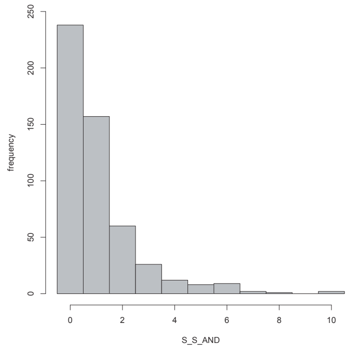
(c) S_{E_E} histogram.



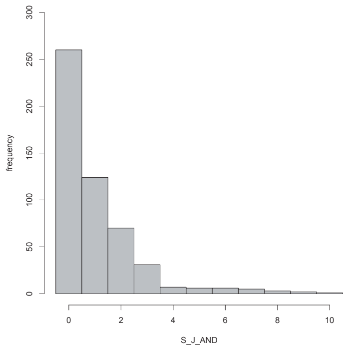
(d) S_E histogram.



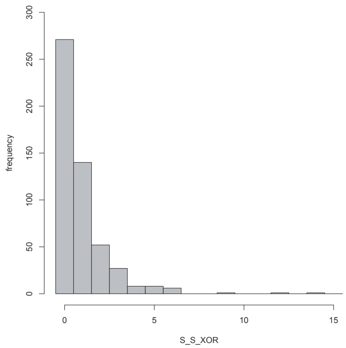
(e) S_F histogram.



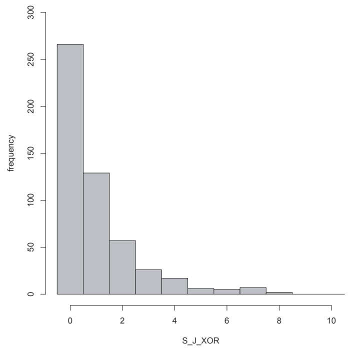
(f) $S_{S_{AND}}$ histogram.



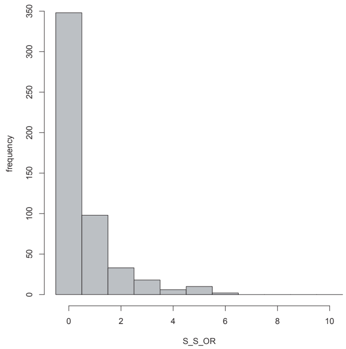
(g) $S_{J_{AND}}$ histogram.



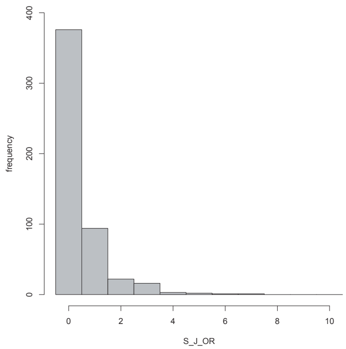
(h) $S_{S_{XOR}}$ histogram.



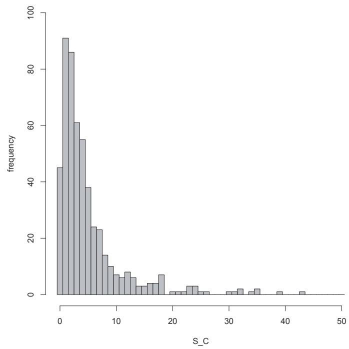
(i) $S_{J_{XOR}}$ histogram.



(j) $S_{S_{OR}}$ histogram.

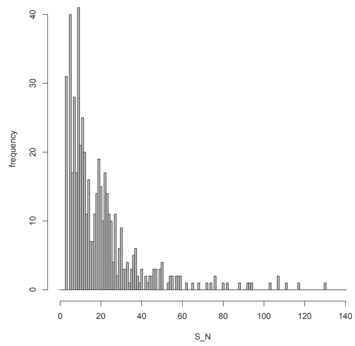


(k) $S_{J_{OR}}$ histogram.

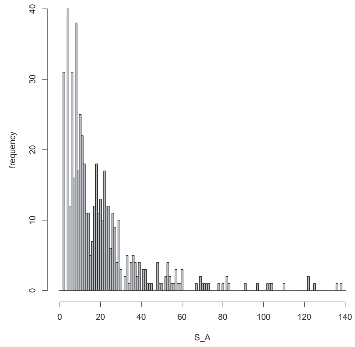


(l) S_C histogram.

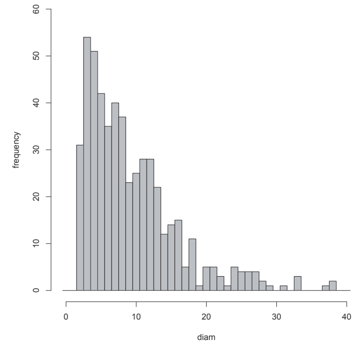
Figure 4.4: Histograms of the selected process model metrics (Part 1 of 3).



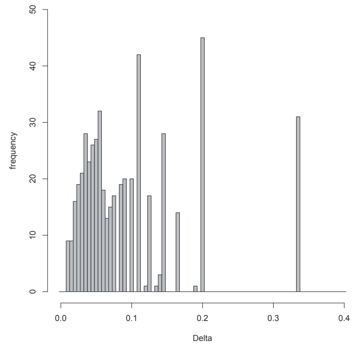
(a) S_N histogram.



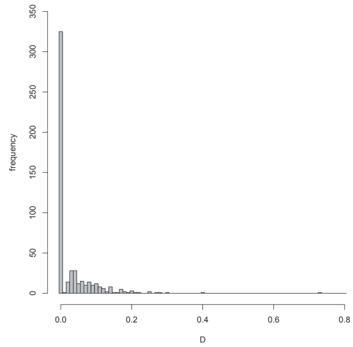
(b) S_A histogram.



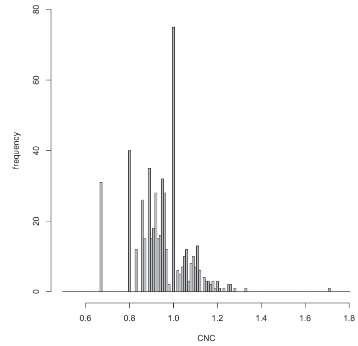
(c) $diam$ histogram.



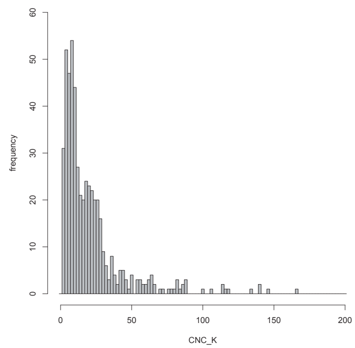
(d) Δ histogram.



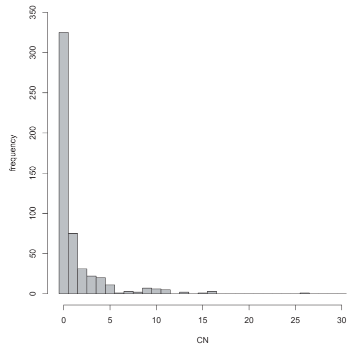
(e) D histogram.



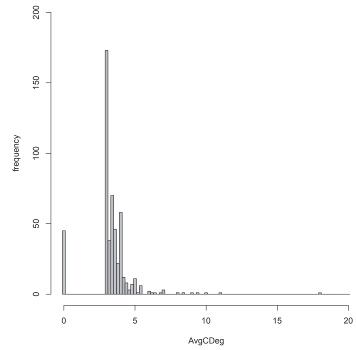
(f) CNC histogram.



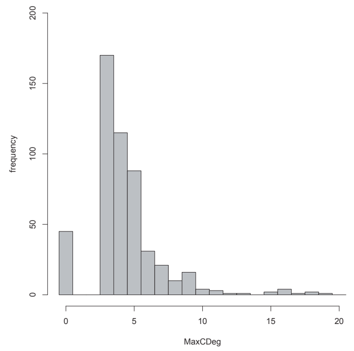
(g) CNC_K histogram.



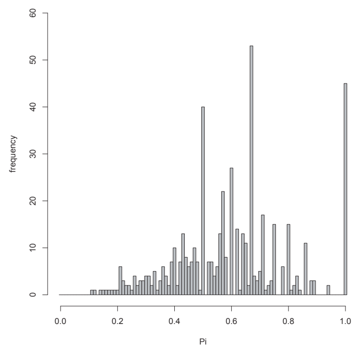
(h) CN histogram.



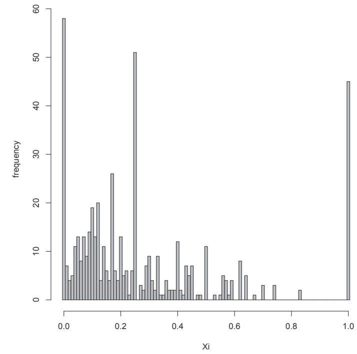
(i) $\overline{d_C}$ histogram.



(j) $\widehat{d_C}$ histogram.

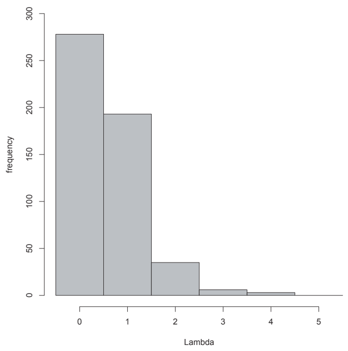


(k) Π histogram.

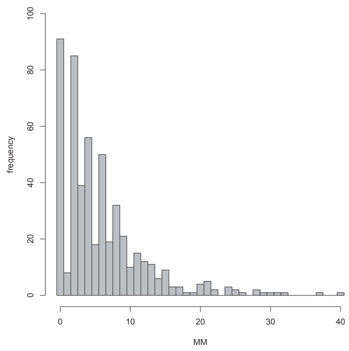


(l) Ξ histogram.

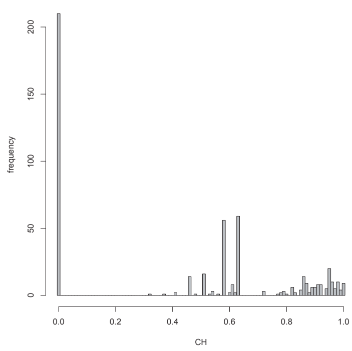
Figure 4.5: Histograms of the selected process model metrics (Part 2 of 3).



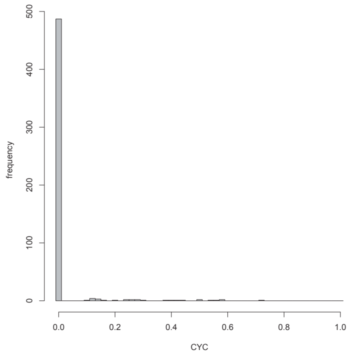
(a) Λ histogram.



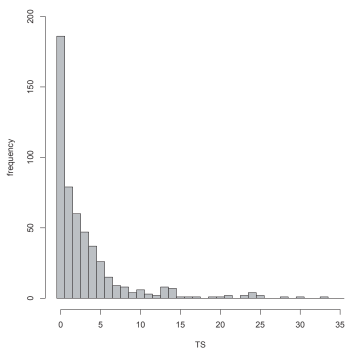
(b) MM histogram.



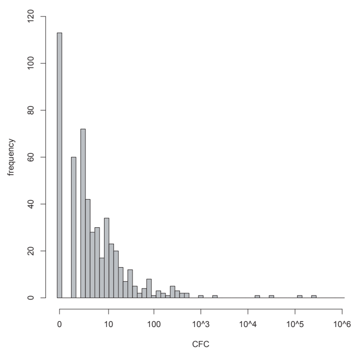
(c) CH histogram.



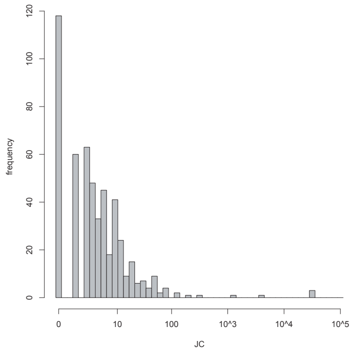
(d) CYC histogram.



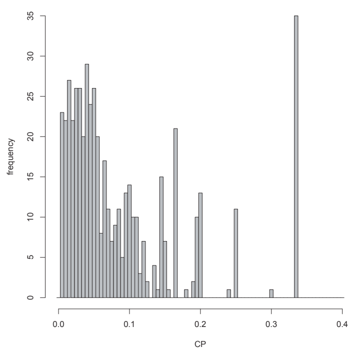
(e) TS histogram.



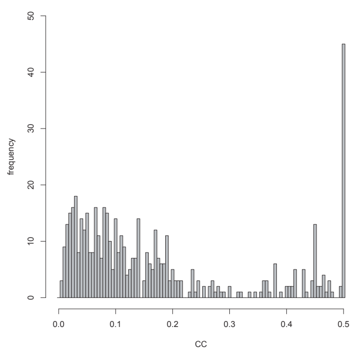
(f) CFC histogram (logarithmic scale on horizontal axis).



(g) JC histogram (logarithmic scale on horizontal axis).



(h) CP histogram.



(i) CC histogram.

Figure 4.6: Histograms of the selected process model metrics (Part 3 of 3).

Looking at the metrics *average* ($\overline{d_C}$) (see Figure 4.5i) and *maximum connector degree* ($\widehat{d_C}$) (see Figure 4.5j), one finds 45 process models (8.7%) with value 0. These are exactly the sequential process models SEQ-n, which do not contain any connectors at all. As a split connector has exactly one incoming and at least two outgoing arcs (vice versa for join connectors), there is no process model with values in the open interval $(0, 3)$. Starting with value 3, the frequencies for increasing metric values are decreasing. As a process model's *maximum* connector degree is at least as high as its *average* connector degree, there are larger metric values for $\widehat{d_C}$ than for $\overline{d_C}$. The value distribution for the maximum connector degree metric is also shifted a little bit upwards compared to the distribution for average connector degree.

210 of 515 process models (40.8%) have a *heterogeneity* metric (CH) (see Figure 4.6c) value of 0 (45 sequential process models SEQ-n and 165 process models with only one connector type) and nine have value 1 (same number of all three connector types). There are 14 process models with value ≈ 0.455 (ratio $0 : 0.2 : 0.8$), 16 process models with value ≈ 0.512 (ratio $0 : 0.25 : 0.75$), 56 process models with value ≈ 0.579 (ratio $0 : \frac{1}{3} : \frac{2}{3}$) and 56 process models with value ≈ 0.631 (ratio $0 : 0.5 : 0.5$). Those process models with higher values up to 1 have an even more equal ratio of connector types.

487 of the 515 process models (94.6%) have a *cyclicity* metric (CYC) value of 0—meaning that they do not contain any (directed) cycles at all (see Figure 4.6d). The remaining process models have values almost equally distributed up to ≈ 0.722.

In contrast, only 325 process models (63.1%) have a *cyclomatic number* metric (CN) value of 0—so, they do not even contain an undirected cycle (see Figure 4.5h).

Correlations

Spearman's rank correlation coefficients (see Section C.2) and Pearson's product-moment correlation coefficients for the selected process model metrics applied to the SAP Reference Model are listed in Table 4.5 and 4.6 respectively. Coefficients which do not differ significantly from 0 (two-sided test, $p = 0.05$) are printed in italics in both tables. Values which are not meaningful because of the process model metric's scale type are skipped. Table cells are colored according to their absolute values—the higher the absolute value the darker the color.

Additional to the correlation coefficients, also the 33×33 scatter plot matrix was looked at. That way, also non-monotonic and non-linear dependencies between two process model metrics can be identified. Because of length restrictions, only scatter plots with an "interesting" behavior are depicted.

The "size metrics" S_E, S_C, S_N and S_A are highly linear correlated with each other (all correlation coefficients > 0.93). Two examples are given in Figure 4.7. The boundaries according to Theorem 4.5 and 4.6 respectively are also drawn within the figures. They indicate the possibility for the existence of a linear correlation—what is now confirmed for the SAP Reference Model.

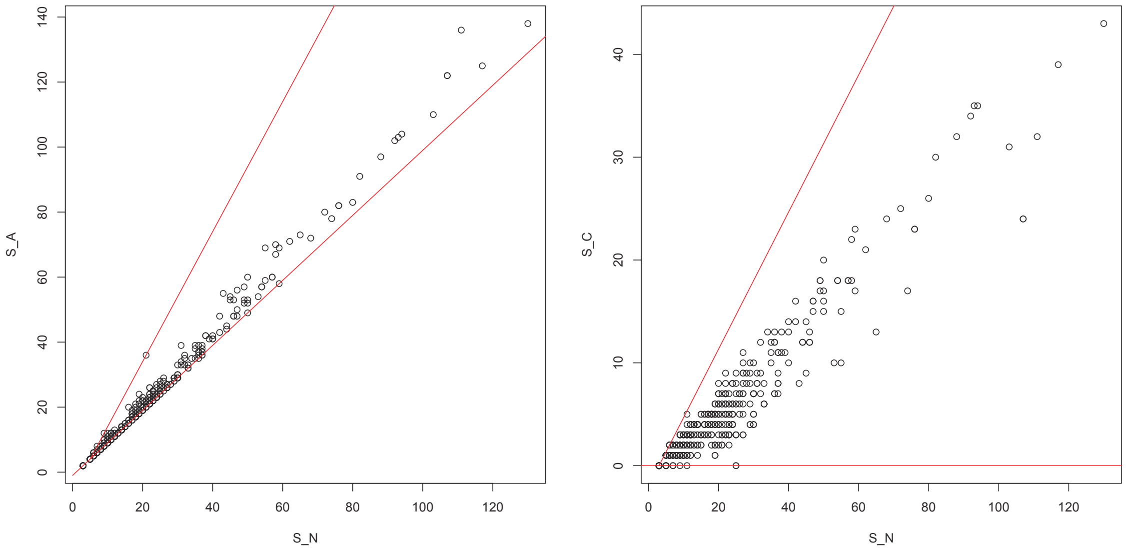
Metric CNC_K is also highly linear correlated with the "size metrics" S_E, S_C, S_N and S_A (all correlation coefficients > 0.94). At first glance, this may be a little surprising because of the power of 2 in the definition ($CNC_K(G) = \frac{(S_A(G))^2}{S_N(G)}$). For the correlation between the metrics CNC_K and S_N, Theorem 4.7 indicates the

Table 4.5: Spearman's rank correlation coefficients of the selected process model metrics for the EPCs of the SAP Reference Model. Coefficients which do not differ significantly from 0 (two-sided test, $\alpha = 0.05$) are printed in italics.

	$S_{E_{Int}}$	S_{E_E}	S_E	S_F	$S_{S_\wedge}$	$S_{J_\wedge}$	$S_{S_\times}$	$S_{J_\times}$	$S_{S_\vee}$	$S_{J_\vee}$	S_C	S_N	S_A	diam	Δ	D	CNC	CNC_K	CN	$\overline{d_C}$	$\widehat{d_C}$	Π	Ξ	Λ	MM	CH	CYC	TS	CFC	JC	CP	CC
S_{E_S}	0.363	0.384	0.702	0.373	0.390	0.649	0.231	0.597	0.279	0.470	0.691	0.670	0.648	0.523	−0.689	0.118	0.440	0.629	0.268	0.429	0.610	−0.508	−0.491	0.398	0.653	0.598	0.147	0.395	0.331	0.852	−0.703	−0.589
$S_{E_{Int}}$		0.431	0.776	0.872	0.505	0.377	0.449	0.514	0.420	0.308	0.671	0.818	0.822	0.922	−0.808	0.547	0.760	0.825	0.651	0.273	0.422	−0.244	*−0.013*	0.648	0.450	0.463	0.264	0.534	0.552	0.543	−0.630	−0.383
S_{E_E}			0.738	0.329	0.657	0.252	0.570	0.417	0.588	0.106	0.695	0.670	0.646	0.545	−0.689	0.096	0.434	0.628	0.250	0.496	0.633	−0.544	−0.536	0.367	0.697	0.597	0.149	0.793	0.835	0.336	−0.713	−0.633
S_E				0.705	0.638	0.513	0.509	0.627	0.530	0.386	0.875	0.973	0.956	0.872	−0.984	0.341	0.741	0.940	0.496	0.574	0.756	−0.580	−0.422	0.576	0.801	0.711	0.205	0.731	0.732	0.742	−0.941	−0.706
S_F					0.505	0.402	0.323	0.465	0.306	0.295	0.631	0.807	0.819	0.847	−0.790	0.548	0.777	0.825	0.628	0.343	0.457	−0.207	*0.039*	0.604	0.403	0.371	0.194	0.513	0.447	0.535	−0.597	−0.330
$S_{S_\wedge}$						0.450	0.358	0.456	0.321	0.173	0.743	0.680	0.681	0.633	−0.674	0.378	0.606	0.681	0.500	0.329	0.500	−0.467	−0.459	0.542	0.497	0.525	0.137	0.796	0.565	0.420	−0.574	−0.320
$S_{J_\wedge}$							0.158	0.289	0.210	0.325	0.624	0.556	0.564	0.493	−0.544	0.375	0.526	0.567	0.479	0.292	0.479	−0.453	−0.432	0.529	0.403	0.457	0.245	0.436	0.280	0.568	−0.449	−0.191
$S_{S_\times}$								0.509	0.264	*0.025*	0.576	0.512	0.508	0.508	−0.512	0.275	0.433	0.504	0.381	0.151	0.298	−0.388	−0.365	0.455	0.431	0.498	0.232	0.301	0.621	0.313	−0.539	−0.589
$S_{J_\times}$									0.387	0.128	0.695	0.649	0.649	0.603	−0.645	0.397	0.570	0.645	0.517	0.243	0.439	−0.473	−0.400	0.534	0.541	0.612	0.249	0.423	0.489	0.688	−0.674	−0.682
$S_{S_\vee}$										0.267	0.539	0.510	0.515	0.442	−0.505	0.313	0.446	0.515	0.441	0.335	0.468	−0.439	−0.367	0.456	0.525	0.560	0.193	0.688	0.761	0.406	−0.553	−0.599
$S_{J_\vee}$											0.399	0.396	0.403	0.315	−0.389	0.309	0.379	0.407	0.396	0.226	0.351	−0.322	−0.203	0.382	0.369	0.397	*0.036*	0.269	0.241	0.646	−0.409	−0.373
S_C												0.909	0.913	0.825	−0.900	0.529	0.821	0.912	0.651	0.449	0.675	−0.670	−0.618	0.732	0.757	0.783	0.256	0.768	0.769	0.791	−0.863	−0.708
S_N													0.996	0.916	−0.996	0.473	0.840	0.989	0.610	0.543	0.733	−0.570	−0.410	0.668	0.748	0.698	0.218	0.748	0.720	0.765	−0.918	−0.669
S_A														0.916	−0.985	0.535	0.878	0.998	0.662	0.554	0.740	−0.594	−0.421	0.702	0.726	0.691	0.233	0.759	0.725	0.772	−0.907	−0.663
diam															−0.910	0.522	0.820	0.913	0.629	0.343	0.514	−0.389	−0.248	0.684	0.575	0.601	0.241	0.632	0.639	0.649	−0.769	−0.490
Δ																−0.409	−0.797	−0.973	−0.556	−0.532	−0.723	0.545	0.396	−0.630	−0.765	−0.701	−0.206	−0.733	−0.713	−0.754	0.924	0.673
D																	0.838	0.577	0.951	0.267	0.361	−0.501	−0.268	0.791	0.162	0.315	0.292	0.440	0.431	0.490	−0.347	−0.271
CNC																		0.902	0.861	0.523	0.653	−0.647	−0.437	0.820	0.513	0.571	0.282	0.686	0.658	0.709	−0.723	−0.540
CNC_K																			0.696	0.562	0.742	−0.609	−0.426	0.722	0.709	0.683	0.243	0.764	0.727	0.774	−0.895	−0.656
CN																				0.303	0.460	−0.556	−0.319	0.838	0.317	0.432	0.337	0.560	0.539	0.597	−0.492	−0.392
$\overline{d_C}$																					0.910	−0.712	−0.431	0.278	0.535	0.353	*0.076*	0.578	0.493	0.471	−0.599	−0.528
$\widehat{d_C}$																						−0.794	−0.549	0.466	0.691	0.545	0.136	0.717	0.635	0.654	−0.777	−0.667
Π																							0.783	−0.524	−0.594	−0.561	−0.266	−0.639	−0.639	−0.628	0.672	0.679
Ξ																								−0.404	−0.524	−0.553	−0.122	−0.533	−0.542	−0.496	0.524	0.587
Λ																									0.382	0.514	0.287	0.571	0.594	0.633	−0.543	−0.455
MM																										0.771	0.163	0.612	0.655	0.639	−0.822	−0.747
CH																											0.211	0.578	0.652	0.663	−0.756	−0.686
CYC																												0.172	0.227	0.241	−0.190	−0.190
TS																													0.804	0.485	−0.691	−0.509
CFC																														0.468	−0.758	−0.744
JC																															−0.779	−0.707
CP																																0.831

Table 4.6: Pearson's product-moment correlation coefficients of the selected process model metrics for the EPCs of the SAP Reference Model. Values which are not meaningful because of the process model metric's scale type are skipped. Coefficients which do not differ significantly from 0 (two-sided test, $\alpha = 0.05$) are printed in italics.

	$S_{E_{Int}}$	S_{E_E}	S_E	S_F	$S_{S_\wedge}$	$S_{J_\wedge}$	$S_{S_\times}$	$S_{J_\times}$	$S_{S_\vee}$	$S_{J_\vee}$	S_C	S_N	S_A	diam	Δ	D	CNC	CNC_K	CN	$\overline{d_C}$	$\widehat{d_C}$	Π	Ξ	Λ	MM	CH	CYC	TS	CFC	JC	CP	CC
S_{E_S}	0.475	0.475	0.767	0.295	0.468	0.723	0.380	0.575	0.396	0.552	0.680	0.714	0.696	0.520	−0.476		0.349	0.674	0.433	0.308	0.593		−0.373	0.463	0.692		*0.016*	0.534	*-0.016*	0.343		
$S_{E_{Int}}$		0.510	0.824	0.750	0.635	0.598	0.564	0.685	0.573	0.498	0.783	0.876	0.895	0.866	−0.522		0.595	0.905	0.831	0.129	0.400		−0.122	0.675	0.581		0.120	0.683	*0.044*	*-0.032*		
S_{E_E}			0.837	0.217	0.732	0.482	0.760	0.628	0.700	0.213	0.791	0.773	0.747	0.529	−0.469		0.349	0.716	0.422	0.261	0.530		−0.397	0.517	0.746		*0.049*	0.874	0.191	*-0.050*		
S_E				0.522	0.767	0.726	0.718	0.778	0.701	0.502	0.930	0.975	0.964	0.791	−0.602		0.534	0.947	0.698	0.281	0.618		−0.362	0.684	0.828		*0.079*	0.874	0.100	*0.084*		
S_F					0.485	0.489	0.274	0.410	0.268	0.326	0.503	0.669	0.678	0.741	−0.471		0.546	0.683	0.599	0.185	0.272		*-0.013*	0.446	0.288		*0.049*	0.417	*0.007*	*-0.041*		
$S_{S_\wedge}$						0.624	0.622	0.640	0.462	0.281	0.829	0.804	0.798	0.695	−0.481		0.477	0.784	0.589	0.142	0.363		−0.366	0.627	0.611		*0.062*	0.775	*-0.029*	*-0.045*		
$S_{J_\wedge}$							0.403	0.520	0.445	0.473	0.778	0.765	0.766	0.648	−0.426		0.452	0.759	0.609	0.156	0.399		−0.348	0.622	0.532		0.125	0.649	*-0.040*	*0.021*		
$S_{S_\times}$								0.685	0.548	0.152	0.779	0.716	0.707	0.559	−0.398		0.373	0.693	0.506	0.089	0.278		−0.320	0.564	0.573		0.123	0.602	*-0.037*	*-0.040*		
$S_{J_\times}$									0.640	0.285	0.850	0.801	0.805	0.705	−0.474		0.495	0.800	0.659	0.125	0.372		−0.348	0.661	0.676		0.120	0.670	*0.002*	*-0.025*		
$S_{S_\vee}$										0.390	0.748	0.694	0.695	0.566	−0.386		0.412	0.689	0.556	0.143	0.383		−0.318	0.601	0.643		*0.074*	0.773	0.116	*-0.033*		
$S_{J_\vee}$											0.510	0.517	0.529	0.453	−0.308		0.352	0.535	0.495	0.155	0.351		−0.208	0.451	0.403		*0.037*	0.407	*0.079*	0.087		
S_C												0.955	0.954	0.805	−0.549		0.565	0.945	0.751	0.174	0.466		−0.426	0.779	0.759		0.124	0.857	*0.006*	*-0.014*		
S_N													0.996	0.864	−0.616		0.600	0.984	0.763	0.253	0.556		−0.345	0.733	0.773		0.096	0.859	*0.059*	*0.033*		
S_A														0.861	−0.598		0.627	0.996	0.818	0.263	0.561		−0.343	0.749	0.759		0.103	0.863	*0.057*	*0.027*		
diam															−0.675		0.648	0.850	0.665	0.148	0.351		−0.237	0.722	0.556		0.121	0.644	*-0.027*	*-0.028*		
Δ																	−0.777	−0.573	−0.352	−0.566	−0.600		0.613	−0.495	−0.569		−0.099	−0.462	*-0.054*	*-0.062*		
D																																
CNC																		0.649	0.676	0.599	0.606		−0.567	0.679	0.454		0.203	0.529	*0.033*	*0.019*		
CNC_K																			0.866	0.275	0.564		−0.338	0.756	0.740		0.107	0.862	*0.055*	*0.021*		
CN																				0.271	0.476		−0.259	0.699	0.515		0.131	0.714	*0.033*	*-0.019*		
$\overline{d_C}$																					0.812		−0.622	0.168	0.357		*0.017*	0.326	0.245	0.255		
$\widehat{d_C}$																							−0.572	0.391	0.639		*0.021*	0.614	0.342	0.342		
Π																																
Ξ																								−0.383	−0.463		−0.087	−0.380	*-0.057*	*-0.063*		
Λ																									0.491		0.211	0.653	*-0.023*	*-0.045*		
MM																											*0.070*	0.743	0.163	0.197		
CH																																
CYC																												*0.077*	*-0.014*	*-0.017*		
TS																													0.217	*-0.042*		
CFC																														*-0.005*		
JC																																
CP																																



(a) S_A/S_N plot with theoretical boundaries according to Theorem 4.5.

(b) S_C/S_N plot with theoretical boundaries according to Theorem 4.6.

Figure 4.7: S_A/S_N and S_C/S_N plots as examples of strong linear correlations between "size metrics".

possibility for such a strong linear correlation. As seen above, the ratio between S_A and S_N is almost a constant c ($\frac{S_A(G)}{S_N(G)} \approx c$), so

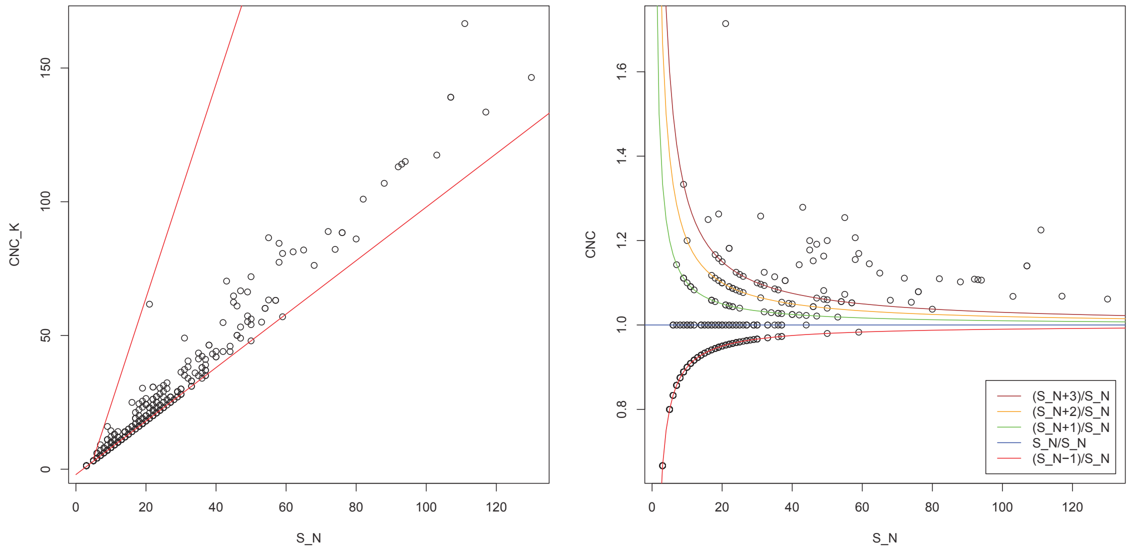
$$\frac{CNC_K(G)}{S_N(G)} = \frac{\frac{(S_A(G))^2}{S_N(G)}}{S_N(G)} \approx \frac{c \cdot S_A(G)}{S_N(G)} \approx c^2 \tag{4.26}$$

holds. The correlation between the metrics CNC_K and S_N as an example is depicted in Figure 4.8a.

For metric CNC, things are different. Here, the correlation coefficients for the pairs of CNC and the "size metrics" are much lower. Also the CNC/S_N plot (see Figure 4.8b) shows no linear or at least monotonic correlation. Yet, there are points forming several curves. Having in mind that $S_A(G) = S_N(G) + a \quad , a \in \{-1, 0, 1, \dots, 2S_N(G) - 6\}$ (Theorem 4.5), these curves are easily explainable.

There are two proposed process model metrics supposed to measure "density": Δ and D. As D is on ordinal scale, only Spearman's rank correlation coefficients can be considered during the following analysis. The rank correlation coefficient between both metrics is -0.409. When one looks at the corresponding Δ/D plot (see Figure 4.9a), many points with a D value of 0 (325 of 515 process models—this equates 63.1%) attract attention. The reason is the definition of D, which assigns $D(G) = 0$ for EPCs G with $S_C(G) \leqslant 1$. If one removes these points from the analysis, the rank correlation coefficient is still only 0.598—what is quite small. So, as a result, one can state that both "density metrics" measure something quite different.

In contrast, the metrics Δ and CP are highly correlated (rank correlation coefficient 0.924, see Figure 4.9b). This is only surprising at first glance as both metrics measure a ratio between existing arcs and possible arcs—the number of arcs



(a) CNC_K / S_N plot with theoretical boundaries according to Theorem 4.7.

(b) CNC / S_N plot.

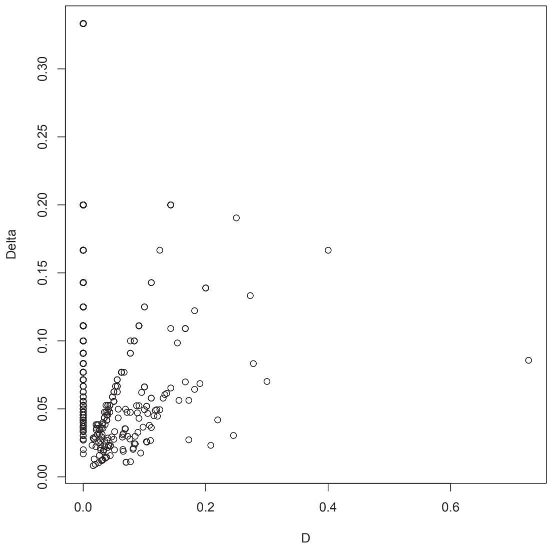
Figure 4.8: CNC_K / S_N and CNC / S_N plots.

divided by the number of possible arcs between the nodes for metric Δ and the weighted number of arcs between tasks and the possible number of arcs between the tasks in the case of metric CP. This fact supports the assumption that metric CP with its relatively complicated computation of weights has little additional benefit compared to metric Δ.

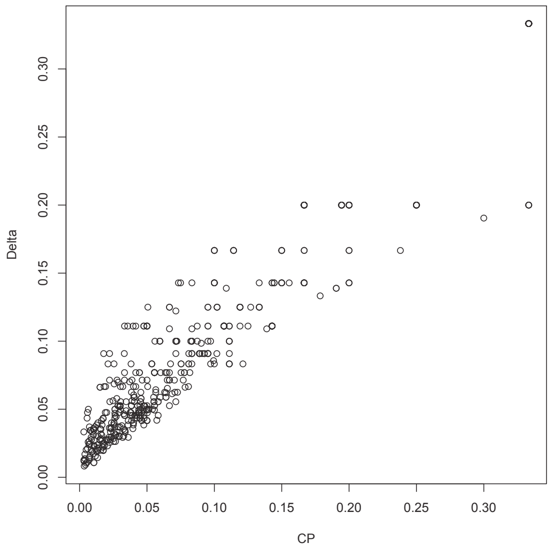
Metric Δ has plots with the "size metrics" S_E, S_C, S_N and S_A very similar to the $\frac{1}{x}/x$ plot. This is also reflected in quite large absolute values of Spearman's rank correlation coefficients (all correlation coefficients < -0.9). For the metrics Δ and S_N (see Figure 4.9c), this was already "predicted" by Theorem 4.8.

Looking at the correlation coefficients of metric *cycling* (CYC), one gets the impression that it is not correlated with any other metric. Yet, this indication is only half the truth. Figure 4.10 depicts the CYC / S_N plot. The overwhelming majority of points has a CYC value of 0 (487 of 515 process models—this equates 94.6%). If one only looks at the remaining points (only 28 points!), one gets a Spearman's rank correlation coefficient of -0.885 and a Pearson's product-moment correlation coefficient of -0.809. So, the result of this analysis is that it does not depend on the number of nodes whether an EPC has a cycle (arc directions are *not* ignored) or not—yet, if it has at least one, the CYC metric values decrease for larger EPCs.

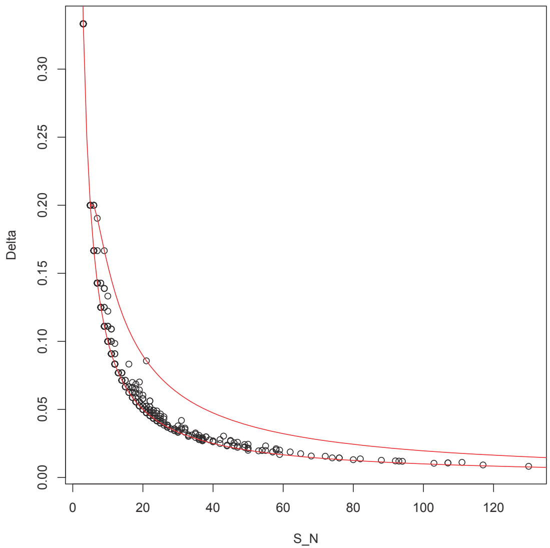
At the end of this paragraph, the CN / S_N (see Figure 4.11a) and CN / S_A (see Figure 4.11b) plots are depicted in order to show how the points for the SAP Reference Model lie between the boundaries of Theorem 4.3 and 4.4 respectively.



(a) Δ/D plot.



(b) Δ/CP plot.



(c) Δ/S_N plot with theoretical boundaries according to Theorem 4.8.

Figure 4.9: Δ/D, Δ/CP and Δ/S_N plots.

Principal Component Analysis

The last step of the analysis is a PCA of the process model metric values. As the metric values of metrics CFC and JC have some few but extreme outliers (see Table 4.4), the values of these two metrics are logarithmized before the PCA.

The result of this PCA is a linear transformation of the 33-dimensional data (33 process model metric values per process model) into a new system of 33 basis vectors (also called "components") which are ordered decreasingly according to their proportion of variance. The corresponding resulting numbers are listed in Table 4.7.

In the previous analysis steps, large correlations between some process model metrics were identified. The results of the PCA reflect this fact: The first compo-

Table 4.7: Proportion of variance and cumulative proportion of the 33 components of the PCA.

component	proportion of variance	cumulative proportion
1	54.402%	54.402%
2	10.290%	64.692%
3	6.761%	71.453%
4	4.756%	76.209%
5	4.166%	80.376%
6	3.530%	83.906%
7	2.629%	86.535%
8	2.379%	88.914%
9	1.975%	90.889%
10	1.312%	92.201%
11	1.163%	93.364%
12	1.125%	94.489%
13	0.875%	95.364%
14	0.783%	96.147%
15	0.677%	96.824%
16	0.598%	97.422%
17	0.462%	97.884%
18	0.437%	98.322%
19	0.415%	98.737%
20	0.278%	99.015%
21	0.255%	99.270%
22	0.221%	99.491%
23	0.172%	99.663%
24	0.126%	99.789%
25	0.119%	99.908%
26	0.052%	99.960%
27	0.035%	99.995%
28	0.005%	100.000%
29	0.000%	100.000%
30	0.000%	100.000%
31	0.000%	100.000%
32	0.000%	100.000%
33	0.000%	100.000%

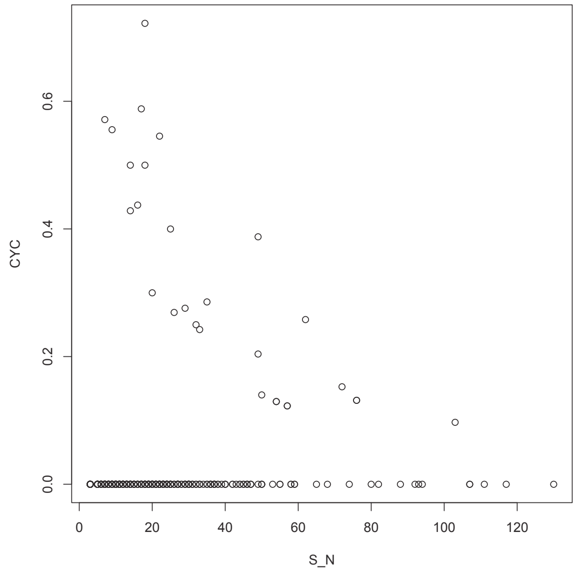
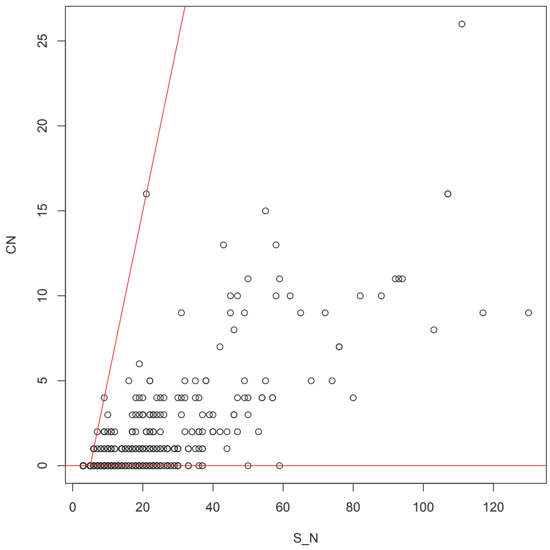
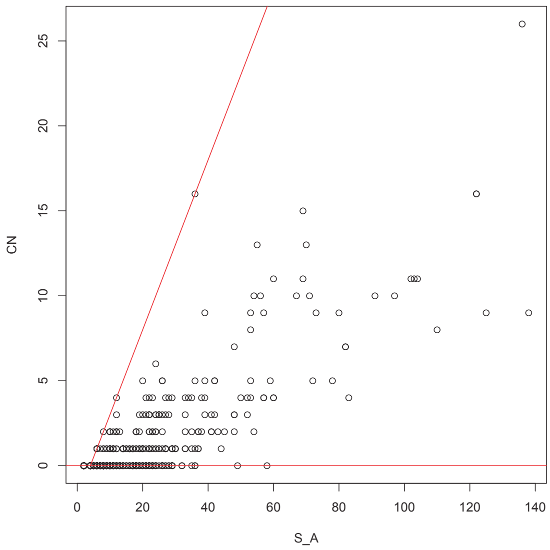


Figure 4.10: CYC/S_N plot.



(a) CN/S_N plot with theoretical boundaries according to Theorem 4.3.

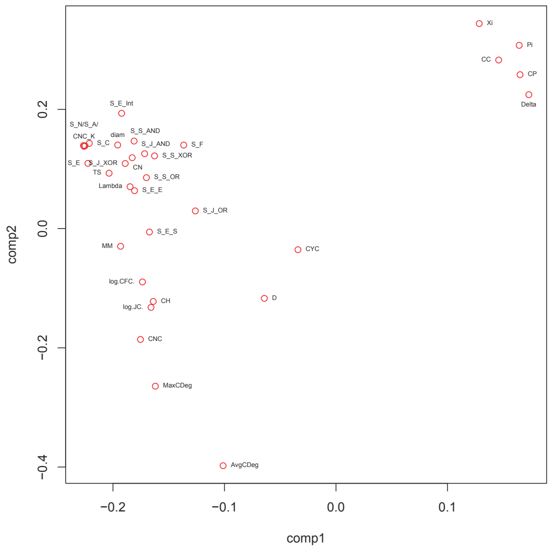


(b) CN/S_A plot with theoretical boundaries according to Theorem 4.4.

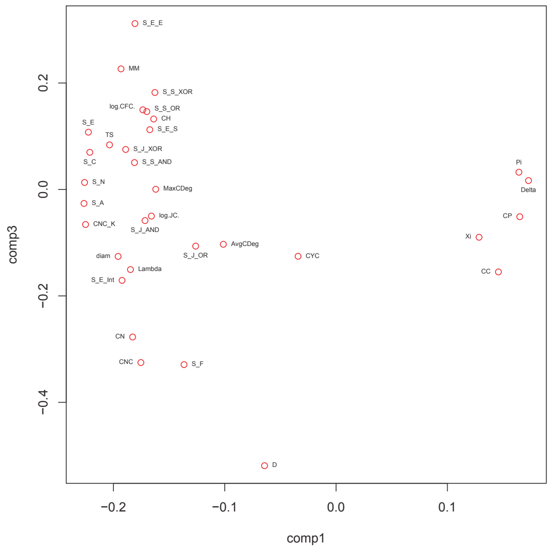
Figure 4.11: CN/S_N and CN/S_A plots.

nent comprises more than half (54.402%) of the total data variance; the first nine components (of 33) together more than 90%. So, there is a lot of "redundancy" in the data of the 33 process model metrics.

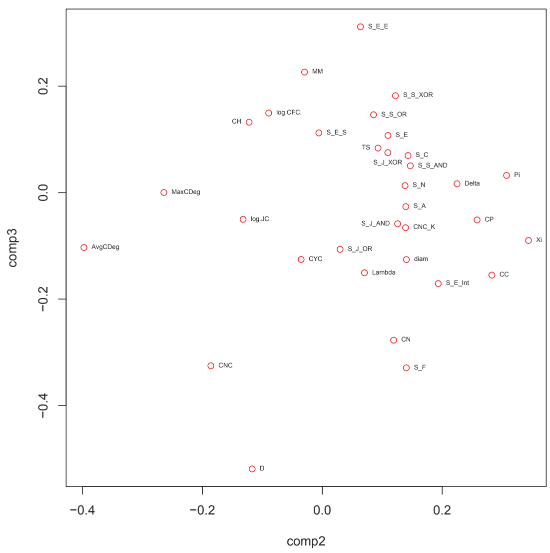
Finally, the location of the original 33 process model metrics in the PCA's new system of basis vectors (called "loadings" in the PCA literature) is to be examined. As the first three components comprise more than 71% of the total data variance, a visualization which is restricted to these first three dimensions of the 33-dimensional data is meaningful. The results are depicted in Figure 4.12. As a 3D visualization is hard to interpret as long as it is not interactively rotatable, three 2D plots are used instead.



(a) comp2/comp1 loadings plot.



(b) comp3/comp1 loadings plot.



(c) comp3/comp2 loadings plot.

Figure 4.12: Selected process model metrics transformed by PCA (first three components).

Looking at these plots (especially Figure 4.12a), one can identify three clusters. The first one consists of almost all “size metrics” (the metrics S_A, S_C, S_N and CNC_K have almost the same comp1 and comp2 values). The second cluster comprises the metrics Δ, Π, Ξ, CP and CC and is clearly separated from the first one. The third cluster contains the remaining metrics. It is not as cohesive as the other two and is located next to the first cluster with the “size metrics”.

4.4 CONCLUSION

In this chapter, an approach for reducing the experimental effort for the validation of prediction systems was introduced.

Its main idea is to add an additional analysis step before the selection of the prediction system which shall be validated. In this preceding step, the behavior as well as important properties of process model metrics which are part of the potential prediction systems which shall be validated are first analyzed. Through this, unfavorable properties of process model metrics (e. g., insufficient dispersion of metric values or strong correlation with other process model metrics) can be identified before the high effort for an experimental validation of the corresponding prediction system occurs.

The approach distinguishes between general properties which hold for all process models because of their definition and process model collection specific properties which are only true for the examined process model collection.

The approach was tested with 33 EPC process model metrics and 515 process models from the SAP Reference Model. During this test, some interesting properties could be found.

As general properties, mathematical boundaries for the value pairs of some "size metrics" could be identified. Furthermore, it could be shown, that $\Delta(G) \approx \frac{1}{S_N(G)}$ holds.

Even more process model collection specific properties could be discovered for the SAP Reference Model:

- The metrics CFC and JC have some few extreme outliers. It is very unlikely that there is a linear dependency between one of these metrics and a process model quality measure. The existence of a threshold over which a process model has some undesirable properties (e. g., high error probability) is much more likely.
- 94.6% of all process models have a *cyclicity* metric (CYC) value of 0—meaning that they do not contain any (directed) cycle.
- There are many linear or at least monotonic correlations between the examined process model metrics. This was also confirmed by the result of a PCA. There are three major clusters for the metrics: The first one consists of almost all "size metrics". The second cluster comprises the metrics Δ, Π, Ξ, CP and CC and is clearly separated from the first one. The third (not so cohesive) cluster contains the remaining metrics. Consequently, some metrics do not provide much additional information compared to others.
- The density metrics Δ and D have quite different behaviors. So, they do not measure the same concept.
- On the other hand, the metrics Δ and CP, which both measure some sort of ratio between existing arcs and possible arcs, are highly correlated. So, it seems that metric CP, which has a much more complicated computation rule, has little additional benefit compared to metric Δ.

As future work, it should be examined whether the identified process model collection specific properties also hold for other collections of process models.

Furthermore, the approach could also be applied to other process modeling languages than EPC.

The results of this chapter may be helpful for planning future validation experiments for prediction systems. Maybe, it can contribute to decrease the lack of validation in this way.

5 VISUALIZATION AND CLUSTERING OF PROCESS MODEL COLLECTIONS

5.1 INTRODUCTION

In Chapter 4, general properties of some EPC process model metrics and process model collection specific properties of these metrics for the SAP Reference Model were analyzed.

As most humans are visually thinking beings—preferring pictures to large tables of numbers—, a visualization of large process model collections based on process model metric values would be interesting. Yet, the resulting process model metric data would be very high-dimensional making visualization problematic.

A second interesting question is whether there are clusters of (structurally) similar process models among a process model collection.

In this chapter, an approach for these two goals is proposed. It comprises heatmaps, a compact visualization technique for high-dimensional data originally used in genetics, and scatter plots for dimensionally reduced data using PCA for visualizing the process model metric data. Additionally, clustering is used for analyzing

1. the correlations (see Chapter C) between different process model metrics and
2. finding (structurally) similar process models[1] among a process model collection.

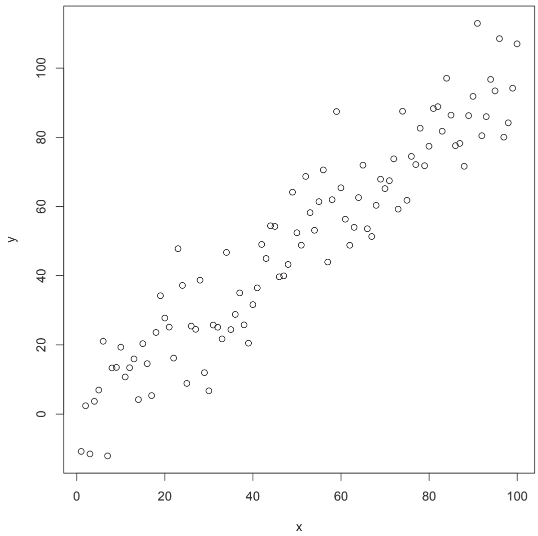
Finally, the proposed approach is applied to the same process model metrics and process model collection as in Chapter 4 to make the findings comparable.

The remainder of this chapter is organized as follows: In Section 5.2, several visualization techniques are presented and assessed for their adequateness for high-dimensional data. Basics of clustering are given in Section 5.3. The visualization and clustering approach is introduced in Section 5.4. Afterwards, Section 5.5 shows the results of an experimental application of the approach. The chapter closes with a conclusion (Section 5.6).

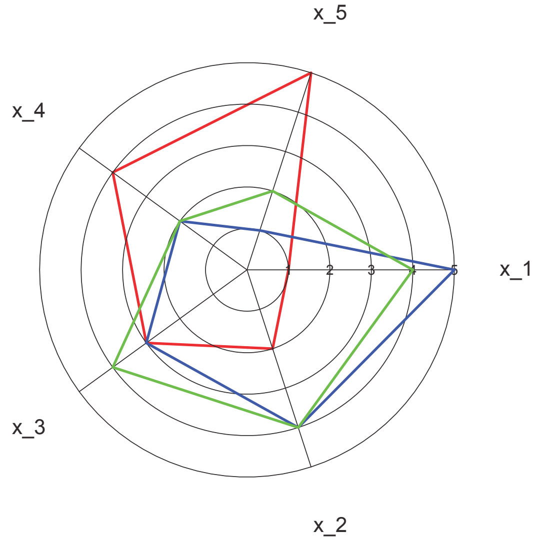
5.2 VISUALIZATION OF HIGH-DIMENSIONAL DATA

A set of n process model metric values of a process model can be represented as a real-valued vector $\vec{x} \in \mathbb{R}^n$. So, the process model metric data of (large) process model collections is high-dimensional data consisting of many data vectors. Consequently, the problem arises how to visualize this data.

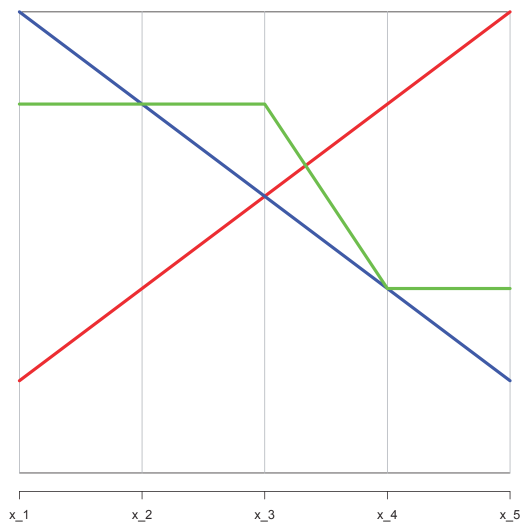
1 The clustering does not consider *behavioral* similarity as, for example, in [164].



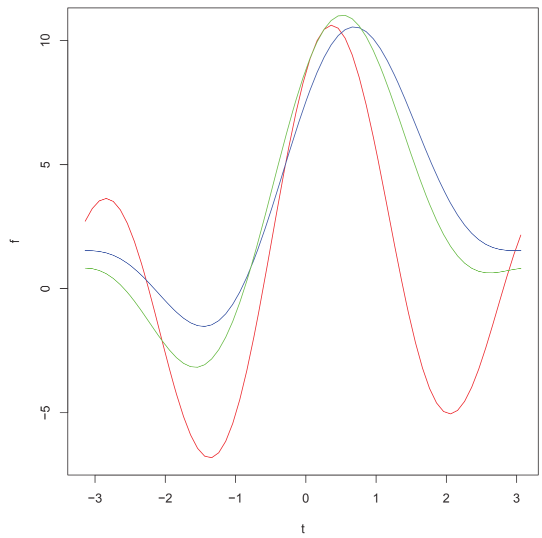
(a) 2-dimensional scatter plot.



(b) Radar chart.



(c) Parallel coordinates plot.



(d) Andrews plot.

Figure 5.1: Examples of several visualization techniques. In (b)–(d), always the same three vectors $\vec{x}_{\mathtt{red}} = (1,2,3,4,5)^\mathsf{T}$, $\vec{x}_{\mathtt{blue}} = (5,4,3,2,1)^\mathsf{T}$ and $\vec{x}_{\mathtt{green}} = (4,4,4,2,2)^\mathsf{T}$ are displayed.

In this section, several different visualization techniques which were found in a literature review are presented and assessed for their adequateness for this purpose.

5.2.1 *Inadequate Visualization Techniques*

In this subsection, several potential visualization techniques which were identified are presented. They all have in common that the subsequent assessment showed their inadequateness because of some major disadvantages.

Scatter Plots

Scatter plots (see Figure 5.1a for an example) can display two or three dimensions. In the 3-dimensional case, a projection of the third dimension onto the 2-dimensional plane is used. Each vector is represented as a point at the corresponding place in the coordinate system.

Scatter plots are good for visualizing large amounts of vectors. But they are only applicable for 2D or at most 3D data. Consequently, they are inadequate for the high-dimensional process model metric data.

Radar Charts

The first ideas of radar charts were published a long time ago by Friedmann in 1862 [46] and by Mayr in 1877 [85, p. 78] respectively.

Radar charts[2] (see Figure 5.1b for an example) are drawn in two dimensions and can display vectors with three or more dimensions and non-negative values. For each dimension, there exists an axis. The axes start in one single center point and are uniformly placed around the 360° of a circle. For depicting a vector, the values of the vector components are marked as points on the corresponding axes. These points are linked by (colored) lines forming a polygon for each vector.

If one wants to display several vectors, one can either draw them in one single radar chart using different colors for each vector (as in Figure 5.1b) or one can draw a radar chart for each single vector.

Radar charts soon become confusing when one increases the number of dimensions and/or depicts many vectors. Consequently, they are inadequate for the high-dimensional process model metric data.

Parallel Coordinates Plots

Parallel coordinates plots (see Figure 5.1c for an example) were proposed by Inselberg in [61]. They can display vectors with two or more dimensions. For each dimension, a parallel coordinate axis is drawn in the plane. For depicting a vector, the values of its components are marked as points on the corresponding axes. These points are linked by (colored) zig-zag lines.

Even though Inselberg states that the largest data set which he has effectively worked with had about 800 dimensions and 10,000 vectors [62, p. 663], one needs a lot of experience with this technique and the use of assisting software (e. g., an automatic classifier [62, p. 664–668]) is advisable.

As for radar charts, parallel coordinates plots soon become confusing for a normal observer when one increases the number of dimensions and/or depicts many vectors. Consequently, they are inadequate for the high-dimensional process model metric data.

2 Radar charts are also called spider charts or star charts.

Figure 5.2: Color legend for heatmaps (blue for minimum and red for maximum values).

Andrews Plots

Andrews plots (see Figure 5.1d for an example) were proposed by Andrews in [2]. They can display vectors with an arbitrary number of dimensions. Each vector $\vec{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ is mapped to a function

$$f_{\vec{x}}(t) := \frac{x_1}{\sqrt{2}} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \dots \quad , -\pi < t < \pi \quad (5.1)$$

in the plot.

Andrews plots preserve the Euclidean distances between the vectors. That means that vectors which lie close together in the n-dimensional space are represented by lines (functions) which also lie close to each other. This property can be used for identifying clusters or outliers among the vectors.

In Figure 5.1d, for example, the close blue and green lines (functions) indicate that the corresponding vectors lie close to each other.

Nevertheless, Andrews plots soon become confusing when one increases the number of dimensions and/or depicts many vectors. Consequently, they are inadequate for the high-dimensional process model metric data.

5.2.2 *Heatmaps*

Heatmaps surmount the described problems of the inadequate visualization techniques. Based on earlier developments in statistics, this visualization technique originally became popular in genetics for depicting microarray data (see [177] for a short overview of its historical development). Recently, this method was adapted by Pryke *et al.* to visualize the individuals (i. e., possible solutions) of population-based multi-objective algorithms (e. g., genetic algorithms) [123].

A heatmap displays the data as a matrix: one row per vector and one column per dimension (see Figure 5.4 for an example). The values of the cells are color-coded—blue for minimum values and red for maximum values (see Figure 5.2). The different dimensions can be individually normalized into the interval $[0, 1]$ if their domains are too different.

Heatmaps have many advantages compared to other visualization techniques for high-dimensional data: Large amounts of data can be clearly displayed on one page. Correlations between different dimensions and the distribution of the values of the different dimensions become visible.

5.2.3 *Principal Component Analysis Visualization*

As already described in Subsection 4.2.2, a PCA searches for a linear transformation (change of basis) so that the new basis vectors—the principal components—are ordered decreasingly according to their proportion of total variance. If the

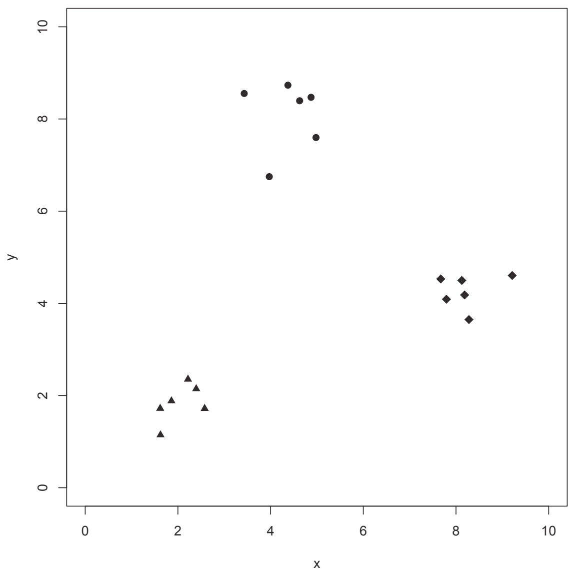


Figure 5.3: Example of a clustering with three clusters.

first few components comprise a high proportion of total variance, one can omit the remaining components (dimension reduction) without loosing much of the original information of the data set.

If the resulting dimensionally reduced data has at least three dimensions, scatter plots can be used for visualization.

Nevertheless, one has to keep in mind that this is a lossy visualization technique and that the original data was transformed into a new coordinate system.

5.3 clustering

In this section, the foundations of clustering are presented. A good overview is given by Berkhin in [12]. Vesanto and Alhoniemi present important facts in [169, pp. 586–588].

5.3.1 *Basics*

The general goal of clustering is to partition a set $X \subseteq \mathbb{R}^n$ of vectors into k disjoint subsets (clusters) $\mathcal{C} = \{C_1, \ldots, C_k\}$ with

$$X = C_1 \cup C_2 \cup \cdots \cup C_k, \quad C_i \cap C_j = \emptyset,\ i \neq j \quad . \tag{5.2}$$

In a good clustering, vectors which fall in the same cluster are quite similar—while those which fall in different clusters are quite dissimilar. Figure 5.3 gives an example of a clustering with three clusters. The three different symbols represent the membership of the vectors to one of these clusters.

There are several different clustering methods. Two often used methods in practice are hierarchical and partitive clustering. These are explained in more detail in the following two subsections.

5.3.2 *Hierarchical Clustering*

Instead of determining a single set of clusters as in the example of Figure 5.3, hierarchical clustering creates a hierarchy of clusters in form of a tree structure—the so-called *dendrogram* (see the top of Figure 5.4 for an example). Each node of this tree represents a cluster. The corresponding cluster of a node is the union of all clusters belonging to this node's child nodes. The root node represents the cluster with all vectors, the leaf nodes the clusters which contain exactly one single vector.

Hierarchical clustering can be divided into agglomerative (bottom-up) and divisive (top-down) algorithms for constructing the dendrogram. In this thesis, only agglomerative hierarchical clustering is used. The approach is described in pseudo code in Algorithm 5.1.

Algorithm 5.1 Agglomerative hierarchical clustering.

Function AGGLOMERATIVEHIERARCHICALCLUSTERING(X)

Input: set X of data vectors

Output: clustering tree (dendrogram) D

```
𝒞 ← ∅
for i = 1 to |X| do
    {initialize: assign each vector to its own cluster}
    C_i ← {x⃗_i}
    𝒞 ← 𝒞 ∪ {C_i}
end for
numberClusters ← |X|
repeat
    {compute distances between all clusters}
    for all C_i ∈ 𝒞 do
        for all C_j ∈ 𝒞 do
            if C_i ≠ C_j then
                compute distance d(C_i, C_j) between clusters C_i and C_j
            end if
        end for
    end for
    {merge the two clusters C_i and C_j that are closest to each other}
    C_i,j ← C_i ∪ C_j
    𝒞 ← 𝒞\{C_i, C_j} ∪ {C_i,j}
    numberClusters ← numberClusters −1
    {store information about two sub-clusters}
    C_i,j.child1 ← C_i
    C_i,j.child2 ← C_j
    D ← C_i,j
until numberClusters = 1
return D
```

At the beginning, each vector is assigned to its own cluster. Then, the algorithm works iteratively. In each step, those two clusters which are closest to each other are merged into a new cluster which is located at the next higher level in the dendrogram.

For the inter-cluster distance $d(C_i, C_j)$ in line 13 of Algorithm 5.1, several measures exist. Among these are

- single linkage

$$d_s(C_i, C_j) := \min_{\substack{\vec{x}_i \in C_i \\ \vec{x}_j \in C_j}} \{d(\vec{x}_i, \vec{x}_j)\} \quad , \tag{5.3}$$

- complete linkage

$$d_{co}(C_i, C_j) := \max_{\substack{\vec{x}_i \in C_i \\ \vec{x}_j \in C_j}} \{d(\vec{x}_i, \vec{x}_j)\} \quad \text{and} \tag{5.4}$$

- average linkage

$$d_a(C_i, C_j) := \frac{1}{|C_i||C_j|} \sum_{\substack{\vec{x}_i \in C_i \\ \vec{x}_j \in C_j}} d(\vec{x}_i, \vec{x}_j) \quad . \tag{5.5}$$

In each of these measures, $d(\vec{x}_i, \vec{x}_j)$ is a distance measure between the two vectors $\vec{x}_i$ and $\vec{x}_j$. This could be, for example, the Euclidean distance $\|\vec{x}\|_2$ defined as

$$\|\vec{x}\|_2 := \sqrt{\sum_{i=1}^{n} |x_i|^2} \quad , \vec{x} \in \mathbb{R}^n \quad . \tag{5.6}$$

5.3.3 *Partitive Clustering:* k*-means*

The k-means clustering algorithm is a randomized clustering approach that generates a disjoint, non-hierarchical partitioning consisting of k clusters. The algorithm is described in pseudo code in Algorithm 5.2.

It minimizes the error $E(\mathcal{C})$ with

$$E(\mathcal{C}) = \sum_{i=1}^{k} \sum_{\vec{x_j} \in C_i} \|\vec{x_j} - \vec{c_i}\|_2^2 \quad . \tag{5.7}$$

As the k-means algorithm does not depend on previously found sub-clusters, it often results in better clusterings than gained with hierarchical approaches. Yet, as it is a randomized algorithm, its execution is nondeterministic—possibly resulting in several different clusterings for the same data set X and value k. So, the question arises how to choose the number k of clusters and how to choose from the different clusterings potentially found for the same number of clusters.

Algorithm 5.2 k-means clustering.

Function KMEANS(X, k)

Input: set X of data vectors, number of clusters k

Output: clustering $\mathcal{C}$ with k clusters

```
C ← ∅
for i = 1 to k do
    C_i ← ∅
    C ← C ∪ {C_i}
    randomly initialize cluster center (centroid) c⃗_i
end for
repeat
    {compute partitioning for data}
    for i = 1 to k do
        C_i ← ∅
    end for
    for j = 1 to |X| do
        add x⃗_j to that C_i with shortest Euclidean distance between x⃗_j and c⃗_i
    end for
    {update cluster centers}
    for i = 1 to k do
        c⃗_i := 1/|C_i| Σ_{x⃗_j ∈ C_i} x⃗_j
    end for
until partitioning stays unchanged or the algorithm has converged
return C
```

In [105], Milligan and Cooper present and assess 30 procedures for determining the number of clusters of a data set. One possible solution to this problem is the Davies-Bouldin index [36] defined as

$$DB(\mathcal{C}) := \frac{1}{k}\sum_{i=1}^{k} \max_{\substack{j\in\{1,\dots,k\}\\ i\neq j}} \left\{ \frac{S_c(C_i)+S_c(C_j)}{d_{ce}(C_i,C_j)} \right\} \quad . \tag{5.8}$$

Thereby, S_c is defined as

$$S_c(C_i) := \frac{1}{|C_i|}\sum_{\vec{x}_j\in C_i} \|\vec{x}_j-\vec{c}_i\|_2 \tag{5.9}$$

and acts as a dispersion measure quantifying the average centroid distance of the cluster's vectors. The measure d_{ce} is defined as

$$d_{ce}(C_i,C_j) := \|\vec{c}_i-\vec{c}_j\|_2 \tag{5.10}$$

and quantifies the distance between two clusters (centroid linkage).

An optimal clustering consists of "compact" clusters with small dispersion and large distances between the single clusters. Looking at (5.8), one can easily notice that such an optimal clustering minimizes the value of the Davies-Bouldin index.

5.4 APPROACH FOR VISUALIZATION AND CLUSTERING OF PROCESS MODEL COLLECTIONS

Based on the results of the previous two sections, an approach for visualization and clustering of the high-dimensional process model metric data of process model collections is proposed in this section.

It is divided into three steps which are subsequently explained.

5.4.1 *Heatmap Visualization*

In the first step, a heatmap is used for depicting the process model metric values of the process models.

The process model metric values of a process model are displayed in one row. The different process model metrics form the columns of the matrix. External attributes as duration, costs, number of errors or understandability (see Subsection 3.4.2) can be added as additional columns of the heatmap if desired.

In order to get a better insight into the correlations between the different process model metrics, the columns of the heatmap can be hierarchically clustered.

5.4.2 *Principal Component Analysis Visualization*

The second step requires the application of a PCA on the process model metric data. If the resulting first three components comprise a sufficiently large proportion of total variance, the dimensionally reduced 3-dimensional data can be visualized using either a 3D scatter plot or three 2D scatter plots.

5.4.3 *Clustering*

In the third and last step, clusters of structurally similar process models within the process model collection are searched for. For that purpose, a partitive clustering algorithm is applied to the process model metric data. The results can be visualized within a heatmap again.

5.5 EXPERIMENTAL APPLICATION

In this section, the abstract approach presented in the previous Section 5.4 is applied to a set of selected process model metrics and a process model collection.

5.5.1 *Selected Process Model Metrics and Process Model Collection*

For the experimental application of the visualization and clustering approach, the same 33 EPC process model metrics and 515 EPC process models of the SAP Reference Model as in Subsection 4.3.1 were used in order to make the findings comparable with those of Chapter 4.

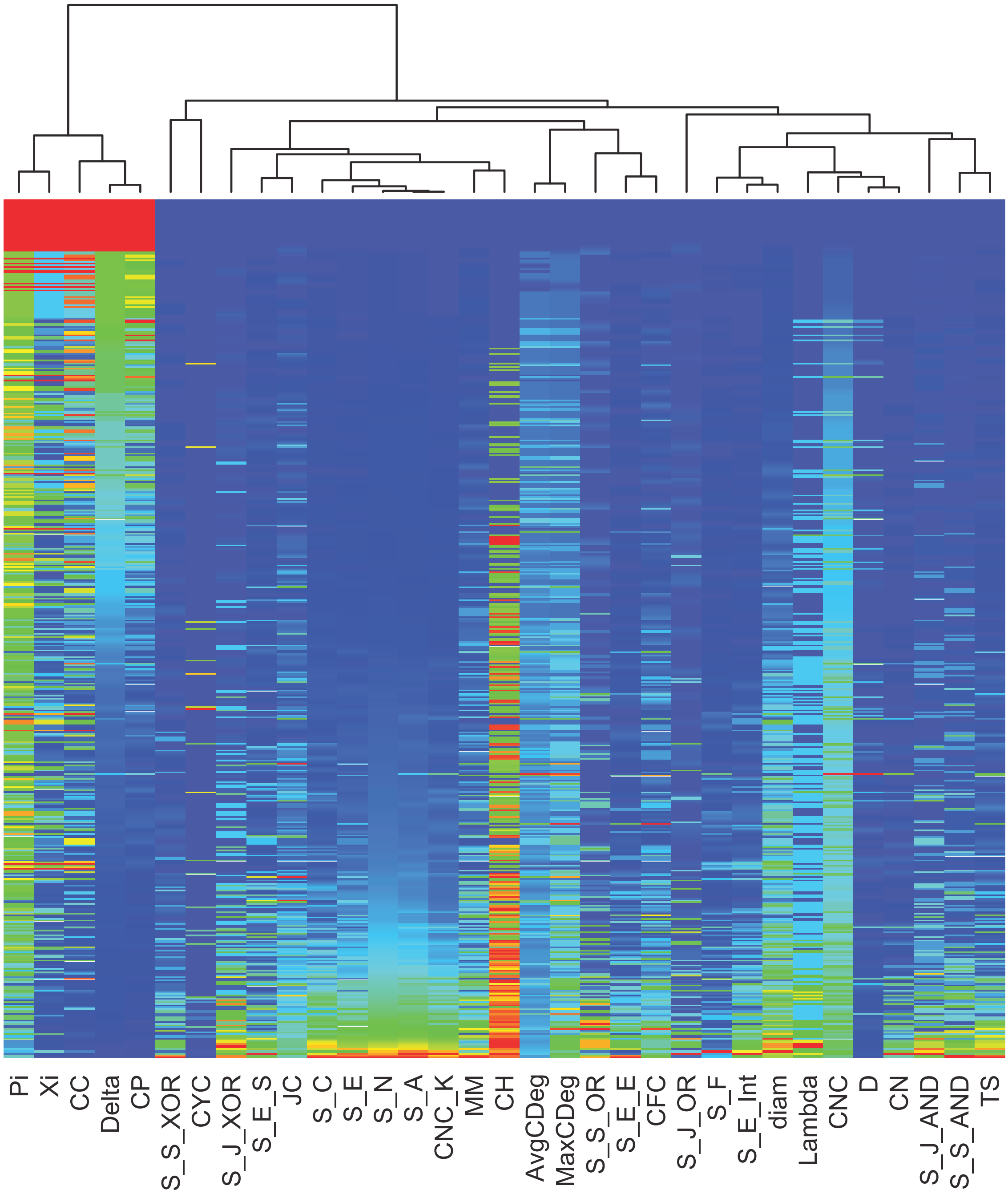


Figure 5.4: Heatmap displaying the values of the 33 selected process model metrics for the 515 selected process models. The rows (i. e., process models) are ordered by the *number of nodes* metric (S_N). The columns (i. e., process model metrics) are hierarchically clustered using (1 − Spearman's rank correlation coefficient) as distance measure.

5.5.2 *Results Concerning Heatmap Visualization*

The values of the 33 selected process model metrics for the 515 selected process models are depicted in the heatmap of Figure 5.4.

The values of each process model metric are normalized into the interval $[0, 1]$ as their domains are too different (see Table 4.4). The metrics *control flow complexity* (CFC) and *join complexity* (JC) are logarithmically normalized as both have some outliers with extremely high values compared to the large rest of the values (see Table 4.4 as well as Figure 4.6f and 4.6g).

The rows (i. e., process models) are ordered by the *number of nodes* metric (S_N). The columns (i. e., process metrics) are hierarchically clustered using $(1-$ Spearman's rank correlation coefficient$)$ (see Section C.2) as distance between two columns (i. e., process model metrics) within the complete linkage inter-cluster distance measure of (5.4).

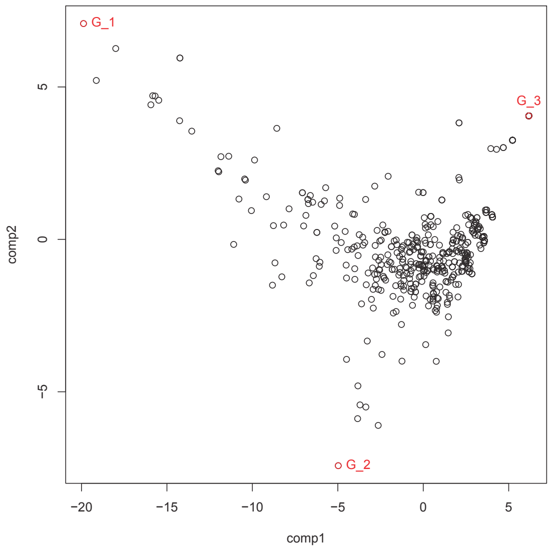
The data is clearly displayed in the heatmap on one page. So, the main goal of the visualization is fulfilled. Furthermore, several observations can be made, which are consistent with the findings of Chapter 4:

- There is a strong positive correlation between the "size metrics" *number of connectors* (S_C), *number of events* (S_E), *number of nodes* (S_N) and *number of arcs* (S_A).
- There is a negative correlation between most metrics (e. g., the "size metrics") and the metrics *separability* (Π), *sequentiality* (Ξ), *cross-connectivity* (CC), *density* (Δ) and *weighted coupling* (CP). The negative correlation is especially strong between S_C, S_E, S_N and S_A on the one hand and Δ and CP on the other.
- Most metrics have many small and only some large values. Metric *heterogeneity* (CH) shows about one third to one half very small values—the remaining values are relatively large. For the metrics *separability* (Π) and *coefficient of connectivity* (CNC), most process models have values in the middle of the domain.

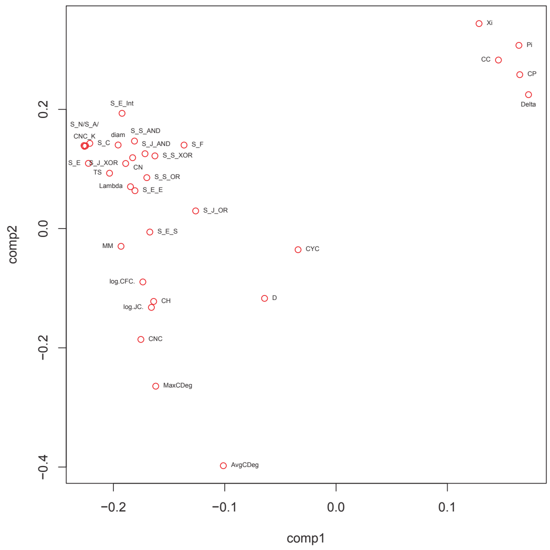
5.5.3 *Results Concerning Principal Component Analysis Visualization*

As shown in Table 4.7 of Subsection 4.3.3, the first three components of the PCA of the selected process model collection comprise more than 71% of the total data variance. So, a visualization of the scores which is restricted to these first three dimensions of the 33-dimensional data is meaningful. The results are depicted in the left column of Figure 5.5. As a 3D visualization is hard to interpret as long as it is not interactively rotatable, three 2D plots are used instead.

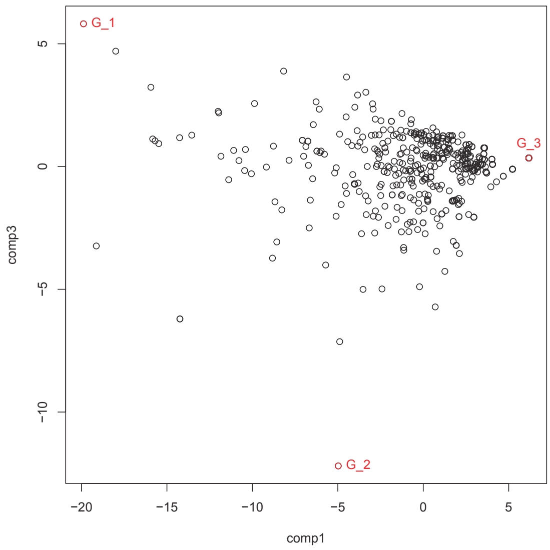
Most process models are located at the origin of the PCA coordinate system. Especially in Figure 5.5a, one can also note three branches which show to the top left (in direction of process model G_1), to the bottom (G_2) and to the top right (G_3). These findings correlate with the loadings of the 33 selected process model metrics (see right column of Figure 5.5). Process model G_1 is the EPC with the most nodes ($S_N(G_1) = 130$) and arcs ($S_A(G_1) = 138$) within the selected process model collection. Consequently, it is located in the direction of the "size metrics". G_2 as an AND-17 has the largest *average connector degree* ($\overline{d_C}$) value. So,



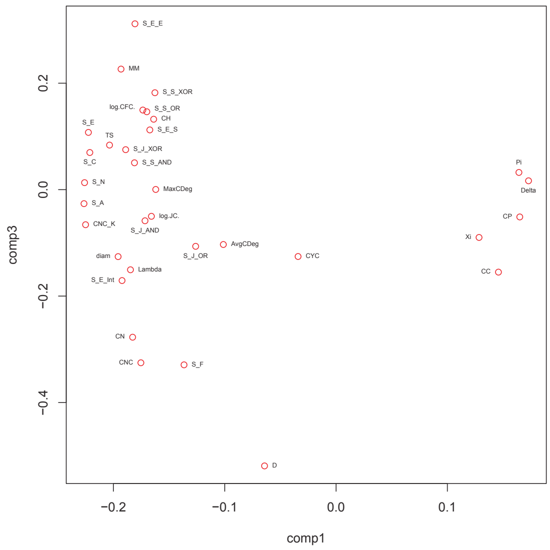
(a) comp2/comp1 scores plot.



(b) comp2/comp1 loadings plot.



(c) comp3/comp1 scores plot.



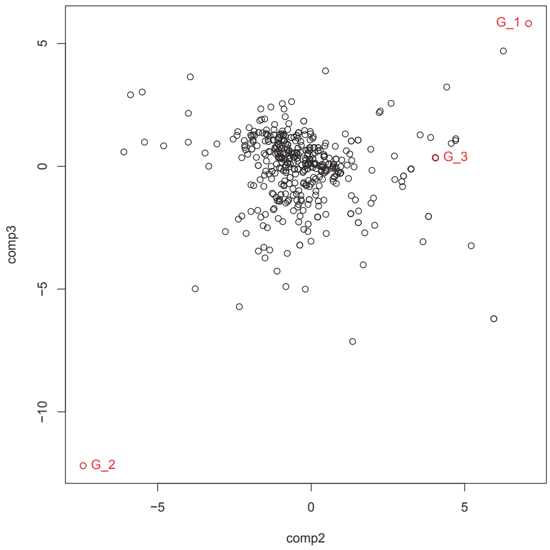
(d) comp3/comp1 loadings plot.

Figure 5.5: Scores and loadings plots of the PCA (first three components). The left column shows the scores plots of the selected process models. The right column shows the loadings plots of the selected process model metrics (as in Figure 4.12) (Part 1 of 2).

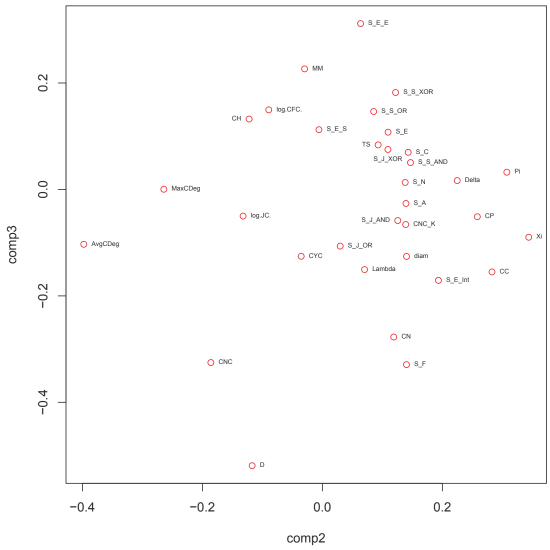
it is located at the bottom. And, finally, G_3 is a SEQ-1 whose Δ, Ξ, Π, CP and CC metric values are relatively large—resulting in a location in direction of these five metrics at the top right.

5.5.4 *Results Concerning Clustering*

A clustered version of the heatmap of Subsection 5.5.2 is depicted in Figure 5.6. The clustering was done using the k-means clustering algorithm for three clusters.



(e) comp3/comp2 scores plot.



(f) comp3/comp2 loadings plot.

Figure 5.5: Scores and loadings plots of the PCA (first three components). The left column shows the scores plots of the selected process models. The right column shows the loadings plots of the selected process model metrics (as in Figure 4.12) (Part 2 of 2).

Before clustering, the input data (normalized metric values from the non-clustered heatmap) was scaled to mean 0 and variance 1 for each dimension. The selection of the optimal number of clusters and the optimal clustering with this cluster number for the input data was done using the Davies-Bouldin index. The depicted clustering has a Davies-Bouldin index of 1.001289 based on the scaled values.

The three clusters are not that "spectacular". They simply segregate the process models into three sets with middle (top), large (center) and small (bottom) metric values for the first five metrics. If one looks at the results of the PCA visualization in Subsection 5.5.3, this observation is not that surprising. In the left column of Figure 5.5, also no clear cluster structure of the scores is identifiable.

5.6 Conclusion

In this chapter, an approach for visualization and clustering of high-dimensional process model metric data of process model collections was proposed.

First, different visualization techniques were examined for their suitability for visualizing many high-dimensional data points. Next, basic clustering methods were presented.

The approach comprises

1. a compact heatmap visualization of the metric data,
2. a 3D scatter plot visualization of the outcome of a PCA of the data and

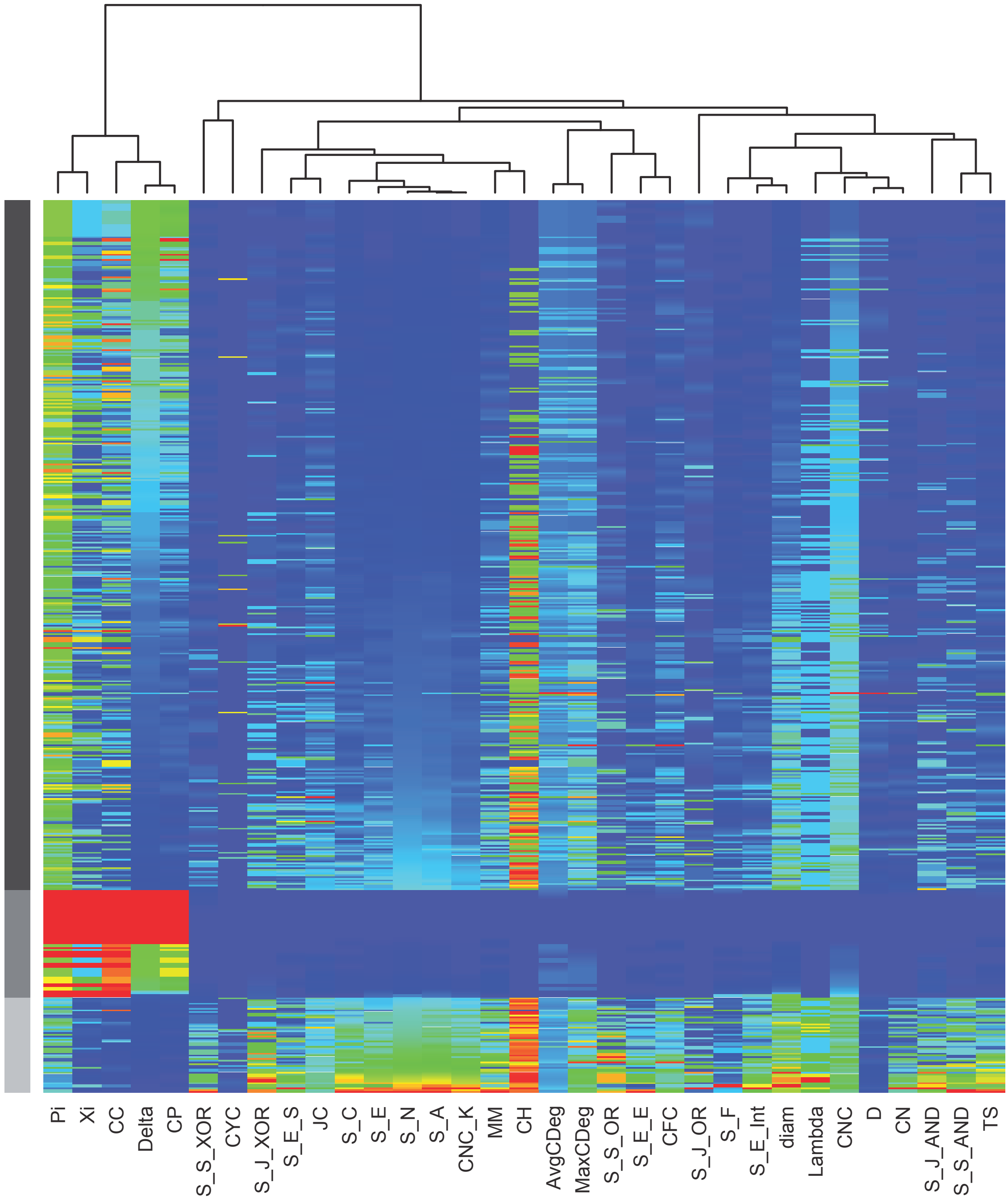


Figure 5.6: Clustered heatmap displaying the values of the 33 selected process model metrics for the 515 selected process models. The rows (i. e., process models) are separated into three clusters (see bar with gray scale values at the left). The columns (i. e., process model metrics) are hierarchically clustered using $(1 - \text{Spearman's rank correlation coefficient})$ as distance measure.

3. a clustered heatmap visualization where the metric data is clustered for (structurally) similar process models within a process model collection.

The approach was successfully applied to the same EPC process model metrics and process models as in Chapter 4.

It could be demonstrated that the visualization of 33 process model metric values for 515 process models using heatmaps is possible and still clear for a human observer. Furthermore, the findings on the correlations between process model metrics and on their value ranges which could be gained visually are also consistent with the statistical results of Chapter 4.

Using the results of a PCA of the process model metric data, it was possible to visualize the data within the three-dimensional coordinate system induced by the first three components of the PCA (comprising more than 71% of the total data variance).

In contrast, the three clusters of structurally similar process models which were found were not that "spectacular". It should be examined in future work whether other process model collections have more interesting clusterings.

6 MEASURING STRUCTURAL PROCESS MODEL UNDERSTANDABILITY

6.1 INTRODUCTION

Following the process measurement approach of Subsection 3.4.2, the values of external attributes of process models are predicted based on the values of internal attributes.

One very important external attribute is process model understandability of involved humans (e. g., process designers, process analysts, process implementers or people executing a process). Understandability influences other quality aspects of process models like error-proneness and maintainability. Even though the importance of understandability is undoubted, Mendling *et al.* state that "we know surprisingly little about the act of modeling and which factors contribute to a 'good' process model in terms of human understandability" [101, p. 48].

Some published studies try to examine the dependencies between some influencing factors and process model understandability: In [138, 139], Sarshar *et al.* compare the understandability of different process modeling languages. Recker and Dreiling examine whether somebody's experience with one process modeling language can be helpful for understanding process models based on another modeling language he/she is not familiar with [125]. Mendling *et al.* search for dependencies between personal and process model specific (structural) properties and process model understandability [101]. In [102], Mendling and Strembeck also examine the influence of content related factors on process model understandability. Reijers and Mendling test the effect of process model modularization on process model understandability [126].

One can distinguish at least two aspects of process model understandability—structural and semantic process model understandability. *Structural* process model understandability entirely abstracts from a process's real goal (e. g., an insurance claim process). Instead, only the understandability relating to its structure (i. e., the process model is only seen as a graph with nodes and edges) is considered. Here, questions like order of activities, number of times an activity can be executed, possible parallelism between activity execution, etc. are of interest. *Semantic* process model understandability—on the other hand—also considers what is actually done during the process model's activities, how these activities are interrelated and how they contribute to the process's final goal.

Only the measurement of structural process model understandability is examined in this chapter (i. e., finding a valid measuring system as defined in Subsection 3.4.1). The analysis of possible influencing factors on structural process model understandability (i. e., possible prediction systems) is beyond the scope of this thesis.

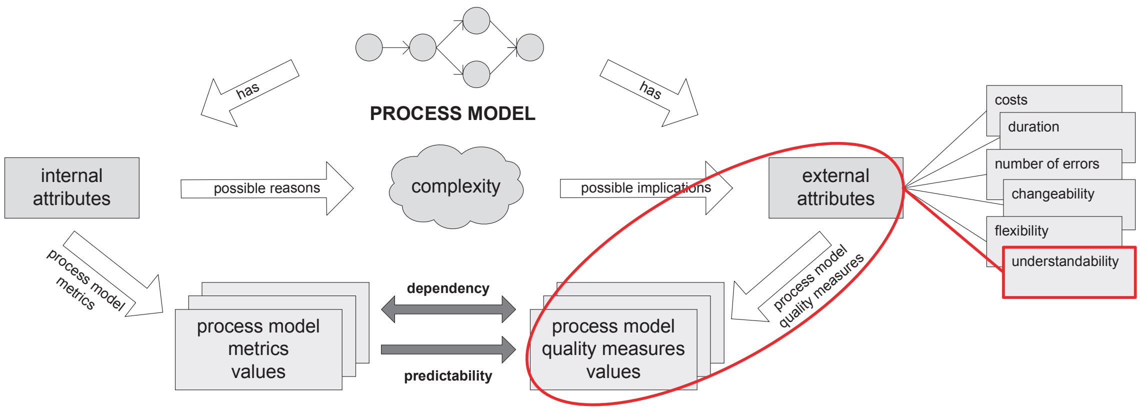


Figure 6.1: Chapter visually located within the measurement approach of Subsection 3.4.2.

For examining structural process model understandability and validating appropriate prediction systems (as done in the above mentioned studies), one first has to quantify structural process model understandability. Thus, a proper measure for structural process model understandability which fulfills the reliability and validity requirements for measuring systems (see Subsection 3.4.3) has to be found. Looking at the few proposed measures for structural process model understandability, serious doubts about this necessary validation arise.

In this chapter, concrete and detailed definitions for measuring structural process model understandability are given which exceed those in existing publications. Using these definitions, hypotheses about effects of measuring structural process model understandability are formulated which have to be considered in the measuring process. Finally, an experimental evaluation is conducted to examine these hypotheses.

Figure 6.1 shows where the chapter is visually located within the measurement approach of Subsection 3.4.2.

The remainder of this chapter is organized as follows: In Section 6.2, existing approaches for measuring structural process model understandability are presented and some major points of criticism on these approaches are introduced. A general framework for the evaluation of modeling technique understanding, which is used as a basis of an own measurement approach, is shown in Section 6.3. This own measurement approach for measuring structural process model understandability and related hypotheses about some effects of measurement are given in Section 6.4. Section 6.5 shows the results of an experimental evaluation of this measurement approach and the related hypotheses. The chapter closes with a conclusion (Section 6.6).

6.2 RELATED WORK

In this section, related work on measuring structural process model understandability is presented and discussed. Subsection 6.2.1 shows existing measurement approaches. Criticism on these approaches is given in Subsection 6.2.2.

6.2.1 *Existing Approaches*

Structural process model understandability is a non-physical property as described in Section A.3. Thus, a concrete operationalization (see Definition A.13) has to be found in order to measure this property. In the literature, one can find some propositions for such an operationalization.

In [138, 139], Sarshar *et al.* examined the influence of different process modeling languages (EPCs vs. Petri nets) on structural process model understandability. They selected a process model which was depicted in an EPC as well as a Petri net version (both with more than 80 nodes). For both versions, a questionnaire with ten multiple-choice questions was created. 50 students participated in the experiment. They were randomly assigned either to the EPC or the Petri net group. The number of correct answers to the process model related questions served as operationalization of structural process model understandability.

In [101], Mendling *et al.* searched for possible dependencies between personal and process model specific (structural) properties and structural process model understandability. They used a questionnaire which was answered by 73 students who had followed courses on process modeling. For the questionnaire, they selected 12 process models (each with 25 tasks). The process models were depicted in a simplified EPC-like notation (without events) in a top-to-bottom-style. The tasks were just labeled with capital letters. As operationalization of structural process model understandability, Mendling *et al.* created the SCORE measure: Each student had to answer eight closed questions on order, concurrency, exclusiveness or repetition of tasks as well as one open question on possible errors for each process model. The sum of correct answers (at most nine) gives the SCORE value.

In addition to the goals of [101], Mendling and Strembeck also examined the influence of content related factors on structural process model understandability in [102]. For that purpose, they designed an online questionnaire which was answered by 42 students and practitioners. Six process models with an equal number of tasks—each in two variants (one with tasks labeled with capital letters and a second one with tasks labeled with normal descriptive text)—were selected. The process models were depicted in the same notation as in [101]. For each process model, six yes/no questions on process model structure and behavior were chosen. The subjects of the experiment were randomly assigned to one of two questionnaire variants (capital letter labels and text labels). The measure PSCORE was calculated as the sum of correctly answered questions on the six process models (at most 36) and served as an operationalization of structural process model understandability related to a person. The measure MSCORE—on the other hand—was calculated as the sum of correct answers from all participants to one process model. It served as an operationalization of structural process model understandability related to a process model.

In [126], Reijers and Mendling analyzed the influence of process model modularity on structural process model understandability. For that purpose, they selected two large process models A and B (with 105 and 120 tasks respectively in the flattened version) and additionally constructed a modularized version for both. A questionnaire with 12 questions for each of the four process models (A/B and flattened/modularized) was used. The questionnaire was answered by 28 experienced consultants. The percentage of correctly answered questions given by a subject was used as a measure for his/her level of structural process model understandability.

6.2.2 *Criticism on Existing Approaches*

The proposed operationalizations for structural process model understandability are intended to serve as measurement systems as defined in Subsection 3.4.1. Consequently, they have to fulfill the validation requirements reliability and validity of Subsection 3.4.3 and—with more details— Subsection A.3.2.

Looking at the proposed measures, large doubts arise whether these requirements are really fulfilled.

Content Validity

Content validity is concerned with whether a measure covers the range of meanings included in the underlying concept (see Subsection 3.4.3 and A.3.2).

Sarshar *et al.* ask questions on the states before and after special events (EPC) or transitions (Petri net) are reached [139, pp. 32–34, 39–41]. Mendling *et al.* name four aspects of structural process model understandability: understanding of order, concurrency, exclusiveness and repetition [101, p. 52]. In [102, p. 146], Mendling and Strembeck ask questions on choices, concurrency, loops and deadlocks. But these are not used to compute the MSCORE measure for process models. For [126], the asked questions are not given. It is only stated [126, p. 26] that the whole measurement approach is similar to that of [101].

Looking at these publications, some questions arise: Do other important aspects of structural process model understandability exist? How different is the understanding based on the different aspects? How can "overall structural process model understandability" be computed?

Reliability

Reliability requires that measure values obtained by different observers of the same process model have to be consistent (see Subsection 3.4.3 and A.3.2).

In [138, 139], Sarshar *et al.* ask only ten questions per process model version (EPC or Petri net)—even though these process models have more than 80 nodes. In [101, 102], only eight and six questions per process model are asked, respectively. And in [101], these questions are even distributed to four different aspects. Reijers and Mendling ask 12 questions on process models with more than 100 tasks in [126]. In all mentioned publications, it does not become clear how the nodes (tasks, functions, events or transitions depending on the used process modeling language) involved in the questions are selected.

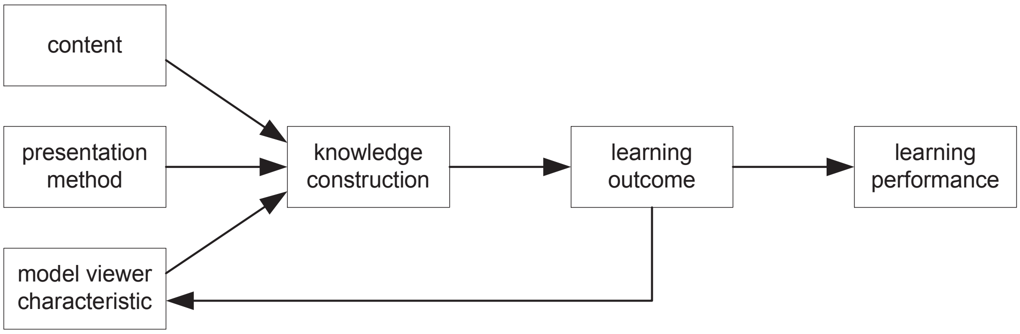


Figure 6.2: "Reading" a model as a form of knowledge construction [51, p. 82].

So, the question arises whether this selection is representative for the process model. It is possible that complicated parts of the process model have been omitted or only especially complex parts have been selected. This selection could have a big influence on the measured values.

6.3 Framework for evaluating modeling technique understanding

In this section, the framework for evaluating modeling technique understanding by Gemino and Wand [51] is presented. As it is applicable to arbitrary modeling techniques—including process modeling techniques, it can serve as a basis for developing an own measurement system for structural process model understandability in this chapter.

According to Gemino and Wand, (graphical) modeling techniques are used to communicate about the specification and the application domain of an information system in order to improve the common understanding among the involved stakeholders. In their view, "requirements development can be viewed as a process of accumulating valid information and communicating it clearly to others. This makes the process of requirements development analogous to the process of learning [. . .]." [51, p. 80]

Consequently, they use Mayer's framework of learning [84] as one basis for their own evaluation framework [51, p. 82].

Gemino and Wand distinguish two tasks when using modeling techniques: "writing" a model (creating a model to represent parts of the real world) and "reading" a model (creating a mental representation from a model) [51, p. 80]. In this chapter, only the second task is of interest.

For evaluating the viewer's understanding when "reading" a model, Gemino and Wand suggest to see "reading" (i. e., interpreting and understanding) a model as a form of knowledge construction (as depicted in Figure 6.2).

The starting points of this knowledge construction are

- the content (part of the real world represented in the model),
- the presentation method (used modeling technique) and

- the model viewer characteristic (attributes of the viewer before looking at the model, including knowledge of and experience with the domain and the modeling technique).

They influence the knowledge construction and consequently the learning outcome which—again—changes the model viewer characteristic. This cognitive process is not directly observable, but has to be observed indirectly through learning performance tasks. [51, pp. 82–83]

Gemino and Wand list two types of such tasks: comprehension and problem-solving tasks. The former include questions regarding attributes of and relationships between model items—while the latter include questions going beyond the information given originally in the model. [51, p. 83]

For measuring *semantic* process model understandability, problem-solving tasks could be a good way. One could conduct experiments in which process models have to be modified/corrected—similar to the field of software engineering where source code has to be modified or errors have to be found and corrected. Yet, as mentioned in the introduction, measuring *semantic* process model understandability is not topic of this thesis.

Comprehension tasks—on the other hand—seem to be useful for measuring *structural* process model understandability in this chapter. The questions asked in [101, 102, 126, 138, 139] fall into this category.

Finally, Gemino and Wand point out that one has to control some of the influencing factors content, presentation method and model viewer characteristic when studying the others [51, p. 83].

Thus, for measuring structural process model understandability in this chapter, one has to decide which of the three influencing factors content (i. e., process model), presentation method (i. e., process modeling language) and model viewer characteristic (e. g., knowledge of process domain and process modeling language) one is interest in.

6.4 APPROACH FOR MEASURING STRUCTURAL PROCESS MODEL UNDERSTANDABILITY

In this section, an own approach for measuring structural process model understandability is introduced. It tries to overcome the existing measures' potential problems with reliability and validity which were identified in Subsection 6.2.2. At the same time, it is as similar to the existing measurement approaches of Subsection 6.2.1 as possible so that hypotheses about the existing approaches' potential problems can be formulated and examined in subsequent experiments (see Section 6.5).

The approach is based on Gemino and Wand's framework for evaluating modeling technique understanding which was presented in the previous Section 6.3. The understandability is measured using comprehension tasks—consistent with the existing approaches. The major advantages of this approach are:

- The approach systematically creates the asked questions ("comprehension tasks" in Gemino and Wand's framework) for all possible process models

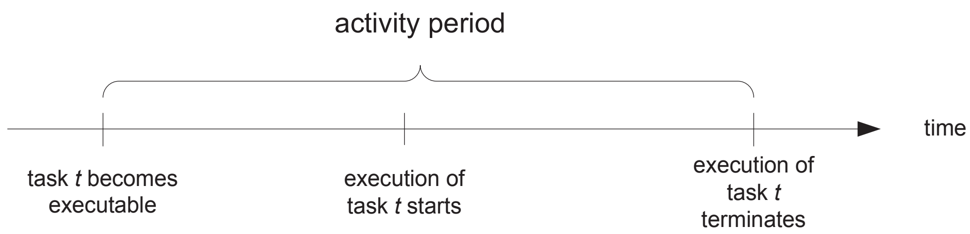


Figure 6.3: Activity period.

without human interaction. So, the potential question selection problem is avoided. Consequently, the retrieved numbers for different process models should really be comparable.

In order to make this practicable, only process models using a special process modeling language ("presentation method" in Gemino and Wand's framework) are allowed. The same EPC-like notation without events as in [101] is used.

- Furthermore, the approach tries to measure understandability, which is a subjective property of a viewer (cf. "model viewer characteristic" in Gemino and Wand's framework), as objective as possible.

In Subsection 6.4.1, different aspects of structural process model understandability are given. Furthermore, the generic questions used as comprehension tasks are defined. Subjective and objective measures for structural process model understandability are introduced in Subsection 6.4.2. Subsection 6.4.3 and 6.4.4 propose two techniques for reducing the number of asked questions.

6.4.1 *Aspects of Structural Process Model Understandability*

As already discussed in Subsection 6.2.2, it is important to cover the different aspects of structural process model understandability to fulfill the content validity criteria for measures. In this chapter, the aspects *concurrency*, *exclusiveness*, *order* and *repetition* identified by Mendling *et al.* in [101, p. 52] are used. In doing so, the possible existence of other aspects is not denied. Unlike in [101], detailed definitions of the questions for the different aspects are given.

First, the term "activity period", which is later used in questions, is defined.

Definition 6.1 (Activity period) *An activity period of task* t *is the period between a point in time when* t *becomes executable and the next point in time when the actual execution of* t *terminates (see Figure 6.3).*

Now, relations for the four aspects of structural process model understandability can be defined.

Definition 6.2 (Concurrency) *For the questions on task concurrency, the relations* $c_{\nexists}, c_{\exists}, c_{\forall} \subseteq T \times T$ *with the following meanings are used.*

$(t_1, t_2) \in c_{\nexists} \Leftrightarrow$ *There is* no *process instance for which the activity periods of tasks* t_1 *and* t_2 *overlap.*

$(t_1, t_2) \in c_{\exists} \Leftrightarrow$ *There is* a *process instance for which the activity periods of tasks* t_1 *and* t_2 *overlap at least once (Several executions of* t_1 *and* t_2 *per process instance are possible!).—But there also exists a process instance for which this does not hold.*

$(t_1, t_2) \in c_{\forall} \Leftrightarrow$ *For* each *process instance, the activity periods of tasks* t_1 *and* t_2 *overlap at least once.*

Definition 6.3 (Exclusiveness) *For the questions on task exclusiveness, the relations* $e_{\nexists}, e_{\exists}, e_{\forall} \subseteq T \times T$ *with the following meanings are used.*

$(t_1, t_2) \in e_{\nexists} \Leftrightarrow$ *There is* no *process instance, for which tasks* t_1 *and* t_2 *are both executed.*

$(t_1, t_2) \in e_{\exists} \Leftrightarrow$ *There is* a *process instance, for which tasks* t_1 *and* t_2 *are both executed.—But there also exists a process instance for which this does not hold.*

$(t_1, t_2) \in e_{\forall} \Leftrightarrow$ *For* each *process instance, the tasks* t_1 *and* t_2 *are both executed.*

Definition 6.4 (Order) *For the questions on task order, the relations* $o_{\nexists}, o_{\exists}, o_{\forall} \subseteq T \times T$ *with the following meanings are used.*

$(t_1, t_2) \in o_{\nexists} \Leftrightarrow$ *There is* no *process instance for which an activity period of task* t_1 *ends before an activity period of task* t_2 *starts.*

$(t_1, t_2) \in o_{\exists} \Leftrightarrow$ *There is* a *process instance for which an activity period of task* t_1 *ends before an activity period of task* t_2 *starts.—But there also exists a process instance for which this does not hold.*

$(t_1, t_2) \in o_{\forall} \Leftrightarrow$ *For* each *process instance, an activity period of task* t_1 *ends before an activity period of task* t_2 *starts.*

Definition 6.5 (Repetition) *For the questions on task repetition, the relations* $r_{=1}, r_{?}, r_{*}, r_{+} \subseteq T$ *with the following meanings are used.*

$t \in r_{=1} \Leftrightarrow$ *For* each *process instance, task* t *is executed* exactly once.

$t \in r_{?} \Leftrightarrow$ *For* each *process instance, task* t *is executed* not once or exactly once. *Both cases really occur.*

$t \in r_{*} \Leftrightarrow$ *For* each *process instance, task* t *is executed* not once, exactly once or more than once. *There exists a process instance for which* t *is executed not once and another one for which* t *is executed more than once.*

$t \in r_{+} \Leftrightarrow$ *For* each *process instance, task* t *is executed* at least once. *There exists a process instance for which* t *is executed more than once.*

At first glance, the definitions of the relations might look a little complicated. But they are constructed in such a way that the properties of Theorem 6.1 hold, which is beneficial for the measurement process.

Theorem 6.1 (Properties of relations) *The relations have the following properties:*

1. *The relations* $c_{\nexists}$, $c_{\exists}$, $c_{\forall}$ *and* $e_{\nexists}$, $e_{\exists}$, $e_{\forall}$ *are symmetric.*

2. *For all possible task combinations, exactly one relation per aspect is true.*

Proof. The two prepositions are proved separately.

Regarding 1.) Within the definitions of the relations $c_{\nexists}$, $c_{\exists}$, $c_{\forall}$, $e_{\nexists}$, $e_{\exists}$ and $e_{\forall}$, t_1 and t_2 can be interchanged without changing the meaning.

Regarding 2.) Looking at the definitions, this can easily be seen. ∎

Because of property 2 of Theorem 6.1, the different relations for an aspect can be grouped to questions on the process model: The question $q_r(t)$, for example, asks which of the relations $r_{=1}, r_?, r_*, r_+$ holds for task t. Because of property 1 of Theorem 6.1, $q_c(t_1, t_2) = q_c(t_2, t_1)$ and $q_e(t_1, t_2) = q_e(t_2, t_1)$ hold.

Theorem 6.2 (Maximum number of questions) *The maximum number* $|Q_{a,max}(p)|$ *of possible different questions for aspect* $a \in \{c, e, o, r\}$ *on a process model* p *with* n *tasks is*

$$|Q_{c,max}(p)| = |Q_{e,max}(p)| = \frac{n(n-1)}{2} \tag{6.1}$$

$$|Q_{o,max}(p)| = n(n-1) \tag{6.2}$$

$$|Q_{r,max}(p)| = n \quad . \tag{6.3}$$

Proof. The three prepositions are proved separately.

Regarding (6.1): The questions on the aspects *concurrency* and *exclusiveness* each depend on two tasks. As there are n tasks per process model p, $n(n-1)$ different pairs of tasks exist. According to property 1 of Theorem 6.1, the relations for both aspects are symmetric. So, for example, the questions with task order t_2 and t_1 can be omitted if the questions with the reverse task order t_1 and t_2 are asked. Consequently, $|Q_{c,max}(p)| = |Q_{e,max}(p)| = \frac{n(n-1)}{2}$.

Regarding (6.2): The questions on aspect *order* each depend on two tasks. As there are n tasks per process model p, $n(n-1)$ different pairs of tasks exist. Consequently, $|Q_{o,max}(p)| = n(n-1)$.

Regarding (6.3): The questions on aspect *repetition* each depend on one task. As there are n tasks per process model p, $|Q_{r,max}(p)| = n$. ∎

As one can see, the maximum number of questions for *concurrency*, *exclusiveness* and *order* grows quadratically with the number of tasks, while the maximum number of questions for *repetition* grows only linearly.

6.4.2 *Structural Process Model Understandability*

Based on the questions on the four aspects of understandability presented in Subsection 6.4.1, measures for structural process model understandability can now be defined.

In a first step, the understanding of a single subject (personal structural process model understandability) is measured by systematically asking him/her all possible questions for one aspect of the process model in question.

Definition 6.6 (Personal structural process model understandability) *The personal structural process model understandability* $U_a(p, s)$ *for aspect* a *of process model* p *by subject* s *is defined as the fraction of correct answers given by* s *to the* $|Q_{a,max}(p)|$ *different questions for aspect* a *about* p.

$$U_a(p, s) := \frac{\textit{\# correct answers to } Q_{a,max}(p)}{|Q_{a,max}(p)|} \quad , a \in \{c, e, o, r\} \tag{6.4}$$

As persons differ in their knowledge, experience and capabilities, one can expect to measure different values of personal structural process model understandability for different persons. As many human properties (e. g., body height, weight and IQ value) are approximately normally distributed, a similar behavior can be assumed here.

Hypothesis 6.1 *The personal structural process model understandability measure values* $U_a(p, s_i)$ *of a process model* p *are normally distributed.*

In a second step, one is interested in a less "subjective" but more "objective" quantification of understandability. The different values of personal structural process model understandability can be seen as outcomes of a random variable. The expected value of this variable can be estimated according to Definition 6.7. The resulting value can be used as the desired "objective" quantification.

Definition 6.7 (Estimated structural process model understandability) *The estimated structural process model understandability* $\widehat{U}_a(p, S)$ *for aspect* a *of process model* p *and set* S *of subjects is defined as the average personal structural process model understandability of* p *by the subjects of* S.

$$\widehat{U}_a(p, S) := \frac{1}{|S|} \sum_{s \in S} U_a(p, s) \quad , a \in \{c, e, o, r\} \tag{6.5}$$

Additionally, confidence intervals for the true expected values of the random variables for the different aspects of structural process model understandability can be computed. The width of these intervals will decrease for higher numbers of subjects—meanwhile, the certainty of the true expected value will increase.

It can be expected that the different aspects *concurrency*, *exclusiveness*, *order* and *repetition* of structural process model understandability are of varying difficulty. Consequently, it is important to measure all aspects to get "overall understandability".

Hypothesis 6.2 *The different aspects of structural process model understandability result in different values of the* $\widehat{U}_a(p, S)$ *of a process model* p.

The measures proposed in this subsection require that all possible questions on an aspect are asked to each subject. As Theorem 6.2 shows, this number of questions grows quadratically for the aspects *concurrency*, *exclusiveness* and *order*. Even for practically relevant process models with, for example, ten tasks, each subject would have to answer 45 questions on *concurrency* and *exclusiveness* respectively as well as 90 questions on *order*. In the following subsections 6.4.3 and 6.4.4, two possible solutions for this problem are suggested.

6.4.3 *Partial Structural Process Model Understandability*

A first possibility to reduce the effort for measuring structural process model understandability is to select only a subset of all possible questions on the different aspects for being answered by each subject. This approach was also used in [101, 102].

Definition 6.8 (Personal partial structural process model understandability) *The personal partial structural process model understandability* $U_a(p, s, Q_a)$ *for aspect* a*, process model* p*, subject* s *and questions* $Q_a \subseteq Q_{a,max}(p)$ *is defined as the fraction of correct answers given by* s *to the questions* Q_a *for aspect* a *on* p*.*

$$U_a(p, s, Q_a) := \frac{\#\ \textit{correct answers to}\ Q_a}{|Q_a|} \quad , a \in \{c, e, o, r\} \tag{6.6}$$

Here again, the different values of personal partial structural process understandability can be seen as outcomes of a random variable. The expected value of this variable can be estimated according to Definition 6.9.

Definition 6.9 (Estimated partial structural process model understandability) *The estimated partial structural process model understandability* $\widehat{U}_a(p, S, Q_a)$ *for aspect* a*, process model* p*, set* S *of subjects and questions* Q_a *is defined as the average personal partial structural process model understandability of* p *and* Q_a *by the subjects of* S*.*

$$\widehat{U}_a(p, S, Q_a) := \frac{1}{|S|} \sum_{s \in S} U_a(p, s, Q_a) \quad , a \in \{c, e, o, r\} \tag{6.7}$$

In order to measure the number of actually asked questions Q_a relative to the number of possible questions $Q_{a,max}(p)$ on a process model p, one can define the term *coverage rate*.

Definition 6.10 (Coverage rate) *The coverage rate of a set of questions* $Q_a \subseteq Q_{a,max}(p)$ *on aspect* a *of process model* p *is defined as*

$$r_a(Q_a, p) := \frac{|Q_a|}{|Q_{a,max}(p)|} \quad , a \in \{c, e, o, r\} \quad . \tag{6.8}$$

There are several possibilities to select a certain number of questions (i. e., equal coverage rate).

Theorem 6.3 *The number of different sets of questions* $Q_a \subseteq Q_{a,max}(p)$ *with* $|Q_a| = m$ *questions is*

$$\binom{|Q_{a,max}(p)|}{m} = \frac{|Q_{a,max}(p)|!}{m!\,(|Q_{a,max}(p)| - m)!} \quad . \tag{6.9}$$

Proof. According to the rules of combinatorics, the number of ways of choosing a set of m symbols from a set of n distinct symbols without repetition is $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ [150, p. 86]. Here, $n := |Q_{a,max}(p)|$. ■

One can assume that the different questions are not equally difficult to be answered. This would have some important implications.

Hypothesis 6.3 *The different questions of* $Q_{a,max}(p)$ *are not equally difficult. This has two consequences:*

1. *For the same coverage rate, one gets different values for estimated partial structural process model understandability depending on the selected questions* Q_a*.*
2. *The smaller the coverage rate, the bigger the standard deviation of the different values of estimated partial structural process model understandability for that coverage rate.*

As a consequence, the coverage rate should not be selected too small. Furthermore, the questions for the set Q_a should be chosen randomly in order to minimize the risk of intentionally or unintentionally selecting especially easy or difficult questions when done by a human. The two recommendations shall assure that the estimated partial structural process model understandability does not differ much from the true value of structural process model understandability.

6.4.4 *Virtual Subjects' Structural Process Model Understandability*

Besides using partial structural process model understandability (Subsection 6.4.3), there is a second possibility to reduce the number of asked questions per subject—virtual subjects. This approach is based on the following hypothesis.

Hypothesis 6.4 *Randomly dividing a set of questions answered by a group of subjects into two subsets of approximately the same size results in a strong correlation between the rates of correct answers given by the same subject to the questions of the two subsets.*

Roughly speaking this means that a subject with good results for one subset of questions will also be good for the second subset. This is used in inverse direction in order to "construct" new virtual subject's answers out of the answers given by several real subjects.

The set of all possible questions for one aspect is divided into different subsets which are each answered by different groups of subjects. Afterwards, in each group the subjects are ordered by their personal partial structural process understandability values. Now, new virtual subjects are "created" by combining the answers of one subject from each group. For this step, the best subjects from each group are combined to the best new virtual subject, the second best subjects to the second best new virtual subject—and so on.

Using the answers of these so "constructed" virtual subjects, (virtual) personal structural process model understandability and (virtual) estimated structural process model understandability can be computed as defined in Definition 6.6 and 6.7 respectively.

6.5 experimental evaluation

Besides an own approach for measuring structural process model understandability, also some hypotheses about effects of measuring were formulated in the previous Section 6.4. If they were really true, they would have to be considered in the measuring process.

In this section, the approach's applicability and the postulated hypotheses are experimentally examined. For that purpose, two experiments were conducted.

1. Subsection 6.5.1 presents a rather small experiment. Here, the subjective and objective measures for structural process model understandability introduced in Subsection 6.4.2 as well as the partial structural process model understandability approach of Subsection 6.4.3 are examined.

2. A larger experiment which was also conducted as a kind of re-test is shown in Subsection 6.5.2. Here, also the virtual subjects approach of Subsection 6.4.4 is studied.

Both subsections are equally structured: First, the experiment design is explained (see Subsection B.3.2 for an explanation of the used terminology). Then, the experiment's results are given and analyzed. Finally, a validity evaluation (see paragraph *Validity Evaluation* in Subsection B.3.3) is carried out.

6.5.1 *Experiment 1*

Experiment Design

For a first experiment, Hypothesis 6.1 (normally distributed personal structural process model understandability values), 6.2 (different estimated structural process model understandability values for different aspects) and 6.3 (influences of coverage rate on estimated partial structural process model understandability) were selected for examination.

Consequently, a rather small process model had to be chosen in order to be able to let one single person answer all possible questions (for one aspect). So,

- personal structural process model understandability values could be computed from a subject's answers,

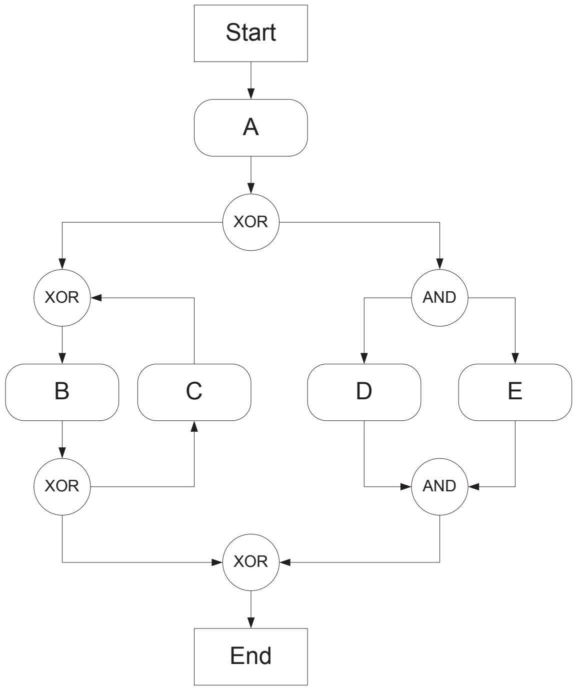


Figure 6.4: Process model used in experiment 1.

- estimated structural process model understandability values for the four aspects could be calculated as averages of the corresponding personal values and
- partial structural process model understandability values could be derived by considering only parts of the given answers.

OBJECT Finally, the process model depicted in Figure 6.4 was selected. It was presented to the subjects in the same top-to-bottom-style EPC-like notation as in [101, 102].

As the process model has only five tasks, all $|Q_{c,max}(p)| = |Q_{e,max}(p)| = 10$, $|Q_{o,max}(p)| = 20$ and $|Q_{r,max}(p)| = 5$ possible questions for the four aspects (cf. Theorem 6.2) could be asked to single subjects.

MEASUREMENT INSTRUMENTATION A questionnaire for the two groups A and B was created. The questionnaire for group A consisted of the 20 questions on *order* and the five questions on *repetition* (25 questions in total)—that of group B consisted of the ten questions on *concurrency* and the ten questions on *exclusiveness* (20 questions in total).

SUBJECTS Students attending the "Workflow Management" lecture at the *Universität Karlsruhe (TH)* were asked to participate in the experiment. Participation was voluntary. Finally, 18 students took part in the experiment. They were randomly assigned to one of the two questionnaire groups A and B—resulting in nine subjects per group.

Table 6.1: Questionnaire for experiment 1.

	group A	group B	total
# questions *concurrency*	–	10	10
# questions *exclusiveness*	–	10	10
# questions *order*	20	–	10
# questions *repetition*	5	–	10
# asked questions	25	20	45
# subjects	9	9	18

Table 6.2: Answers given for aspect *concurrency*.

subject	$q_c(A,B)$	$q_c(A,C)$	$q_c(A,D)$	$q_c(A,E)$	$q_c(B,C)$	$q_c(B,D)$	$q_c(B,E)$	$q_c(C,D)$	$q_c(C,E)$	$q_c(D,E)$	$U_c(p,s)$
solution	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	
s_2	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	0.9
s_4	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	1.0
s_6	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	1.0
s_{34}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	1.0
s_{42}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	0.9
s_{50}	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	0.3
s_{52}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	-	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	0.8
s_{56}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	1.0
s_{60}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	1.0
correct	100%	89%	100%	100%	78%	78%	89%	89%	89%	67%	

An overview of the questionnaire's structure and the involved subjects is given in Table 6.1.

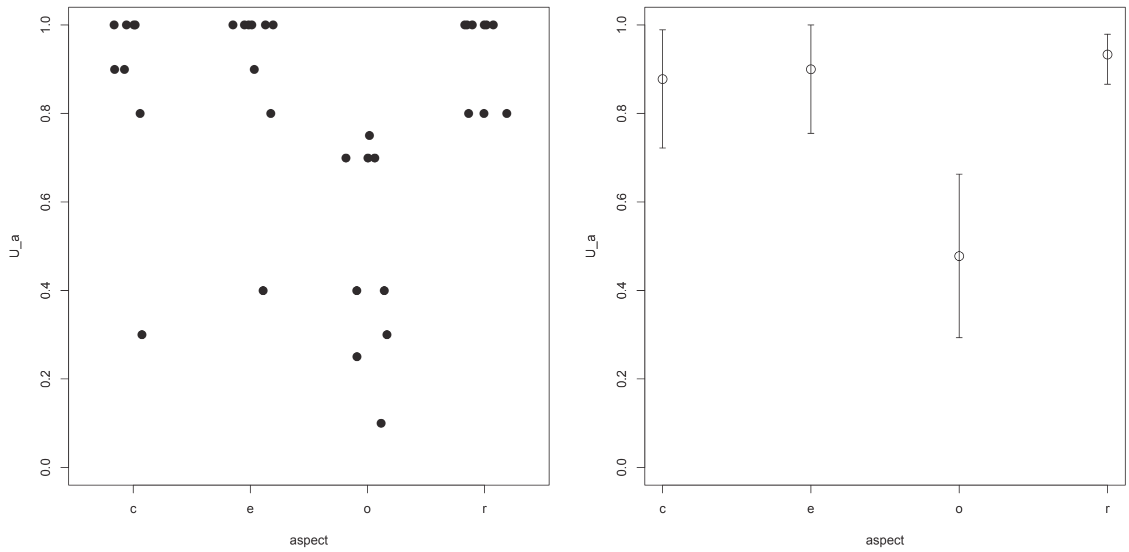
Results

The answers to the questionnaire are given in Table 6.2 (aspect *concurrency*), Table 6.3 (aspect *exclusiveness*), Table 6.4 (aspect *order*) and Table 6.5 (aspect *repetition*).

PERSONAL STRUCTURAL PROCESS MODEL UNDERSTANDABILITY The personal structural process model understandability values of the subjects for the

Table 6.3: Answers given for aspect *exclusiveness*.

subject	$q_e(A,B)$	$q_e(A,C)$	$q_e(A,D)$	$q_e(A,E)$	$q_e(B,C)$	$q_e(B,D)$	$q_e(B,E)$	$q_e(C,D)$	$q_e(C,E)$	$q_e(D,E)$	$U_e(p,s)$
solution	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\exists$	
s_2	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\exists$	1.0
s_4	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\exists$	1.0
s_6	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\exists$	1.0
s_{34}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\exists$	1.0
s_{42}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\forall$	0.9
s_{50}	$e_\forall$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\exists$	0.8
s_{52}	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\forall$	$e_\exists$	0.4
s_{56}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\exists$	1.0
s_{60}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\exists$	1.0
correct	78%	78%	89%	89%	89%	100%	100%	100%	89%	89%	



(a) Personal structural process model understandability values for the four aspects.

(b) Estimated structural process model understandability values and 95% confidence intervals for the four aspects.

Figure 6.5: Visualizations of values for experiment 1.

four aspects *concurrency* (c), *exclusiveness* (e), *order* (o) and *repetition* (r) are depicted in Figure 6.5a.

In order to test the hypothesis that the personal structural process model understandability values are normally distributed for each aspect (Hypothesis 6.1),

Table 6.4: Answers given for aspect *order*.

subject	$q_o(A,B)$	$q_o(A,C)$	$q_o(A,D)$	$q_o(A,E)$	$q_o(B,A)$	$q_o(B,C)$	$q_o(B,D)$	$q_o(B,E)$	$q_o(C,A)$	$q_o(C,B)$	$q_o(C,D)$	$q_o(C,E)$	$q_o(D,A)$	$q_o(D,B)$	$q_o(D,C)$	$q_o(D,E)$	$q_o(E,A)$	$q_o(E,B)$	$q_o(E,C)$	$q_o(E,D)$	$U_o(p,s)$
solution	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	
s_1	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.70
s_3	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\exists$	0.10
s_5	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	0.40
s_{11}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	0.40
s_{35}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	0.30
s_{51}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	0.25
s_{53}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.70
s_{55}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.75
s_{57}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.70
correct	0%	0%	0%	0%	100%	44%	56%	44%	89%	11%	44%	44%	89%	44%	44%	89%	78%	44%	44%	89%	

Table 6.5: Answers given for aspect *repetition*.

subject	$q_r(A)$	$q_r(B)$	$q_r(C)$	$q_r(D)$	$q_r(E)$	$U_r(p,s)$
solution	$r_{=1}$	r_*	r_*	$r_?$	$r_?$	
s_1	$r_{=1}$	r_*	r_*	$r_?$	$r_?$	1.0
s_3	$r_{=1}$	r_*	r_*	$r_?$	$r_?$	1.0
s_5	$r_{=1}$	r_+	r_*	$r_?$	$r_?$	0.8
s_{11}	$r_{=1}$	r_*	r_*	$r_?$	$r_?$	1.0
s_{35}	$r_{=1}$	r_*	r_*	$r_?$	$r_?$	1.0
s_{51}	$r_{=1}$	r_*	r_*	$r_?$	$r_?$	1.0
s_{53}	$r_{=1}$	r_*	r_*	$r_?$	$r_?$	1.0
s_{55}	$r_{=1}$	$r_?$	r_*	$r_?$	$r_?$	0.8
s_{57}	$r_{=1}$	r_+	r_*	$r_?$	$r_?$	0.8
correct	100%	67%	100%	100%	100%	

a Shapiro-Wilk test [142] was conducted for each of the four data sets. For *concurrency*, *exclusiveness* and *repetition*, the null-hypothesis that the data is normally distributed had to be rejected ($p \ll 0.05$). Only for *order*, this null-hypothesis could not be rejected on the $\alpha = 0.05$ level.

Possible reasons for not finding a normal distribution for *concurrency*, *exclusiveness* and *repetition* are:

- The process model is too "easy". So, most values are near 1.0. As the value range ends there, there cannot exist any bigger values "symmetric" to the values lower than 1.0.
- The process model is too "small". Only five and ten questions were asked respectively. Consequently, personal structural process model understandability values have a "step size" of 0.2 and 0.1 respectively.
- The number of subjects is too low. Only data from nine participants per aspect could be collected.

ESTIMATED STRUCTURAL PROCESS MODEL UNDERSTANDABILITY Based on the data on personal structural process model understandability, the estimated structural process model understandability values (together with the standard deviations of the corresponding personal structural process model understandability values) were computed (Table 6.6).

Additionally, also the 95% confidence intervals for the expected structural process model understandability values of the four aspects were calculated. For *order*, the method for estimating confidence intervals for means of normal distributions [117, pp. 446–447] was used. For the other three aspects, the bootstrap approach

Table 6.6: Estimated structural process model understandability values, standard deviations and 95% confidence intervals for the four aspects.

	concurrency	*exclusiveness*	*order*	*repetition*
$\widehat{U}_a(p, S)$	0.878	0.900	0.478	0.933
standard deviation	0.228	0.200	0.240	0.100
lower conf. interval bound	0.722	0.755	0.293	0.866
upper conf. interval bound	0.989	1.000	0.663	0.979

[42], which does not require normally distributed data, was applied. The lower and upper confidence interval bounds are also listed in Table 6.6.

Furthermore, the estimated structural process model understandability values and the 95% confidence intervals for the four aspects are also depicted in Figure 6.5b.

For testing the hypothesis that the structural process model understandability values for the four aspects are different (Hypothesis 6.2), Wilcoxon rank-sum tests for independent values (aspects asked for in different questionnaire groups) [117, pp. 590–597] and Wilcoxon signed-rank tests for paired values (aspects asked for in one single questionnaire group) [117, pp. 599–603] were used. Both tests do not require normally distributed data. Only for the combinations *order-concurrency*, *order-exclusiveness* and *order-repetition*, the null-hypothesis (data belongs to the same distribution) could be rejected on the $\alpha = 0.05$ level.

Here again, a possible reason that the values for *concurrency*, *exclusiveness* and *repetition* are so equal could be that the process model is too "small" and "easy" so that no really difficult parts which are of varying difficulty for the different aspects are included.

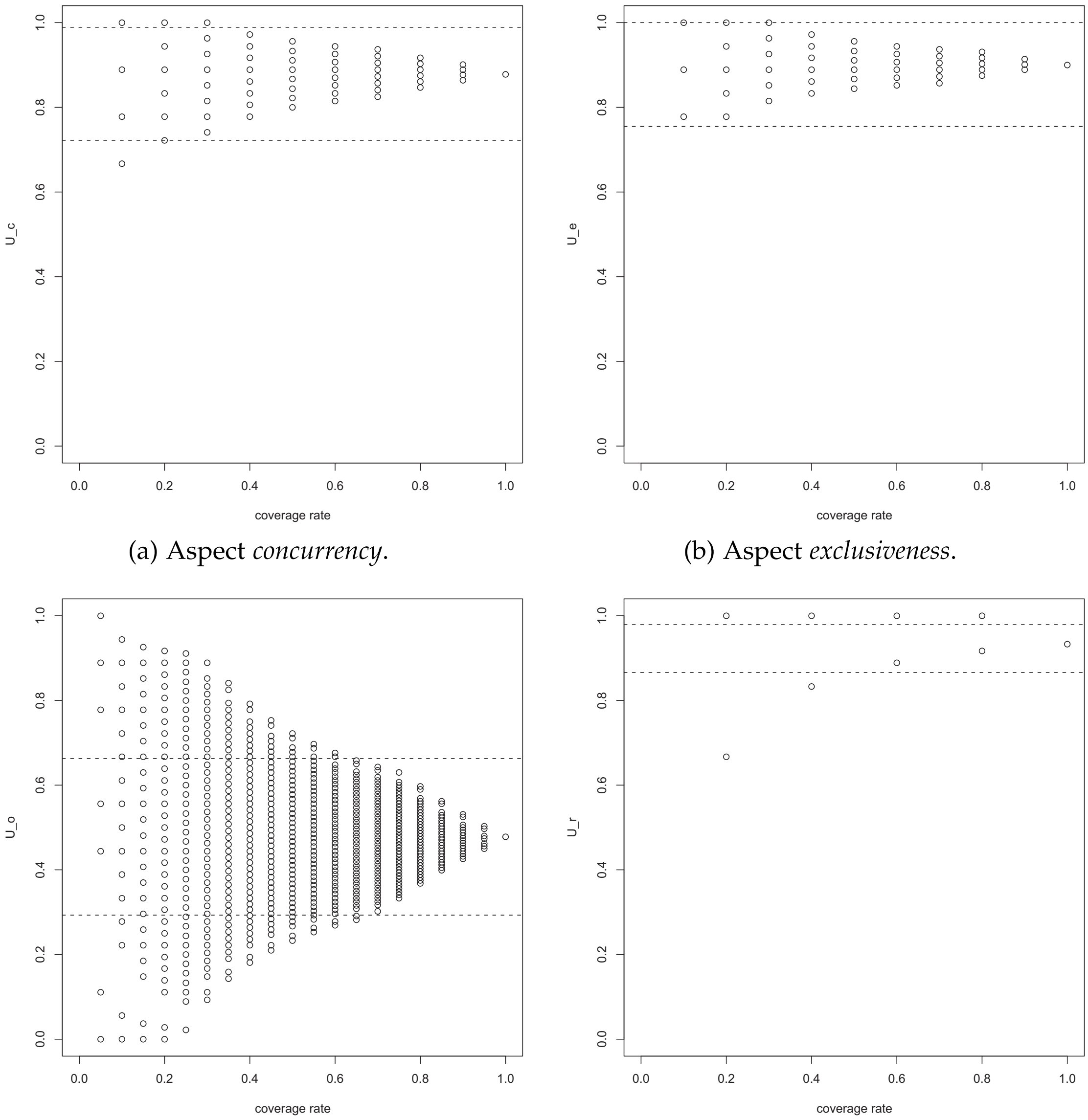
estimated partial structural process model understandability

In order to test the hypothesis about partial structural process model understandability (Hypothesis 6.3), all estimated partial structural process model understandability values for the four aspects were computed.

The values depending on the coverage rate are depicted in Figure 6.6. The dashed horizontal lines are the lower and upper 95% confidence interval bounds for the estimated structural process model understandability values of the four aspects.

In Table 6.7, the mean estimated partial structural process model understandability, the standard deviation of the estimated partial structural process model understandability values and the rate of values lower and higher than the confidence interval bounds of the four aspects are listed for all different coverage rates.

Table 6.7 and Figure 6.6 support the hypothesis—aspect *order* having the strongest effect: For the same coverage rate, many different estimated partial



(a) Aspect *concurrency*.

(b) Aspect *exclusiveness*.

(c) Aspect *order*.

(d) Aspect *repetition*.

Figure 6.6: Estimated partial structural process model understandability values of the four aspects depending on coverage rate.

structural process model understandability values exist. The smaller the coverage rate, the higher the standard deviation and the number of values outside the confidence interval.

CONCLUSION EXPERIMENT 1 The results of this first, small experiment give some interesting insights on the postulated hypotheses. While Hypothesis 6.3 (effects of asking only some of the possible questions) is fully supported, only aspect *order* has been found to be normally distributed (Hypothesis 6.1) and only some pairs of aspects (*order-concurrency*, *order-exclusiveness* and *order-repetition*) are differently distributed (Hypothesis 6.2).

Table 6.7: Data on estimated partial structural process model understandability values of the four aspects (Part 1 of 2).

(a) Aspect *concurrency*.

# questions	cov. rate	mean	s. d.	rate lower	rate higher
1	0.1	0.878	0.110	10.0%	30.0%
2	0.2	0.878	0.071	0.0%	6.7%
3	0.3	0.878	0.054	0.0%	0.8%
4	0.4	0.878	0.043	0.0%	0.0%
5	0.5	0.878	0.035	0.0%	0.0%
6	0.6	0.878	0.029	0.0%	0.0%
7	0.7	0.878	0.023	0.0%	0.0%
8	0.8	0.878	0.018	0.0%	0.0%
9	0.9	0.878	0.012	0.0%	0.0%
10	1.0	0.878	–	0.0%	0.0%

(b) Aspect *exclusiveness*.

# questions	cov. rate	mean	s. d.	rate lower	rate higher
1	0.1	0.900	0.082	0.0%	0.0%
2	0.2	0.900	0.052	0.0%	0.0%
3	0.3	0.900	0.040	0.0%	0.8%
4	0.4	0.900	0.032	0.0%	0.0%
5	0.5	0.900	0.026	0.0%	0.0%
6	0.6	0.900	0.021	0.0%	0.0%
7	0.7	0.900	0.017	0.0%	0.0%
8	0.8	0.900	0.013	0.0%	0.0%
9	0.9	0.900	0.009	0.0%	0.0%
10	1.0	0.900	–	0.0%	0.0%

Consequently, a second—and larger (concerning both size of process model and number of subjects)—experiment would be useful. In this second experiment, the results of the first experiment should be retested in a larger setting with a special focus on those hypotheses which were not supported by the outcome of the first experiment. Furthermore, the virtual subjects approach (Subsection 6.4.4) could be applied.

Validity Evaluation

Finally, the necessary validity evaluation (see paragraph *Validity Evaluation* in Subsection B.3.3) of experiment 1 has to be carried out.

Table 6.7: Data on estimated partial structural process model understandability values of the four aspects (Part 2 of 2).

(c) Aspect *order*.

# questions	cov. rate	mean	s. d.	rate lower	rate higher
1	0.05	0.478	0.333	25.0%	30.0%
2	0.10	0.478	0.224	28.4%	32.1%
3	0.15	0.478	0.177	9.3%	14.6%
4	0.20	0.478	0.149	10.6%	14.4%
5	0.25	0.478	0.129	10.5%	7.3%
6	0.30	0.478	0.114	3.7%	6.5%
7	0.35	0.478	0.101	3.7%	3.1%
8	0.40	0.478	0.091	2.8%	2.5%
9	0.45	0.478	0.082	0.8%	1.0%
10	0.50	0.478	0.074	0.7%	0.7%
11	0.55	0.478	0.067	0.3%	0.1%
12	0.60	0.478	0.061	0.0%	0.0%
13	0.65	0.478	0.055	0.0%	0.0%
14	0.70	0.478	0.049	0.0%	0.0%
15	0.75	0.478	0.043	0.0%	0.0%
16	0.80	0.478	0.037	0.0%	0.0%
17	0.85	0.478	0.031	0.0%	0.0%
18	0.90	0.478	0.025	0.0%	0.0%
19	0.95	0.478	0.018	0.0%	0.0%
20	1.00	0.478	–	0.0%	0.0%

(d) Aspect *repetition*.

# questions	cov. rate	mean	s. d.	rate lower	rate higher
1	0.2	0.933	0.149	20.0%	80.0%
2	0.4	0.933	0.086	40.0%	60.0%
3	0.6	0.933	0.057	0.0%	40.8%
4	0.8	0.933	0.037	0.0%	20.0%
5	1.0	0.933	–	0.0%	0.0%

INTERNAL VALIDITY Internal validity refers to the fact that the effects observed in the experiment are not caused by a factor which one has no control of or has not measured (see Definition B.9).

Looking at the threats to internal validity mentioned in Subsection B.3.3, one can make the following statements:

- History: During the experiment which lasted less than one hour, no events occurred which were able to strongly influence the subjects.
- Maturation: The experiment was so short that factors as, for example, fatigue, boredom or hunger had no big influence.
- Instrumentation: There was no subjective influence of a human observer of the experiment on the assessment whether a given answer was correct or not.
- Mortality: As all subjects finished the experiment, this threat played no role.
- Selection: As the subjects were randomly assigned to one of the two questionnaire groups, possible personal differences should have been balanced.

external validity External validity refers to the extent to which the results of an experiment can be generalized out of the scope of the study (see Definition B.10).

Looking at the threats to external validity mentioned in Subsection B.3.3, one can make the following statements:

- Population validity: The subjects were students with knowledge in the area of BPM. Nevertheless, the question remains whether professionals would produce the same results. Here, it is believed to be most likely that professionals have maybe a higher personal structural process model understandability than students as they have longer experience—yet resulting in the same qualitative effects as the students (e. g., for partial structural process model understandability). At least the results of an experiment analyzing differences between students and professionals in software engineering support this hypothesis [133].
- Ecological validity: The process model used in experiment 1 was only a very small "toy problem". Consequently, it is questionable whether the results are generalizable to more realistic process models. This is one reason for the conduction of the following second and larger experiment.
- Temporal validity: An influence of the time of the experiment (as long as the subjects are not tired) is hardly imaginable.

6.5.2 *Experiment 2*

Experiment Design

The second experiment was conducted as a cooperation with Jan Mendling (*Humboldt-Universität zu Berlin*) and Hajo A. Reijers (*Technische Universiteit Eindhoven*). Its goal was to (re)test the hypotheses with a larger and more realistic process model as well as with much more subjects.

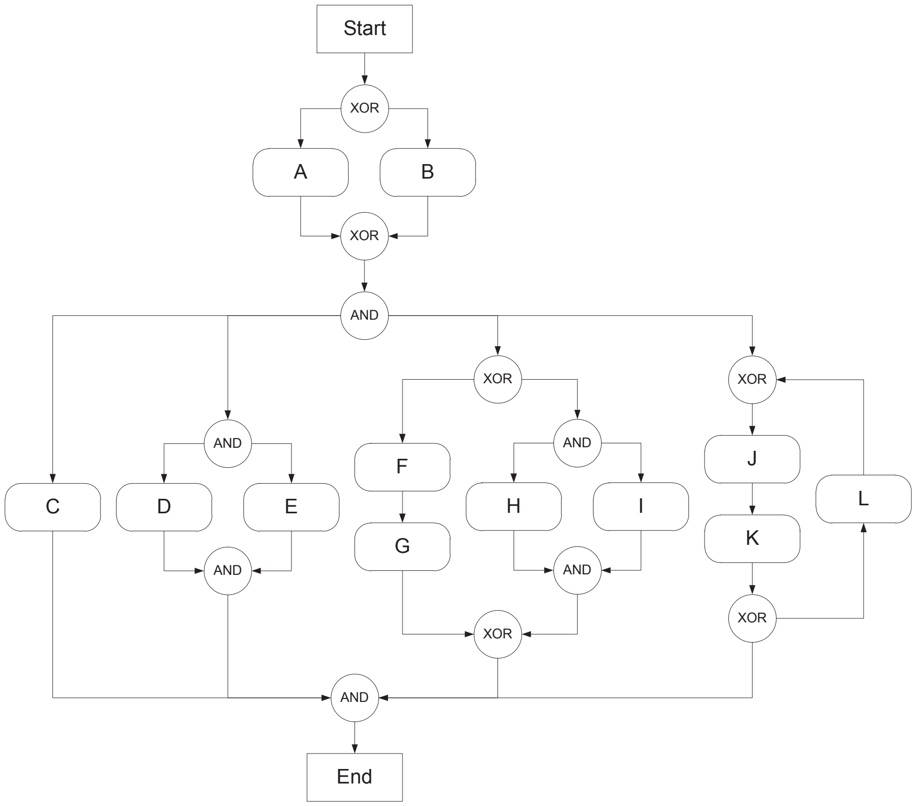


Figure 6.7: Process model used in experiment 2.

OBJECT For experiment 2, the process model depicted in Figure 6.7 was used. It was presented to the subjects in the same notation as in experiment 1.

As the process model has 12 tasks, the number of possible questions for the four aspects are $|Q_{c,max}(p)| = |Q_{e,max}(p)| = 66$, $|Q_{o,max}(p)| = 132$ and $|Q_{r,max}(p)| = 12$ (cf. Theorem 6.2).

MEASUREMENT INSTRUMENTATION As the number of possible questions per aspect is too high (except for aspect *repetition*), not all could be asked to one subject. Instead, the questions for *concurrency*, *exclusiveness* and *order* were divided into different subsets. So, a questionnaire with nine groups (groups 1 to 4: questions on *order* [*o1–o4*]; groups 5 to 8: questions on *concurrency* [*c5–c8*] and *exclusiveness* [*e5–e8*]; group 9: questions on *repetition* [*r9*]) was created—resulting in 13 data sets. In each group, 33 questions were asked (group 9 was filled by 21 "dummy questions" in order to guarantee equal conditions between the groups). The detailed assignment of the questions to the groups is shown in Tables 6.9–6.17.

SUBJECTS Students attending courses on workflow management at the *Humboldt-Universität zu Berlin*, the *Technische Universiteit Eindhoven* and the *Universität Karlsruhe (TH)* were asked to participate in the experiment. Participation was voluntary. Participating students from Berlin and Eindhoven got a bonus for their final exam—students from Karlsruhe could use the questions as a training for similar questions in their exam. Finally, 178 students answered the questionnaire.

The participants were randomly assigned to one group of the questionnaire. As this assignment was done for all students (potential participants) before knowing who would actually participate, the final number of subjects per group is varying.

An overview of the questionnaire's structure and the involved subjects is given in Table 6.8.

Results

The answers to the questionnaire are given in Tables 6.9–6.17.

SINGLE QUESTIONS First, the rate of correctly answered questions on the four aspects were analyzed. These values are given in Table 6.18 and as histograms in Figure 6.8.

As one can see, the aspect *order* has a quite different behavior compared to the other three aspects. While those have narrow peaks near the rate of 1.0, the values for aspect *order* are more spread over the whole interval with the peak near 0.6. So, most questions on the aspect *order* seem to be more difficult to answer by the subjects than those on the other aspects.

Next, it was analyzed whether a connection between the rates of correct answers to questions about *concurrency* and *exclusiveness* for the same pair of tasks exist. As both aspects deal with the execution of task pairs during a process instance execution, such a connection is imaginable. The single value pairs are depicted in Figure 6.9a. As Spearman's rank correlation coefficient (see Section C.2) is only 0.465, there is no strong connection.

Afterwards, the same analysis was done for the aspect *order* and *order* in reverse ordering. The single value pairs are depicted in Figure 6.9b. Here, Spearman's rank correlation coefficient is -0.209. So, knowing the rate of correct answers to question $q_o(t_1, t_2)$, no prediction for $q_o(t_2, t_1)$ can be given.

PERSONAL (PARTIAL) STRUCTURAL PROCESS MODEL UNDERSTANDABILITY The personal (partial) structural process model understandability values of the subjects of the nine groups are depicted in Figure 6.10 and 6.11.

In order to test the hypothesis that the personal structural process model understandability values are normally distributed for each aspect (Hypothesis 6.1), a Shapiro-Wilk test was done for each of the 13 data sets. Only for *o1*, *o2* and *o4*, the null-hypothesis that the data is normally distributed could not be rejected on the $\alpha = 0.05$ level. For the remaining data sets, the null-hypothesis had to be rejected (*o3*: $p = 0.037$; *c5*: $p = 0.035$; all others: $p \ll 0.05$).

ESTIMATED (PARTIAL) STRUCTURAL PROCESS MODEL UNDERSTANDABILITY Based on the data on personal (partial) structural process model understandability, the estimated (partial) structural process model understandability values (together with the standard deviations of the corresponding personal (partial) structural process model understandability values) were computed (Table 6.19).

Table 6.8: Questionnaire for experiment 2.

	group 1 [o1]	group 2 [o2]	group 3 [o3]	group 4 [o4]	group 5 [c5/e5]	group 6 [c6/e6]	group 7 [c7/e7]	group 8 [c8/e8]	group 9 [r9]	total
# questions *concurrency*	–	–	–	–	17	17	16	16	–	66
# questions *exclusiveness*	–	–	–	–	16	16	17	17	–	66
# questions *order*	33	33	33	33	–	–	–	–	–	132
# questions *repetition*	–	–	–	–	–	–	–	–	12	12
# "dummy questions"	–	–	–	–	–	–	–	–	21	21
# asked questions	33	33	33	33	33	33	33	33	33	297
# subjects	18	20	21	20	18	25	21	20	15	178

Table 6.9: Answers given in group 1.

subject	$q_\circ(A,B)$	$q_\circ(A,F)$	$q_\circ(A,J)$	$q_\circ(B,C)$	$q_\circ(B,G)$	$q_\circ(B,K)$	$q_\circ(C,D)$	$q_\circ(C,H)$	$q_\circ(C,L)$	$q_\circ(D,E)$	$q_\circ(D,I)$	$q_\circ(E,A)$	$q_\circ(E,F)$	$q_\circ(E,J)$	$q_\circ(F,B)$	$q_\circ(F,G)$	$q_\circ(F,K)$	$q_\circ(G,C)$	$q_\circ(G,H)$	$q_\circ(G,L)$	$q_\circ(H,D)$	$q_\circ(H,I)$	$q_\circ(I,A)$	$q_\circ(I,E)$	$q_\circ(I,J)$	$q_\circ(J,B)$	$q_\circ(J,F)$	$q_\circ(J,K)$	$q_\circ(K,C)$	$q_\circ(K,G)$	$q_\circ(K,L)$	$q_\circ(L,D)$	$q_\circ(L,H)$	$U_\circ(p,s,Q_\circ)$
solution	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	
s001	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	0.79
s002	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	0.48
s003	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	0.67
s004	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	0.61
s005	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	0.64
s006	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.58
s007	$o_\exists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	0.27
s008	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	0.55
s009	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	0.79
s010	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	0.42
s011	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	0.55
s012	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	0.52
s013	$o_\exists$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	0.58
s014	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	0.52
s015	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	0.67
s016	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	0.70
s017	$o_\exists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	0.24
s018	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	0.85
correct	44%	61%	50%	67%	61%	61%	72%	17%	83%	83%	28%	50%	33%	56%	50%	6%	89%	78%	50%	100%	44%	83%	67%	67%	33%	56%	33%	94%	72%	89%	33%	61%	33%	

Table 6.10: Answers given in group 2.

subject	$q_\circ(A,C)$	$q_\circ(A,G)$	$q_\circ(A,K)$	$q_\circ(B,D)$	$q_\circ(B,H)$	$q_\circ(B,L)$	$q_\circ(C,E)$	$q_\circ(C,I)$	$q_\circ(D,A)$	$q_\circ(D,F)$	$q_\circ(D,J)$	$q_\circ(E,B)$	$q_\circ(E,G)$	$q_\circ(E,K)$	$q_\circ(F,C)$	$q_\circ(F,H)$	$q_\circ(F,L)$	$q_\circ(G,D)$	$q_\circ(G,I)$	$q_\circ(H,A)$	$q_\circ(H,E)$	$q_\circ(H,J)$	$q_\circ(I,B)$	$q_\circ(I,F)$	$q_\circ(I,K)$	$q_\circ(J,C)$	$q_\circ(J,G)$	$q_\circ(J,L)$	$q_\circ(K,D)$	$q_\circ(K,H)$	$q_\circ(L,A)$	$q_\circ(L,E)$	$q_\circ(L,I)$	$U_\circ(p,s,Q_\circ)$
solution	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	
s_{019}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	0.58
s_{020}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	0.45
s_{021}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.76
s_{022}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.94
s_{023}	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\forall$	0.36
s_{024}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	0.64
s_{025}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	0.42
s_{026}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.91
s_{027}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	0.52
s_{028}	$o_\forall$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.58
s_{029}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	0.24
s_{030}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	0.58
s_{031}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	0.58
s_{032}	$o_\forall$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	0.52
s_{033}	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	0.70
s_{034}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	0.55
s_{035}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.88
s_{036}	$o_\forall$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	0.45
s_{037}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.76
s_{038}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	0.76
correct	55%	60%	60%	60%	60%	60%	80%	40%	50%	35%	70%	45%	85%	85%	65%	60%	90%	70%	55%	55%	65%	70%	50%	60%	85%	70%	85%	15%	70%	40%	55%	65%	35%	

Table 6.11: Answers given in group 3.

subject	$q_\circ(A,D)$	$q_\circ(A,H)$	$q_\circ(A,L)$	$q_\circ(B,E)$	$q_\circ(B,I)$	$q_\circ(C,A)$	$q_\circ(C,F)$	$q_\circ(C,J)$	$q_\circ(D,B)$	$q_\circ(D,G)$	$q_\circ(D,K)$	$q_\circ(E,C)$	$q_\circ(E,H)$	$q_\circ(E,L)$	$q_\circ(F,D)$	$q_\circ(F,I)$	$q_\circ(G,A)$	$q_\circ(G,E)$	$q_\circ(G,J)$	$q_\circ(H,B)$	$q_\circ(H,F)$	$q_\circ(H,K)$	$q_\circ(I,C)$	$q_\circ(I,G)$	$q_\circ(I,L)$	$q_\circ(J,D)$	$q_\circ(J,H)$	$q_\circ(K,A)$	$q_\circ(K,E)$	$q_\circ(K,I)$	$q_\circ(L,B)$	$q_\circ(L,F)$	$q_\circ(L,J)$	$U_\circ(p,s,Q_\circ)$
solution	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	
s_{039}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	0.45
s_{040}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\forall$	0.94
s_{041}	$o_\exists$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	0.45
s_{042}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	0.70
s_{043}	$o_\forall$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	0.48
s_{044}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	0.97
s_{045}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	0.61
s_{046}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	0.39
s_{047}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	0.39
s_{048}	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	0.58
s_{049}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	0.48
s_{050}	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	0.73
s_{051}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	0.64
s_{052}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.61
s_{053}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	0.42
s_{054}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	0.52
s_{055}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	0.39
s_{056}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\exists$	0.58
s_{057}	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\forall$	$o_\forall$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	0.55
s_{058}	$o_\exists$	$o_\forall$	$o_\exists$	$o_\forall$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	0.55
s_{059}	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\exists$	$o_\nexists$	$o_\forall$	$o_\nexists$	$o_\nexists$	$o_\nexists$	$o_\exists$	0.82
correct	86%	81%	81%	71%	71%	62%	48%	38%	67%	71%	57%	62%	24%	62%	43%	71%	62%	48%	43%	62%	71%	86%	52%	67%	86%	57%	24%	67%	48%	24%	71%	29%	33%	

Table 6.12: Answers given in group 4.

subject	q_o(A,E)	q_o(A,I)	q_o(B,A)	q_o(B,F)	q_o(B,J)	q_o(C,B)	q_o(C,G)	q_o(C,K)	q_o(D,C)	q_o(D,H)	q_o(D,L)	q_o(E,D)	q_o(E,I)	q_o(F,A)	q_o(F,E)	q_o(F,J)	q_o(G,B)	q_o(G,F)	q_o(G,K)
solution	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{060}	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{061}	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{062}	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$
s_{063}	$o_{\forall}$	$o_{\exists}$	$o_{\exists}$	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{064}	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$
s_{065}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\forall}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$
s_{066}	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{067}	$o_{\exists}$	$o_{\forall}$	$o_{\exists}$	$o_{\forall}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$
s_{068}	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\forall}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$
s_{069}	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{070}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$
s_{071}	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{072}	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{073}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$
s_{074}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$
s_{075}	$o_{\exists}$	$o_{\forall}$	$o_{\exists}$	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$
s_{076}	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{077}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$
s_{078}	$o_{\forall}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\forall}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$
s_{079}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\forall}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$
correct	70%	65%	55%	65%	65%	55%	75%	75%	75%	10%	75%	65%	10%	55%	70%	55%	55%	90%	85%

subject	q_o(H,C)	q_o(H,G)	q_o(H,L)	q_o(I,D)	q_o(I,H)	q_o(J,A)	q_o(J,E)	q_o(J,I)	q_o(K,B)	q_o(K,F)	q_o(K,J)	q_o(L,C)	q_o(L,G)	q_o(L,K)	$U_o(p,s,Q_o)$
solution	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	
s_{060}	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	0.76
s_{061}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	0.36
s_{062}	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\forall}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	0.61
s_{063}	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	0.45
s_{064}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	0.48
s_{065}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	-	$o_{\nexists}$	0.36
s_{066}	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\forall}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	0.67
s_{067}	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	0.52
s_{068}	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	0.58
s_{069}	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	0.58
s_{070}	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	0.58
s_{071}	$o_{\nexists}$	$o_{\forall}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	0.70
s_{072}	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	0.70
s_{073}	$o_{\exists}$	$o_{\forall}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	0.33
s_{074}	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	0.55
s_{075}	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	0.55
s_{076}	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	0.97
s_{077}	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	0.67
s_{078}	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\nexists}$	0.70
s_{079}	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	$o_{\exists}$	$o_{\exists}$	$o_{\nexists}$	0.45
correct	65%	40%	95%	60%	75%	60%	60%	10%	45%	25%	20%	70%	80%	30%	

Table 6.13: Answers given in group 5.

subject	$q_c(A,B)$	$q_c(A,F)$	$q_c(A,J)$	$q_c(B,D)$	$q_c(B,H)$	$q_c(B,L)$	$q_c(C,G)$	$q_c(C,K)$	$q_c(D,G)$	$q_c(D,K)$	$q_c(E,H)$	$q_c(E,L)$	$q_c(F,J)$	$q_c(G,I)$	$q_c(H,I)$	$q_c(I,J)$	$q_c(J,L)$	$U_c(p,s,Q_c)$
solution	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	
s080	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.82
s081	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	0.71
s082	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	0.53
s083	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	0.71
s084	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00
s085	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	0.94
s086	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.82
s087	$c_{\forall}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.82
s088	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.94
s089	$c_{\forall}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.59
s090	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	0.88
s091	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.94
s092	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.88
s093	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.88
s094	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.82
s095	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.88
s096	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	0.53
s097	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00
correct	83%	89%	83%	89%	83%	89%	83%	56%	83%	56%	94%	89%	94%	94%	39%	100%	83%	

subject	$q_e(A,D)$	$q_e(A,H)$	$q_e(A,L)$	$q_e(B,F)$	$q_e(B,J)$	$q_e(C,E)$	$q_e(C,I)$	$q_e(D,E)$	$q_e(D,I)$	$q_e(E,F)$	$q_e(E,J)$	$q_e(F,H)$	$q_e(F,L)$	$q_e(G,K)$	$q_e(H,K)$	$q_e(I,L)$	$U_e(p,s,Q_e)$
solution	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	
s080	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.81
s081	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s082	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s083	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.94
s084	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.94
s085	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.88
s086	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s087	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s088	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s089	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.88
s090	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s091	$e_{\nexists}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.62
s092	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s093	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.94
s094	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s095	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s096	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	1.00
s097	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.94
correct	72%	94%	94%	94%	78%	100%	100%	100%	100%	100%	83%	94%	94%	100%	100%	100%	

Table 6.14: Answers given in group 6.

subject	$q_c(A,C)$	$q_c(A,G)$	$q_c(A,K)$	$q_c(B,E)$	$q_c(B,I)$	$q_c(C,D)$	$q_c(C,H)$	$q_c(C,L)$	$q_c(D,H)$	$q_c(D,L)$	$q_c(E,I)$	$q_c(F,G)$	$q_c(F,K)$	$q_c(G,J)$	$q_c(H,J)$	$q_c(I,K)$	$q_c(K,L)$	$U_c(p,s,Q_c)$	$q_e(A,E)$	$q_e(A,I)$	$q_e(B,C)$	$q_e(B,G)$	$q_e(B,K)$	$q_e(C,F)$	$q_e(C,J)$	$q_e(D,F)$	$q_e(D,J)$	$q_e(E,G)$	$q_e(E,K)$	$q_e(F,I)$	$q_e(G,H)$	$q_e(G,L)$	$q_e(H,L)$	$q_e(J,K)$	$U_e(p,s,Q_e)$
solution	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$		$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	
s_{098}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{099}	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	0.53	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	0.81
s_{100}	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	0.82	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	0.88
s_{101}	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	0.59	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{102}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{103}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{104}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.76	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{105}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{106}	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	0.59	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	0.88
s_{107}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	0.94	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{108}	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	0.82	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{109}	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	0.59	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\forall}$	0.62
s_{110}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	0.71	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	0.94
s_{111}	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	0.76	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{112}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{113}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{114}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.75
s_{115}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	0.82	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	0.94
s_{116}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	0.71	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
s_{117}	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	0.59	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\nexists}$	$e_{\forall}$	0.94
s_{118}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	0.88	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	0.94
s_{119}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\forall}$	$c_{\exists}$	$c_{\nexists}$	0.76	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	0.94
s_{120}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	0.94
s_{121}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	0.94
s_{122}	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\nexists}$	$c_{\forall}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\exists}$	$c_{\nexists}$	1.00	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\exists}$	$e_{\forall}$	$e_{\nexists}$	$e_{\nexists}$	$e_{\exists}$	$e_{\exists}$	$e_{\forall}$	1.00
correct	80%	84%	72%	76%	76%	84%	96%	84%	92%	84%	80%	84%	88%	88%	80%	88%	84%		100%	96%	92%	100%	88%	100%	88%	96%	84%	100%	92%	92%	92%	96%	96%	92%	

Table 6.15: Answers given in group 7.

subject	q_c(A,D)	q_c(A,H)	q_c(A,L)	q_c(B,F)	q_c(B,J)	q_c(C,E)	q_c(C,I)	q_c(D,E)	q_c(D,I)	q_c(E,F)	q_c(E,J)	q_c(F,H)	q_c(F,L)	q_c(G,K)	q_c(H,K)	q_c(I,L)	$U_c(p,s,Q_c)$
solution	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	
s123	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
s124	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	0.69
s125	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
s126	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	-	$c_\exists$	0.62
s127	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	0.81
s128	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	0.69
s129	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	0.88
s130	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	0.88
s131	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
s132	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
s133	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
s134	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
s135	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
s136	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	0.94
s137	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
s138	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\nexists$	$c_\nexists$	0.81
s139	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\nexists$	$c_\nexists$	0.81
s140	$c_\nexists$	$c_\nexists$	-	$c_\nexists$	$c_\nexists$	$c_\forall$	-	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	-	$c_\nexists$	-	$c_\exists$	$c_\nexists$	0.62
s141	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
s142	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	0.75
s143	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	1.00
correct	86%	81%	71%	86%	86%	95%	90%	100%	95%	95%	90%	86%	81%	95%	86%	86%	

subject	q_e(A,B)	q_e(A,F)	q_e(A,J)	q_e(B,D)	q_e(B,H)	q_e(B,L)	q_e(C,G)	q_e(C,K)	q_e(D,G)	q_e(D,K)	q_e(E,H)	q_e(E,L)	q_e(F,J)	q_e(G,I)	q_e(H,I)	q_e(I,J)	q_e(J,L)	$U_e(p,s,Q_e)$
solution	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	
s123	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	0.94
s124	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\forall$	$e_\exists$	$e_\exists$	0.94
s125	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\forall$	$e_\exists$	$e_\exists$	0.88
s126	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s127	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s128	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\forall$	$e_\exists$	$e_\exists$	0.94
s129	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\forall$	$e_\exists$	$e_\exists$	0.94
s130	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s131	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	0.94
s132	$e_\nexists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\forall$	$e_\exists$	$e_\exists$	0.65
s133	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s134	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s135	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	-	$e_\forall$	$e_\exists$	$e_\exists$	0.88
s136	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\forall$	$e_\exists$	$e_\exists$	0.94
s137	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\nexists$	0.94
s138	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s139	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s140	$e_\nexists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	0.88
s141	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s142	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s143	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
correct	100%	100%	90%	100%	90%	100%	95%	90%	100%	86%	100%	100%	100%	95%	67%	100%	95%	

Table 6.16: Answers given in group 8.

subject	$q_c(A,E)$	$q_c(A,I)$	$q_c(B,C)$	$q_c(B,G)$	$q_c(B,K)$	$q_c(C,F)$	$q_c(C,J)$	$q_c(D,F)$	$q_c(D,J)$	$q_c(E,G)$	$q_c(E,K)$	$q_c(F,I)$	$q_c(G,H)$	$q_c(G,L)$	$q_c(H,L)$	$q_c(J,K)$	$U_c(p,s,Q_c)$
solution	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	
s_{144}	$c_\nexists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	0.81
s_{145}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	0.94
s_{146}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	1.00
s_{147}	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	0.81
s_{148}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	1.00
s_{149}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	1.00
s_{150}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	1.00
s_{151}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\exists$	$c_\nexists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	0.75
s_{152}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	1.00
s_{153}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\nexists$	$c_\nexists$	0.94
s_{154}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	1.00
s_{155}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	1.00
s_{156}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\forall$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	0.81
s_{157}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	0.88
s_{158}	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\nexists$	0.88
s_{159}	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\exists$	$c_\nexists$	$c_\exists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\forall$	0.56
s_{160}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	0.94
s_{161}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	0.94
s_{162}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	0.75
s_{163}	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\exists$	$c_\forall$	$c_\nexists$	$c_\nexists$	$c_\exists$	$c_\exists$	$c_\nexists$	0.94
correct	95%	90%	90%	100%	95%	90%	85%	90%	80%	80%	70%	100%	90%	95%	90%	95%	

subject	$q_e(A,C)$	$q_e(A,G)$	$q_e(A,K)$	$q_e(B,E)$	$q_e(B,I)$	$q_e(C,D)$	$q_e(C,H)$	$q_e(C,L)$	$q_e(D,H)$	$q_e(D,L)$	$q_e(E,I)$	$q_e(F,G)$	$q_e(F,K)$	$q_e(G,J)$	$q_e(H,J)$	$q_e(I,K)$	$q_e(K,L)$	$U_e(p,s,Q_e)$
solution	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	
s_{144}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	0.94
s_{145}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{146}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{147}	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\nexists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	0.59
s_{148}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{149}	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	0.88
s_{150}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{151}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	0.82
s_{152}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{153}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{154}	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	0.94
s_{155}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{156}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{157}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	0.82
s_{158}	$e_\nexists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\nexists$	0.76
s_{159}	-	$e_\exists$	$e_\forall$	$e_\nexists$	$e_\forall$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\nexists$	$e_\nexists$	$e_\forall$	$e_\exists$	$e_\nexists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\nexists$	0.41
s_{160}	$e_\exists$	$e_\exists$	-	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	0.94
s_{161}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{162}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
s_{163}	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\forall$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	$e_\exists$	1.00
correct	80%	95%	70%	80%	90%	100%	100%	100%	95%	95%	95%	75%	90%	100%	100%	95%	80%	

Table 6.17: Answers given in group 9.

subject	$q_r(A)$	$q_r(B)$	$q_r(C)$	$q_r(D)$	$q_r(E)$	$q_r(F)$	$q_r(G)$	$q_r(H)$	$q_r(I)$	$q_r(J)$	$q_r(K)$	$q_r(L)$	$U_r(p,s)$
solution	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	
s_{164}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{165}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{166}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{167}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{168}	r_*	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	r_*	$r_?$	$r_?$	r_+	r_+	r_*	0.83
s_{169}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{170}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_*	r_*	0.92
s_{171}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{172}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{173}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{174}	$r_?$	$r_?$	r_+	r_*	r_*	$r_{=1}$	$r_{=1}$	$r_?$	$r_{=1}$	r_+	$r_?$	r_*	0.42
s_{175}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{176}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{177}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
s_{178}	$r_?$	$r_?$	$r_{=1}$	$r_{=1}$	$r_{=1}$	$r_?$	$r_?$	$r_?$	$r_?$	r_+	r_+	r_*	1.00
correct	93%	100%	93%	93%	93%	93%	87%	100%	93%	100%	87%	100%	

Additionally, also the 95% confidence intervals for the estimated (partial) structural process model understandability values of the 13 data sets were computed. For *o1*, *o2* and *o4*, the method for estimating confidence intervals for means of normal distributions (see experiment 1) was used. For the other ten data sets, the bootstrap approach which does not require normally distributed data was applied. The lower and upper confidence interval bounds are also listed in Table 6.19.

The estimated (partial) structural process model understandability values and the 95% confidence intervals for the 13 data sets are also depicted graphically in Figure 6.12.

Finally, it was analyzed whether the distributions of the personal partial structural process understandability values of the four data sets of each of the aspects *concurrency*, *exclusiveness* and *order* are the same. For that purpose, a Kruskal-Wallis rank sum test was conducted for each of these three aspects. The null-hypothesis (same distribution) could not be rejected on the $\alpha = 0.05$ level for all of them. So, the difficulty of the four subsets of questions for each of these aspects seems to be quite equivalent.

VIRTUAL SUBJECTS APPROACH As the process model used in experiment 2 is so large that the high number of questions could not be asked entirely to single subjects, the questions were divided into different subsets (see paragraph *Experiment Design*) for later use of the virtual subjects approach (see Subsection 6.4.4).

In order to show that this approach is legitimate, Hypothesis 6.4 was tested: Using the data on the aspects *concurrency*, *exclusiveness* and *order* from experiment 1, the questions for each aspect were randomly divided into two halves of the same size simulating two groups of questions which could be answered by

Table 6.18: Correct given answers per question for the four aspects (Part 1 of 2).

(a) Aspect *concurrency*.

	B	C	D	E	F	G	H	I	J	K	L
A	83%	80%	86%	95%	89%	84%	81%	90%	83%	72%	71%
B		90%	89%	76%	86%	100%	83%	76%	86%	95%	89%
C			84%	95%	90%	83%	96%	90%	85%	56%	84%
D				100%	90%	83%	92%	95%	80%	56%	84%
E					95%	80%	94%	80%	90%	70%	89%
F						84%	86%	100%	94%	88%	81%
G							90%	94%	88%	95%	95%
H								39%	80%	86%	90%
I									100%	88%	86%
J										95%	83%
K											84%

(b) Aspect *exclusiveness*.

	B	C	D	E	F	G	H	I	J	K	L
A	100%	80%	72%	100%	100%	95%	94%	96%	90%	70%	94%
B		92%	100%	80%	94%	100%	90%	90%	78%	88%	100%
C			100%	100%	100%	95%	100%	100%	88%	90%	100%
D				100%	96%	100%	95%	100%	84%	86%	95%
E					100%	100%	100%	95%	83%	92%	100%
F						75%	94%	92%	100%	90%	94%
G							92%	95%	100%	100%	96%
H								67%	100%	100%	96%
I									100%	95%	100%
J										92%	95%
K											80%

two different groups of subjects. In the next step, Spearman's rank correlation coefficient between the personal partial structural process model understandability values from the two halves was computed. This was repeated 5,000 times for each aspect.

The corresponding empirical cumulative distribution functions are depicted in Figure 6.13. The medians were 0.714 (*concurrency*), 0.818 (*exclusiness*) and 0.933 (*order*). So, the approach seems to be legitimate.

The resulting virtual subjects are listed in Table 6.20.

In the remainder of this section, the resulting virtual personal structural process model understandability values are denoted as $U^*_a(p,s)$, the virtual estimated structural process model understandability values as $\widehat{U}^*_a(p,S)$ and the virtual estimated partial structural process model understandability as $\widehat{U}^*_a(p,S,Q_a)$ ($a \in \{c,e,o,r\}$).

(virtual) personal structural process model understandability The (virtual) personal structural process model understandability values of the

Table 6.18: Correct given answers per question for the four aspects (Part 2 of 2).

(c) Aspect *order*.

	A	B	C	D	E	F	G	H	I	J	K	L
A		44%	55%	86%	70%	61%	60%	81%	65%	50%	60%	81%
B	55%		67%	60%	71%	65%	61%	60%	71%	65%	61%	60%
C	62%	55%		72%	80%	48%	75%	17%	40%	38%	75%	83%
D	50%	67%	75%		83%	35%	71%	10%	28%	70%	57%	75%
E	50%	45%	62%	65%		33%	85%	24%	10%	56%	85%	62%
F	55%	50%	65%	43%	70%		6%	60%	71%	55%	89%	90%
G	62%	55%	78%	70%	48%	90%		50%	55%	43%	85%	100%
H	55%	62%	65%	44%	65%	71%	40%		83%	70%	86%	95%
I	67%	50%	52%	60%	67%	60%	67%	75%		33%	85%	86%
J	60%	56%	70%	57%	60%	33%	85%	24%	10%		94%	15%
K	67%	45%	72%	70%	48%	25%	89%	40%	24%	20%		33%
L	55%	71%	70%	61%	65%	29%	80%	33%	35%	33%	30%	

(d) Aspect *repetition*.

A	B	C	D	E	F	G	H	I	J	K	L
93%	100%	93%	93%	93%	93%	87%	100%	93%	100%	87%	100%

(virtual) subjects for the four aspects *concurrency*, *exclusiveness*, *order* and *repetition* are depicted in Figure 6.14a.

In order to test the hypothesis that the personal structural process model understandability values are normally distributed for each aspect (Hypothesis 6.1), a Shapiro-Wilk test for each of the four data sets was done. For *concurrency*, *exclusiveness* and *repetition*, the null-hypothesis that the data is normally distributed had to be rejected (*concurrency*: $p = 0.023$; all others: $p \ll 0.05$). Only for *order*, this null-hypothesis could not be rejected on the $\alpha = 0.05$ level.

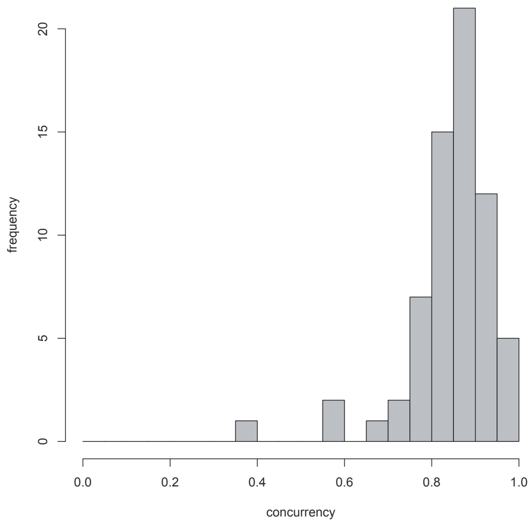
(VIRTUAL) ESTIMATED STRUCTURAL PROCESS MODEL UNDERSTANDABILITY Based on the four data sets, the (virtual) estimated structural process model understandability values (together with the standard deviations of the corresponding (virtual) personal structural process model understandability values) were computed (Table 6.21).

Additionally, also the 95% confidence intervals for the (virtual) estimated structural process model understandability values of the four aspects were computed. For *order*, the method for estimating confidence intervals for means of normal distributions was used. For the other three aspects, the bootstrap approach which does not require normally distributed data was applied. The lower and upper confidence interval bounds are also listed in Table 6.21.

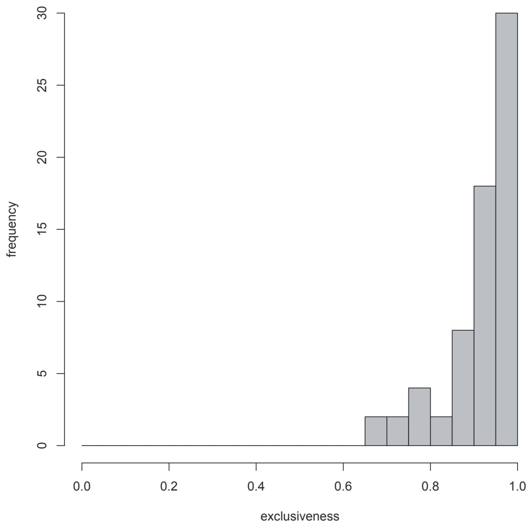
The (virtual) estimated structural process model understandability values and the 95% confidence intervals for the four aspects are also depicted graphically in Figure 6.14b.

Table 6.19: Estimated (partial) structural process model understandability values, standard deviations and 95% confidence intervals for the four aspects and 13 data sets.

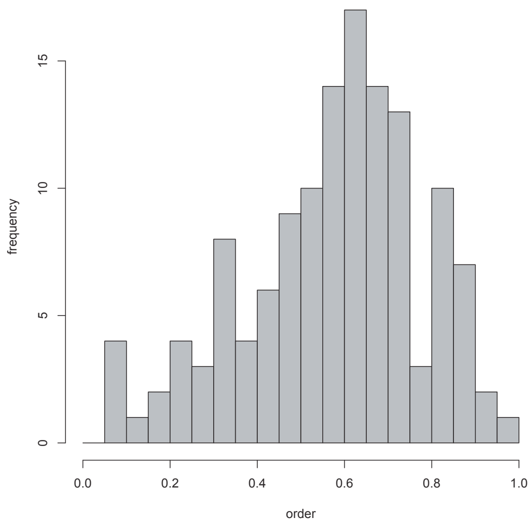
	o1	o2	o3	o4	c5	c6	c7	c8	e5	e6	e7	e8	r9
$\widehat{U}_a(p, S, Q_a)$ or $\widehat{U}_r(p, S)$	0.579	0.609	0.583	0.579	0.816	0.835	0.881	0.898	0.942	0.941	0.946	0.905	0.945
standard deviation	0.163	0.186	0.169	0.154	0.147	0.167	0.139	0.118	0.098	0.093	0.081	0.159	0.153
lower conf. interval bound	0.498	0.521	0.513	0.506	0.747	0.769	0.818	0.844	0.893	0.900	0.907	0.830	0.861
upper conf. interval bound	0.661	0.697	0.653	0.651	0.878	0.898	0.932	0.948	0.983	0.972	0.978	0.965	1.000



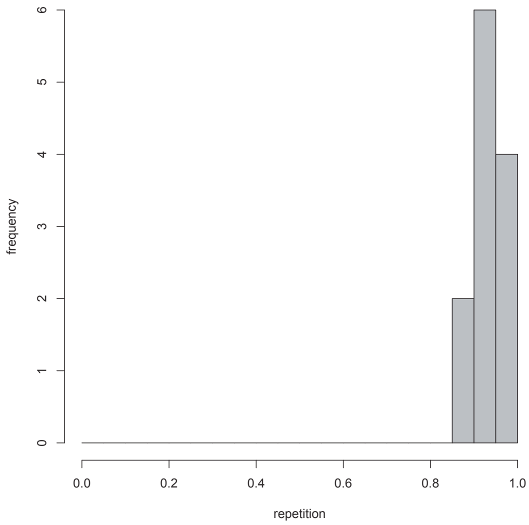
(a) Aspect *concurrency*.



(b) Aspect *exclusiveness*.



(c) Aspect *order*.

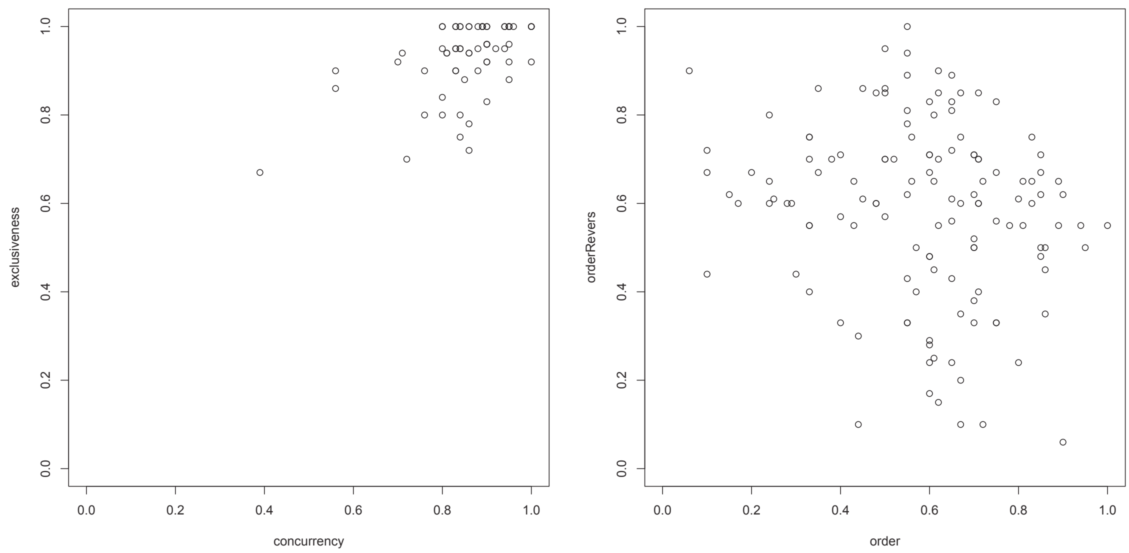


(d) Aspect *repetition*.

Figure 6.8: Histograms for rates of correct answers for the four aspects.

For testing the hypothesis that the structural process model understandability values for the four aspects are different (Hypthesis 6.2), Wilcoxon rank-sum tests for independent values were conducted. This test does not require normally distributed data. Only for the combination *exclusiveness-repetition*, the null-hypothesis (data belongs to same distribution) could not be rejected on the $\alpha = 0.05$ level ($p = 0.110$).

(virtual) estimated partial structural process model understandability In order to test the hypothesis about partial structural process model understandability (Hypothesis 6.3), all (virtual) estimated partial structural process model understandability values for the four aspects were computed. For *concurrency*, *exclusiveness* and *order*, the data of the virtual subjects were used.



(a) Rates of correct answers *concurrency/exclusiveness*.

(b) Rates of correct answers *order/order* in reverse order.

Figure 6.9: Scatter plots with rates of correct answers for experiment 2.

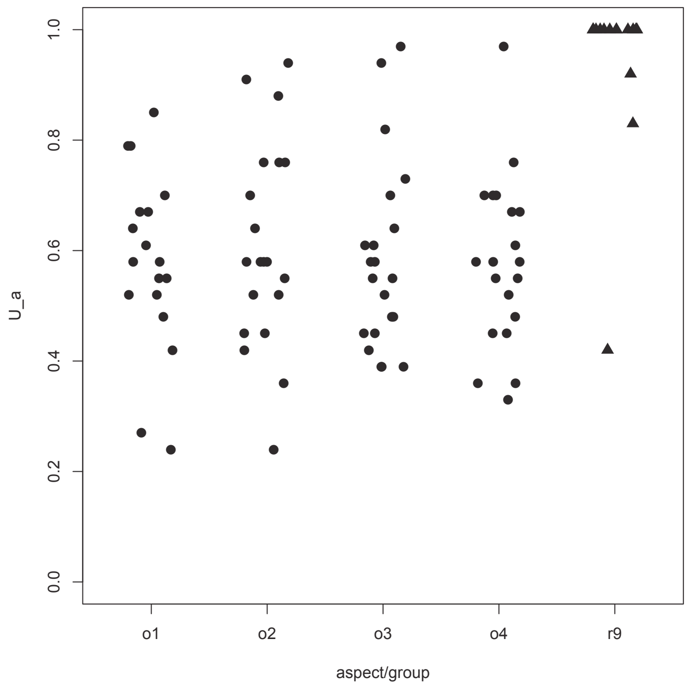


Figure 6.10: Personal (partial) structural process model understandability values for the aspects *order* and *repetition*.

The values depending on the coverage rate are depicted in Figure 6.15. The dashed horizontal lines are the lower and upper 95% confidence interval bounds for the (virtual) estimated structural process model understandability values of the four aspects.

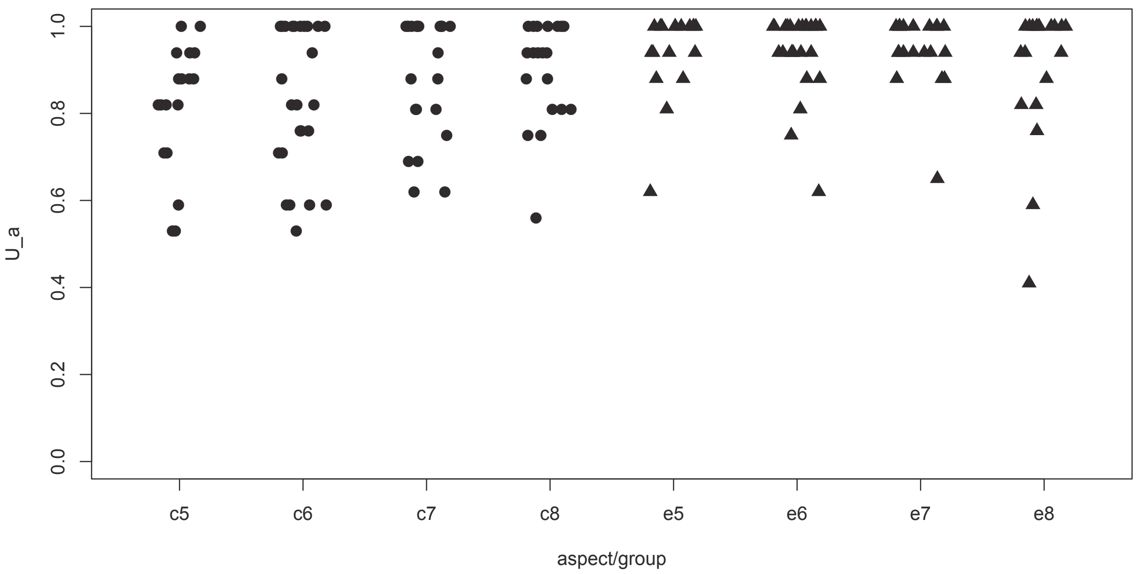


Figure 6.11: Personal partial structural process model understandability values for the aspects *concurrency* and *exclusiveness*.

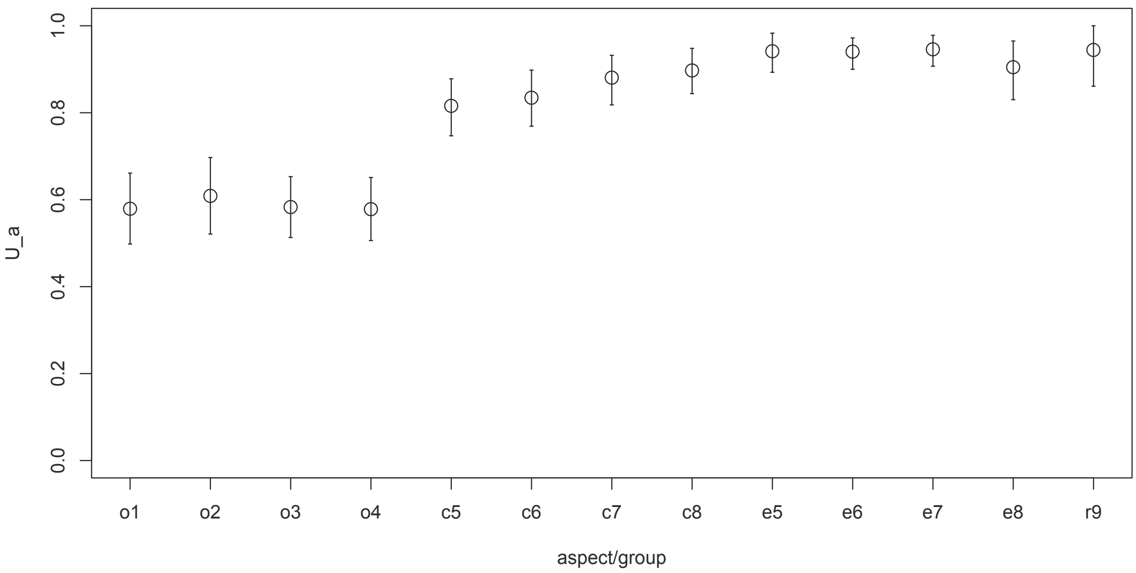
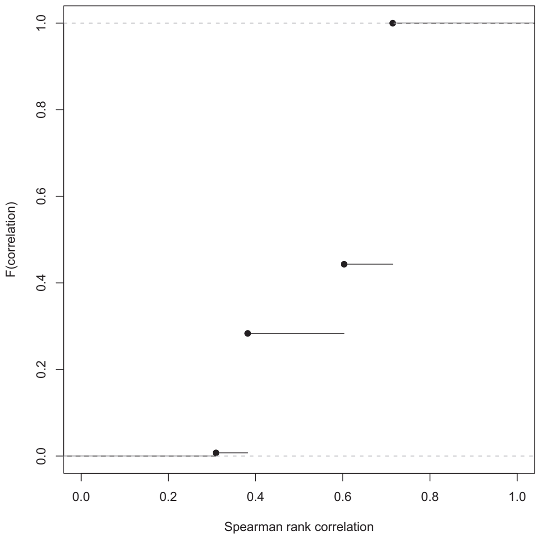


Figure 6.12: Estimated (partial) structural process model understandability values and 95% confidence intervals for the 13 data sets.

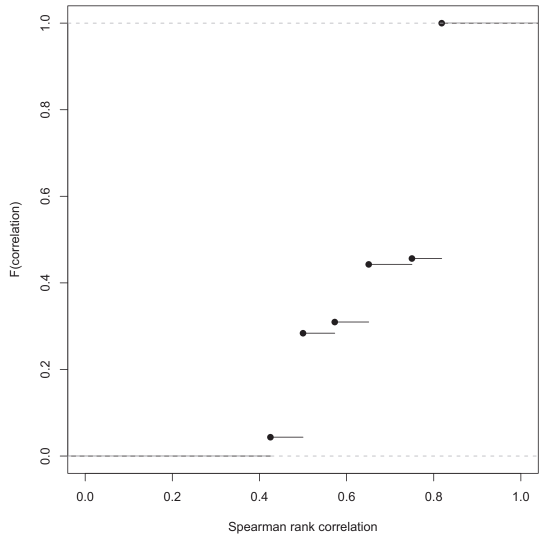
Because of the "combinatoric explosion" (cf. Theorem 6.3)[1], a probabilistic algorithm had to be used for these plots: For each analyzed coverage rate, 1,000,000 (for the aspects *concurrency* and *exclusiveness*) and 5,000,000 (for the aspect *order*) subsets of questions were randomly selected, respectively. Exact values could only be computed for very small and very large coverage rates as well as for the aspect *repetition*.

In Table 6.22, the mean (virtual) estimated partial structural process model understandability, the standard deviation of the (virtual) estimated partial structural

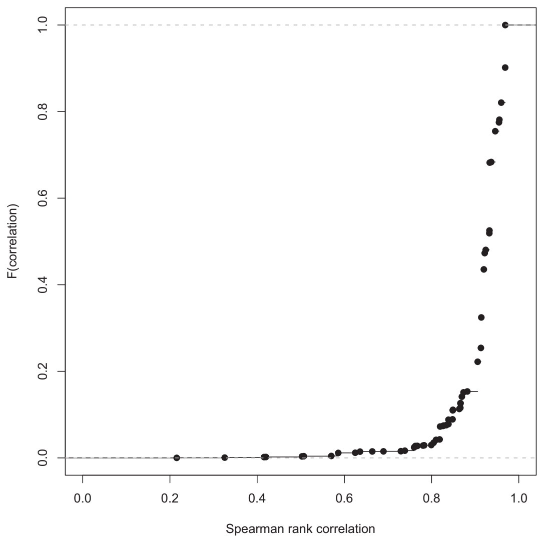
1 The highest number of possible subsets exists for the aspect *order* and coverage rate 0.5. Here, $\binom{132}{66} \approx 3.8 \times 10^{38}$ different subsets exist.



(a) Aspect *concurrency*.



(b) Aspect *exclusiveness*.



(c) Aspect *order*.

Figure 6.13: Empirical cumulative distribution functions for Spearman's rank correlation coefficients between two halves of questions of experiment 1.

process model understandability values and the rate of values lower and higher than the confidence interval bounds of the four aspects are listed for different coverage rates.

Also for these tables, a probabilistic algorithm had to be used: For each analyzed coverage rate, 100,000 subsets of questions were randomly selected. Exact values could only be computed for very small and very large coverage rates as well as for the aspect *repetition*.

The different values for the mean in Table 6.23a, 6.23c and 6.23d compared to Table 6.21 are caused by rounding errors.

Table 6.22 and Figure 6.15 support the hypothesis—the aspect *repetition* having the weakest effect: For the same coverage rate, many different (virtual) estimated

Table 6.20: Construction of virtual subjects for experiment 2 (Part 1 of 3).

(a) Aspect *concurrency*.

virtual subject	subject1	subject2	subject3	subject4	$U_c^*(p, s)$
s_{01}^*	s_{082}	s_{101}	s_{126}	s_{159}	0.58
s_{02}^*	s_{096}	s_{106}	s_{140}	s_{151}	0.62
s_{03}^*	s_{089}	s_{109}	s_{124}	s_{162}	0.65
s_{04}^*	s_{081}	s_{110}	s_{128}	s_{144}	0.73
s_{05}^*	s_{083}	s_{116}	s_{142}	s_{147}	0.74
s_{06}^*	s_{080}	s_{104}	s_{127}	s_{156}	0.80
s_{07}^*	s_{086}	s_{119}	s_{139}	s_{145}	0.83
s_{08}^*	s_{087}	s_{108}	s_{129}	s_{153}	0.86
s_{09}^*	s_{094}	s_{115}	s_{130}	s_{160}	0.86
s_{10}^*	s_{090}	s_{118}	s_{136}	s_{161}	0.91
s_{11}^*	s_{092}	s_{107}	s_{123}	s_{163}	0.94
s_{12}^*	s_{093}	s_{102}	s_{125}	s_{146}	0.97
s_{13}^*	s_{095}	s_{105}	s_{131}	s_{148}	0.97
s_{14}^*	s_{085}	s_{112}	s_{132}	s_{149}	0.98
s_{15}^*	s_{088}	s_{113}	s_{134}	s_{150}	0.98
s_{16}^*	s_{091}	s_{114}	s_{135}	s_{152}	0.98
s_{17}^*	s_{084}	s_{120}	s_{141}	s_{154}	1.00
s_{18}^*	s_{097}	s_{121}	s_{143}	s_{155}	1.00

partial structural process model understandability values exist. The reason for this is the different difficulty of the single questions as already shown in Table 6.18 and Figure 6.8. The smaller the coverage rate, the higher the standard deviation and the number of values outside the confidence interval.

For the process model used in experiment 2, a coverage rate of 0.25 produces less than 1% lower or upper outliers for all four aspects.

Validity Evaluation

Also for the second experiment, the necessary validity evaluation (see paragraph *Validity Evaluation* in Subsection B.3.3) is carried out.

Table 6.20: Construction of virtual subjects for experiment 2 (Part 2 of 3).

(b) Aspect *exclusiveness*.

virtual subject	subject1	subject2	subject3	subject4	$U_e^*(p,s)$
s_{19}^*	s_{091}	s_{109}	s_{132}	s_{159}	0.58
s_{20}^*	s_{080}	s_{114}	s_{125}	s_{147}	0.76
s_{21}^*	s_{085}	s_{106}	s_{135}	s_{151}	0.86
s_{22}^*	s_{089}	s_{110}	s_{140}	s_{149}	0.89
s_{23}^*	s_{083}	s_{115}	s_{123}	s_{144}	0.94
s_{24}^*	s_{084}	s_{118}	s_{124}	s_{154}	0.94
s_{25}^*	s_{093}	s_{119}	s_{128}	s_{160}	0.94
s_{26}^*	s_{097}	s_{120}	s_{129}	s_{145}	0.95
s_{27}^*	s_{081}	s_{121}	s_{131}	s_{146}	0.97
s_{28}^*	s_{082}	s_{101}	s_{136}	s_{148}	0.98
s_{29}^*	s_{086}	s_{102}	s_{126}	s_{150}	1.00
s_{30}^*	s_{087}	s_{104}	s_{127}	s_{152}	1.00
s_{31}^*	s_{088}	s_{105}	s_{130}	s_{153}	1.00
s_{32}^*	s_{090}	s_{107}	s_{134}	s_{155}	1.00
s_{33}^*	s_{092}	s_{108}	s_{139}	s_{156}	1.00
s_{34}^*	s_{094}	s_{112}	s_{141}	s_{161}	1.00
s_{35}^*	s_{095}	s_{113}	s_{142}	s_{162}	1.00
s_{36}^*	s_{096}	s_{116}	s_{143}	s_{163}	1.00

INTERNAL VALIDITY Looking at the threats to internal validity mentioned in Subsection B.3.3, one can make the following statements:

- History: As an online questionnaire was used, it is not known whether strongly influencing events occurred for some of the subjects while they answered the questions.
- Maturation: It was not measured how much time the subjects needed for answering the online questionnaire. As the number of asked questions did not differ that much between experiment 1 (20 or 25 questions per subject) and experiment 2 (33 questions per subject), it is believed that factors as, for example, fatigue, boredom or hunger had also no big influence for experiment 2.

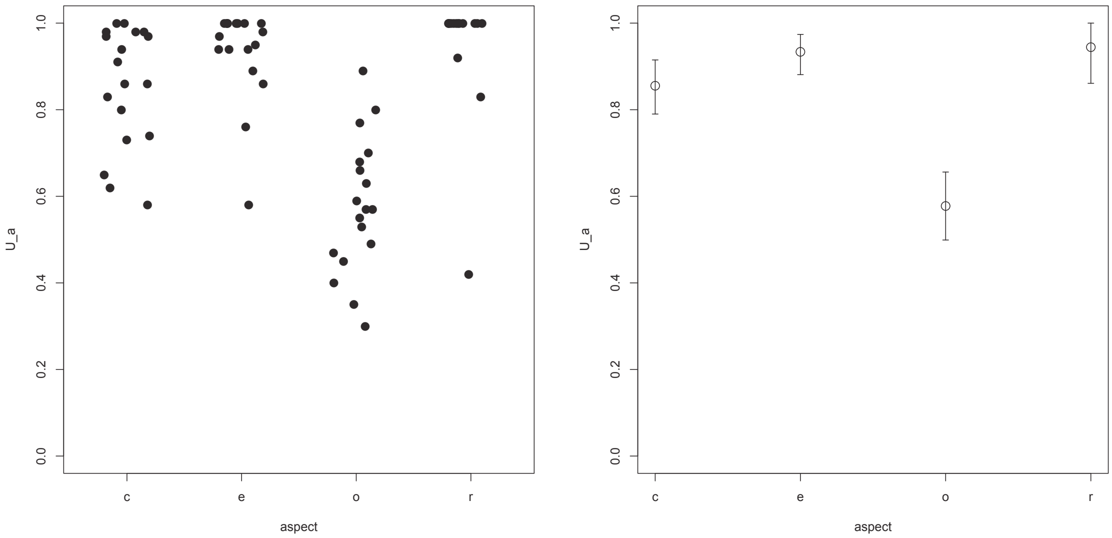
Table 6.20: Construction of virtual subjects for experiment 2 (Part 3 of 3).

(c) Aspect *order*.

virtual subject	subject1	subject2	subject3	subject4	$U_o^*(p, s)$
s_{37}^*	s_{017}	s_{029}	s_{046}	s_{073}	0.30
s_{38}^*	s_{007}	s_{023}	s_{047}	s_{061}	0.35
s_{39}^*	s_{010}	s_{025}	s_{055}	s_{065}	0.40
s_{40}^*	s_{002}	s_{020}	s_{053}	s_{063}	0.45
s_{41}^*	s_{012}	s_{036}	s_{039}	s_{079}	0.47
s_{42}^*	s_{014}	s_{032}	s_{041}	s_{064}	0.49
s_{43}^*	s_{008}	s_{034}	s_{043}	s_{074}	0.53
s_{44}^*	s_{011}	s_{019}	s_{054}	s_{075}	0.55
s_{45}^*	s_{006}	s_{028}	s_{057}	s_{068}	0.57
s_{46}^*	s_{013}	s_{030}	s_{058}	s_{070}	0.57
s_{47}^*	s_{004}	s_{031}	s_{048}	s_{062}	0.59
s_{48}^*	s_{005}	s_{024}	s_{056}	s_{066}	0.63
s_{49}^*	s_{003}	s_{033}	s_{045}	s_{077}	0.66
s_{50}^*	s_{015}	s_{021}	s_{052}	s_{071}	0.68
s_{51}^*	s_{016}	s_{038}	s_{051}	s_{072}	0.70
s_{52}^*	s_{001}	s_{035}	s_{042}	s_{078}	0.77
s_{53}^*	s_{009}	s_{026}	s_{050}	s_{060}	0.80
s_{54}^*	s_{018}	s_{022}	s_{059}	s_{076}	0.89

Table 6.21: (Virtual) estimated structural process model understandability values, standard deviations and 95% confidence intervals for the four aspects.

	concurrency	*exclusiveness*	*order*	*repetition*
$\widehat{U}_a^*(p, S)$ or $\widehat{U}_r(p, S)$	0.856	0.934	0.578	0.945
standard deviation	0.140	0.109	0.157	0.153
lower conf. interval bound	0.790	0.881	0.499	0.861
upper conf. interval bound	0.915	0.974	0.656	1.000



(a) (Virtual) personal structural process model understandability values.

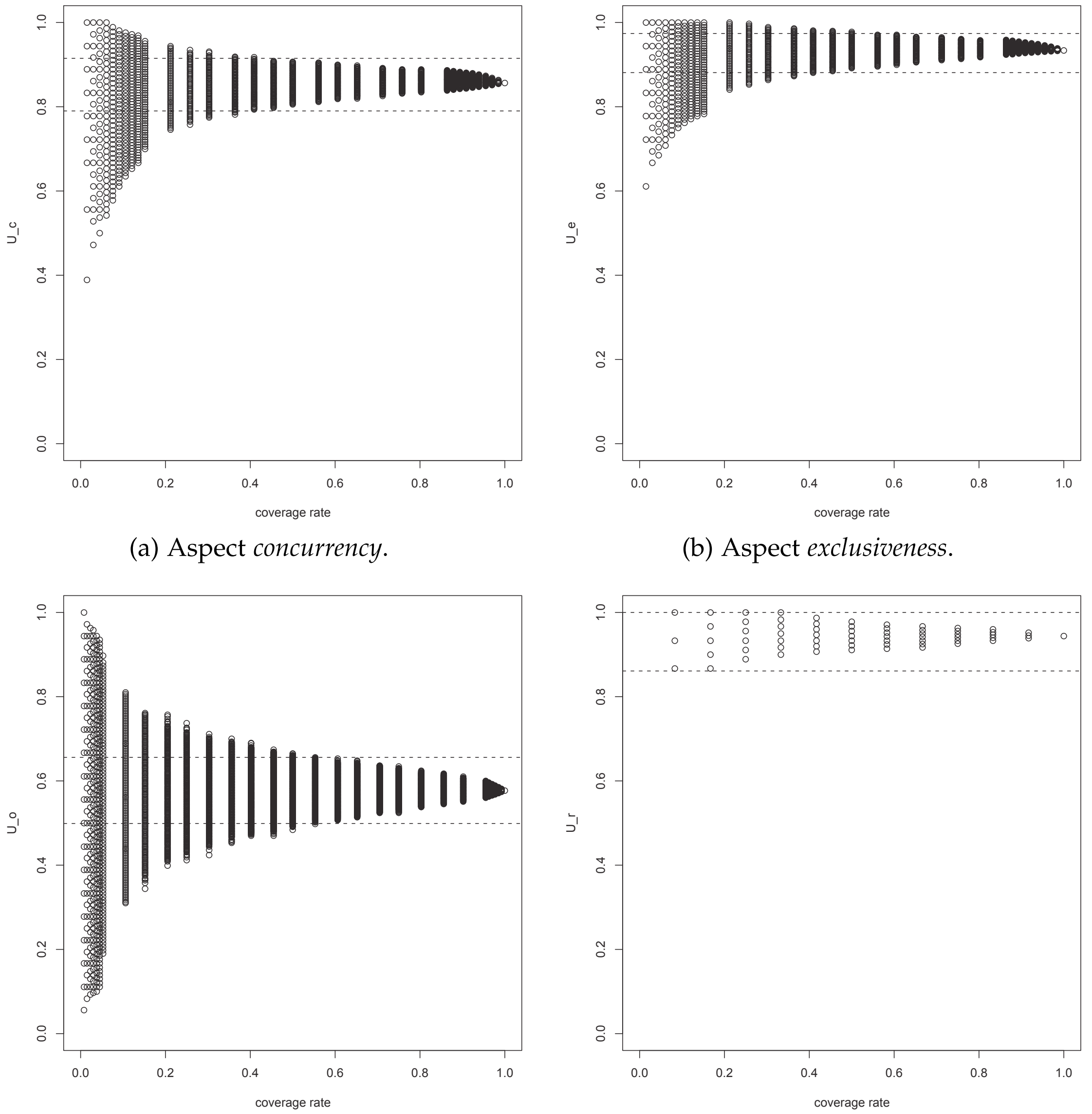
(b) (Virtual) estimated structural process model understandability values and 95% confidence intervals.

Figure 6.14: Visualizations of (virtual) structural process model understandability values of experiment 2.

- Instrumentation: There was no subjective influence of a human observer of the experiment on the assessment whether a given answer was correct or not.
- Mortality: Some students started the online questionnaire without finishing it. Their incomplete data was not used in the evaluation of the experiment. In the text above, only the data of the 178 subjects who finished the questionnaire is utilized.
- Selection: As the subjects were randomly assigned to one of the nine questionnaire groups, possible personal differences should have been balanced.

EXTERNAL VALIDITY Looking at the threats to external validity mentioned in Subsection B.3.3, one can make the following statements:

- Population validity: The same remarks on the students vs. professionals "problem" as for experiment 1 have to be given here.
- Ecological validity: The process model used in experiment 2 was quite realistic and much larger than that of experiment 1. Nevertheless, the effects observed in the second experiment are consistent with those of experiment 1. So, it is most likely that other process models would produce similar results.
- Temporal validity: Each participating student was able to answer the online questionnaire at any time during a specified period of time he/she wanted to. An influence of the time of the experiment (as long as the subjects are not tired) is hardly imaginable.



(a) Aspect *concurrency*.

(b) Aspect *exclusiveness*.

(c) Aspect *order*.

(d) Aspect *repetition*.

Figure 6.15: (Virtual) estimated partial structural process model understandability values of the four aspects depending on coverage rate.

6.6 CONCLUSION

In this chapter, an approach for measuring structural process model understandability was proposed and several hypotheses about effects which have to be considered during measurement were postulated.

First, the importance of measuring process model understandability was motivated. It was distinguished between structural and semantic process model understandability—only structural process model understandability was further considered. Next, an overview of existing measures for structural process model understandability was given. Looking at these measures, serious doubts about their validity arised.

Table 6.22: Data on (virtual) estimated partial structural process model understandability values for the four aspects (Part 1 of 4).

(a) Aspect *concurrency*.

# questions	cov. rate	mean	s. d.	rate lower	rate higher
1	0.02	0.857	0.105	16.7%	27.3%
2	0.03	0.857	0.073	12.9%	25.1%
3	0.05	0.857	0.059	12.2%	11.4%
4	0.06	0.857	0.051	10.0%	11.4%
5	0.08	0.857	0.045	10.2%	5.4%
7	0.11	0.857	0.037	5.3%	2.6%
10	0.15	0.857	0.031	2.8%	1.2%
14	0.21	0.857	0.025	0.7%	0.2%
17	0.26	0.857	0.022	0.1%	0.1%
20	0.30	0.857	0.019	0.0%	0.0%
24	0.36	0.857	0.017	0.0%	0.0%
27	0.41	0.857	0.015	0.0%	0.0%
30	0.45	0.857	0.014	0.0%	0.0%
33	0.50	0.857	0.013	0.0%	0.0%
37	0.56	0.857	0.011	0.0%	0.0%
40	0.61	0.857	0.010	0.0%	0.0%
43	0.65	0.857	0.009	0.0%	0.0%
47	0.71	0.857	0.008	0.0%	0.0%
50	0.76	0.857	0.007	0.0%	0.0%
53	0.80	0.857	0.006	0.0%	0.0%
57	0.86	0.857	0.005	0.0%	0.0%
60	0.91	0.857	0.004	0.0%	0.0%
62	0.94	0.857	0.003	0.0%	0.0%
63	0.95	0.857	0.003	0.0%	0.0%
64	0.97	0.857	0.002	0.0%	0.0%
65	0.98	0.857	0.002	0.0%	0.0%
66	1.00	0.857	–	0.0%	0.0%

Based on a framework for evaluating modeling technique understanding, concrete and detailed definitions for measuring structural process model understandability were given. Using these definitions, hypotheses about effects of measuring structural process model understandability were formulated which have to be considered in the measuring process. Finally, two experiments (experiment 1:

Table 6.22: Data on (virtual) estimated partial structural process model understandability values for the four aspects (Part 2 of 4).

(b) Aspect *exclusiveness*.

# questions	cov. rate	mean	s. d.	rate lower	rate higher
1	0.02	0.934	0.082	16.7%	43.9%
2	0.03	0.934	0.057	14.7%	18.9%
3	0.05	0.934	0.046	12.7%	22.2%
4	0.06	0.934	0.040	10.7%	11.4%
5	0.08	0.934	0.035	8.9%	13.1%
7	0.11	0.934	0.029	5.9%	8.1%
10	0.15	0.934	0.024	2.0%	2.8%
14	0.21	0.934	0.019	0.7%	1.0%
17	0.26	0.934	0.017	0.1%	0.3%
20	0.30	0.934	0.015	0.0%	0.2%
24	0.36	0.934	0.013	0.0%	0.0%
27	0.41	0.934	0.012	0.0%	0.0%
30	0.45	0.934	0.011	0.0%	0.0%
33	0.50	0.934	0.010	0.0%	0.0%
37	0.56	0.934	0.009	0.0%	0.0%
40	0.61	0.934	0.008	0.0%	0.0%
43	0.65	0.934	0.007	0.0%	0.0%
47	0.71	0.934	0.006	0.0%	0.0%
50	0.76	0.934	0.006	0.0%	0.0%
53	0.80	0.934	0.005	0.0%	0.0%
57	0.86	0.934	0.004	0.0%	0.0%
60	0.91	0.934	0.003	0.0%	0.0%
62	0.94	0.934	0.003	0.0%	0.0%
63	0.95	0.934	0.002	0.0%	0.0%
64	0.97	0.934	0.002	0.0%	0.0%
65	0.98	0.934	0.001	0.0%	0.0%
66	1.00	0.934	–	0.0%	0.0%

process model with five tasks, 18 subjects; experiment 2: process model with 12 tasks, 178 subjects) were conducted in order to examine these hypotheses. The results of these experiments are quite consistent.

They support the hypothesis that different aspects of structural process model understandability are of varying difficulty (only *exclusiveness* and *repetition* are

Table 6.22: Data on (virtual) estimated partial structural process model understandability values for the four aspects (Part 3 of 4).

(c) Aspect *order*.

# questions	cov. rate	mean	s. d.	rate lower	rate higher
1	0.01	0.577	0.210	28.8%	47.0%
2	0.02	0.577	0.147	26.2%	33.6%
3	0.02	0.577	0.120	23.3%	26.9%
4	0.03	0.577	0.103	20.7%	22.2%
7	0.05	0.577	0.077	14.9%	16.1%
14	0.11	0.577	0.053	7.2%	6.4%
20	0.15	0.577	0.043	3.7%	3.0%
27	0.20	0.577	0.036	1.6%	1.3%
33	0.25	0.578	0.032	0.7%	0.5%
40	0.30	0.578	0.028	0.3%	0.2%
47	0.36	0.578	0.025	0.1%	0.0%
53	0.40	0.577	0.022	0.0%	0.0%
60	0.45	0.577	0.020	0.0%	0.0%
66	0.50	0.578	0.018	0.0%	0.0%
73	0.55	0.577	0.016	0.0%	0.0%
80	0.61	0.577	0.015	0.0%	0.0%
86	0.65	0.577	0.013	0.0%	0.0%
93	0.70	0.577	0.012	0.0%	0.0%
99	0.75	0.577	0.011	0.0%	0.0%
106	0.80	0.577	0.009	0.0%	0.0%
113	0.86	0.577	0.007	0.0%	0.0%
119	0.90	0.577	0.006	0.0%	0.0%
126	0.95	0.577	0.004	0.0%	0.0%
128	0.97	0.577	0.003	0.0%	0.0%
129	0.98	0.577	0.003	0.0%	0.0%
130	0.98	0.577	0.002	0.0%	0.0%
131	0.99	0.577	0.002	0.0%	0.0%
132	1.00	0.577	—	0.0%	0.0%

quite similar in case of the second and larger process model). Thus, all different aspects have to be measured in order to get a feeling of the “overall structural process model understandability”.

Table 6.22: Data on (virtual) estimated partial structural process model understandability values for the four aspects (Part 4 of 4).

(d) Aspect *repetition*.

# questions	cov. rate	mean	s. d.	rate lower	rate higher
1	0.08	0.944	0.048	0.0%	0.0%
2	0.17	0.944	0.031	0.0%	0.0%
3	0.25	0.944	0.024	0.0%	0.0%
4	0.33	0.944	0.020	0.0%	0.0%
5	0.42	0.944	0.016	0.0%	0.0%
6	0.50	0.944	0.014	0.0%	0.0%
7	0.58	0.944	0.012	0.0%	0.0%
8	0.67	0.944	0.010	0.0%	0.0%
9	0.75	0.944	0.008	0.0%	0.0%
10	0.83	0.944	0.006	0.0%	0.0%
11	0.92	0.944	0.004	0.0%	0.0%
12	1.00	0.944	–	0.0%	0.0%

Furthermore, the hypothesis that asking only a small part of the set of possible questions for one aspect can cause values to differ substantially from the real value was strongly confirmed. Consequently, the coverage rate of asked questions should not be too small. With respect to the larger process model of experiment 2, a coverage rate of 0.25 resulted in less than 1% outliers (higher or lower than 95% confidence interval) for all four aspects. Finally, the asked questions should be selected randomly in order to minimize the risk of choosing particularly easy or difficult questions.

In both experiments, only the aspect *order* was normally distributed. This aspect also had the lowest values of all examined aspects—what is not directly intuitive. Arguably, *concurrency* and *exclusiveness* are more complicated matters than *order*. This fact should be further examined in future work.

Another future issue is the selection of suitable coverage rates which minimize the measuring effort *and* the differences from the real structural process model understandability value. It should be investigated whether the ideal coverage rate is indicated relative or absolute to the process model size and whether it depends on other (structural) process model properties.

Furthermore, it should also be examined whether other aspects of structural process model understandability exist.

7

EFFECTS OF PROCESS MODEL GRANULARITY

7.1 INTRODUCTION

The goal of the process measurement approach of Subsection 3.4.2 is to establish a valid prediction system which can predict the values of an external attribute depending on the values of one or more internal attribute(s). In this chapter, a postulated—yet not validated—prediction system is experimentally evaluated.

During the design phase of a process model (see paragraph *Business Process Management Lifecycle* in Subsection 2.1.2), choosing the adequate size of process activities (process model granularity) is a well-known problem. Vanderfeesten *et al.* have proposed a heuristic for this problem which is inspired by the concepts of *coupling* and *cohesion* in software engineering [129, 168].

In this field of study, the influence of coupling and cohesion on structural software complexity has been examined for some decades (see, for example, [34, pp. 984–985] for a short literature review). The first ideas about coupling and cohesion for the procedural programming paradigm were published in the 1970s under the name "structured design" [149, 179]. Basic coupling and cohesion metrics for the object-oriented paradigm can be found, for example, in the classic Chidamber and Kemerer metrics suite [26]. Empirical evaluations showed the influence of coupling and cohesion metrics on structural software complexity (e. g., [34, 151]).

Motivated by these results from software engineering, Vanderfeesten *et al.* introduced a process model granularity metric. This metric measures the ratio between process model coupling and cohesion. Based on this metric, they suggested a heuristic for selecting between different process model alternatives. It prefers models with high cohesion and low coupling. Vanderfeesten *et al.* also postulated the hypothesis that those process models are less error-prone during process instance execution. As they do not give an empirical validation of their heuristic and hypothesis, it is still no valid prediction system as explained in Subsection 3.4.3.

In this chapter, an experimentation system for analyzing the hypothesis is presented and the results of a conducted experiment with 165 students using this experimentation system are reported. Additionally, an alternative error probability model is suggested which can explain the results of the experiment.

Figure 7.1 shows where the chapter is visually located within the measurement approach of Subsection 3.4.2.

The remainder of this chapter is organized as follows: In Section 7.2, a short introduction into the process model granularity heuristic proposed by Vanderfeesten *et al.* is given. The experimentation system for analyzing a hypothesis about error probability postulated by Vanderfeesten *et al.* is presented in Sec-

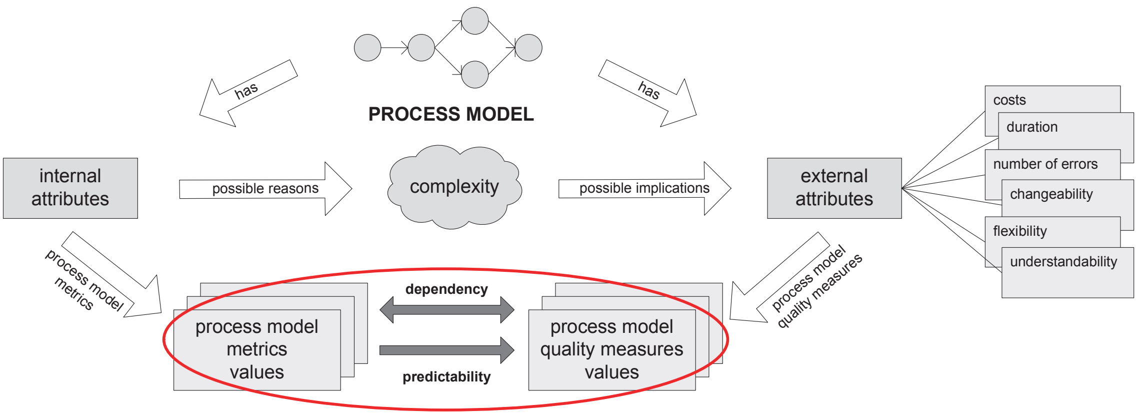


Figure 7.1: Chapter visually located within the measurement approach of Subsection 3.4.2.

tion 7.3. The conducted experiment and its results are shown in Section 7.4. The chapter closes with a conclusion (Section 7.5).

7.2 PROCESS MODEL GRANULARITY HEURISTIC

In this section, the process model granularity heuristic, which was introduced by Vanderfeesten *et al.*, is presented. In Subsection 7.2.1, the underlying process model granularity metric is first explained. The actual heuristic is presented in Subsection 7.2.2.

7.2.1 *Process Model Granularity Metric*

In this chapter, the definitions from [168, pp. 426–429] are used with some modifications:

- In [168], all definitions are based on so-called *operations structures* (see Definition 1 in [168, p. 426]). This process modeling language is almost equivalent to the PDMs of Definition 2.11 presented in Subsection 2.3.2. The operations of a operations structure correspond to the production rules of a PDM. As PDMs are more frequently used, the original definitions of [168] are adapted to PDMs in this chapter.
- In [168], references to resource classes or roles which are able to execute the operations and activities are given. As they are not relevant for the analysis here, they are omitted in this chapter.

Based on a PDM, the contained data elements and production rules are partitioned into different activities.

Definition 7.1 (Activitiy) *An activity* $T \subseteq F$ *based on a PDM* $(D, C, \mathrm{pre}, \mathrm{constr}, \mathrm{cst}, \mathrm{flow})$ *is a set of production rules.*[1]

1 Remember: F is the set of production rules of a PDM $(D, C, \mathrm{pre}, \mathrm{constr}, \mathrm{cst}, \mathrm{flow})$ (see Definition 2.11).

As a shorthand, the notation $\hat{T} := \bigcup_{(p,cs)\in T}(\{p\} \cup cs)$ for the data elements processed in an activity T is introduced.

The different activities can be combined to a process model which processes and computes the data elements of the PDM in a valid sequence. For details on how to specify the control flow or how to check the correctness and soundness of the process model, the reader is referred to [127]. For the purpose here, the following definition is sufficient.

Definition 7.2 (Process model) *A process model[2] based on a PDM* $(D, C, pre, constr, cst, flow)$ *is a set* S *of activities on this PDM.*

Based on these notations, metrics for *process model cohesion* and *coupling* can be defined.

Process model cohesion measures "to what extent operations[3] 'belong' to each other within one activity" [168, pp. 420–421]. It consists of two components.

The first one, *activity relation cohesion*, quantifies how much the production rules of an activity are related. For that purpose, it measures the average overlap of production rules. Two production rules overlap if they share input or output data elements.

Definition 7.3 (Activity relation cohesion) *For an activity* T *based on a PDM* $(D, C, pre, constr, cst, flow)$, *the activity relation cohesion* $\lambda(T)$ *is defined as*

$$\lambda(T) := \begin{cases} \frac{|\{((p_1,cs_1),(p_2,cs_2))\in T\times T|((\{p_1\}\cup cs_1)\cap(\{p_2\}\cup cs_2))\neq\emptyset\wedge p_1\neq p_2\}|}{|T|\cdot(|T|-1)} & \textit{for } |T| > 1 \\ 0 & \textit{for } |T| \leqslant 1 \end{cases} \quad . \tag{7.1}$$

The second cohesion component, *activity information cohesion*, measures which fraction of data elements of an activity are used in more than one production rule.

Definition 7.4 (Activity information cohesion) *For an activity* T *based on a PDM* $(D, C, pre, constr, cst, flow)$, *the activity information cohesion* $\mu(T)$ *is defined as*

$$\mu(T) := \begin{cases} \frac{|\{d\in D|\exists((p_1,cs_1),(p_2,cs_2))\in T\times T:(d\in((\{p_1\}\cup cs_1)\cap(\{p_2\}\cup cs_2))\wedge p_1\neq p_2\}|}{|\hat{T}|} & \textit{for } |\hat{T}| > 0 \\ 0 & \textit{for } |\hat{T}| = 0 \end{cases} \quad . \tag{7.2}$$

The total cohesion of an activity is simply the product of its relation and information cohesion.

Definition 7.5 (Activity cohesion) *For an activity* T *based on a PDM* $(D, C, pre, constr, cst, flow)$, *the activity cohesion* $c(T)$ *is defined as*

$$c(T) := \lambda(T) \cdot \mu(T) \quad . \tag{7.3}$$

2 "Process model" in the nomenclature of this thesis instead of "process" in [168, p. 427].
3 "Production rules" in the nomenclature of this thesis.

The overall cohesion of a process model is computed by the average activity cohesion.

Definition 7.6 (Process model cohesion) *For a process model with a set* S *of activities based on a PDM* $(D, C, \mathrm{pre}, \mathrm{constr}, \mathrm{cst}, \mathrm{flow})$*, the process model cohesion* ch[4] *is defined as*

$$\mathrm{ch} := \frac{\sum_{T \in S} c(T)}{|S|} \quad . \tag{7.4}$$

Process model coupling quantifies how strong the activities of a process model are connected to each other. Two activities are connected if they share at least one data element. The coupling metric measures the fraction of connected activity pairs.

Definition 7.7 (Process model coupling) *For a process model with a set* S *of activities based on a PDM* $(D, C, \mathrm{pre}, \mathrm{constr}, \mathrm{cst}, \mathrm{flow})$*, the process model coupling* cp[5] *is defined as*

$$\mathrm{cp} := \begin{cases} \frac{|\{(T_1,T_2) \in S \times S \mid T_1 \neq T_2 \wedge (\hat{T}_1 \cap \hat{T}_2) \neq \emptyset\}|}{|S| \cdot (|S|-1)} & \textit{for } |S| > 1 \\ 0 & \textit{for } |S| \leqslant 1 \end{cases} \quad . \tag{7.5}$$

Finally, Vanderfeesten *et al.* define a process model coupling/cohesion ratio which serves as a process model granularity metric.

Definition 7.8 (Process model coupling/cohesion ratio) *For a process model with a set* S *of activities based on a PDM* $(D, C, \mathrm{pre}, \mathrm{constr}, \mathrm{cst}, \mathrm{flow})$*, the process model coupling/cohesion ratio* ρ[6] *is defined as*

$$\rho := \frac{\mathrm{cp}}{\mathrm{ch}} \quad . \tag{7.6}$$

7.2.2 *Process Model Granularity Heuristic*

According to Vanderfeesten and Reijers, an important issue in process model design is "the proper size of the individual activities in a process[7] (the process granularity[8])" [129, p. 290]. The heuristic presented in [129, 168] is thought to help designers "to select from several alternatives the process design[9] that is strongly cohesive and weakly coupled" [168, p. 420].

4 "Process model cohesion" in the nomenclature of this thesis instead of "process cohesion" in [168, p. 428].
5 "Process model coupling" in the nomenclature of this thesis instead of "process coupling" in [168, p. 428].
6 "Process model coupling/cohesion ratio" in the nomenclature of this thesis instead of "process coupling/cohesion ratio" in [168, pp. 428–429].
7 "Process model" in the nomenclature of this thesis.
8 "Process model granularity" in the nomenclature of this thesis.
9 "Process model" in the nomenclature of this thesis.

Vanderfeesten *et al.* state that the proposed metrics and the heuristic are inspired by software engineering "where an old design aphorism is to strive for *strong cohesion, and loose coupling*" [168, p. 421].

Consequently, the statement of the heuristic is that a process model with a smaller value of the process model granularity metric (process model coupling/ cohesion ratio) of Definition 7.8 is to be preferred over another one with a larger value. Yet, it does not describe how different alternative process models can be found. [168, p. 429]

Vanderfeesten *et al.* establish the following two hypotheses about the implications of their heuristic [168, pp. 425–426]:

Hypothesis 7.1 *The smaller the value of the process model granularity metric of a process model, the smaller the probability of run-time mistakes.*

Hypothesis 7.2 *The smaller the value of the process model granularity metric of a process model, the larger the understandability and—consequently—the maintainability.*

Instead of an empirical validation of these hypotheses, they only give some arguments as a motivation [168, p. 426]:

- "A *loose coupling* of activities will result in few information elements[10] that need to be exchanged between activities [...], reducing the probability of run-time mistakes."
- "*Highly cohesive* activities [...] are likely to be understood and performed better by people than large chunks of unrelated work being grouped together."

7.3 EXPERIMENTATION SYSTEM

For testing Hypothesis 7.2, experiments in which subjects have to change existing process models (e. g., finding and correcting errors or realizing special modification tasks) would have to be conducted. These experiments would be very time-consuming for the participants and therefore making it hard to find volunteers among students of a university. Consequently, only Hypothesis 7.1 (error probability) is looked at in this section.

For an appropriate experimentation system, three main requirements were identified:

1. automatization of the experiments and the subsequent analysis,
2. comparability of different experiment runs with different process models (consequently, special domain knowledge must not be a necessary requirement and the actual process goal has to be abstracted from—concentrating only on process model structure and granularity) and

10 "Data elements" in the nomenclature of this thesis.

3. cooperation of several subjects with different roles during process instance execution.

A computer-based (cf. requirement 1) experimentation system was created which is described in the remainder of this section.

In this system, very abstract PDMs are used (cf. requirement 2): Each data element represents a single variable of type boolean, integer or double. The production rules are functions with the variables corresponding to the production rule's input data elements as input parameters. According to the variable types, these functions consist of addition, subtraction, multiplication or logical AND, OR, XOR and negation. Activities consist of sets of corresponding functions which can depend on each other in a non-cyclic manner. See Figure 7.4 for an example of such abstract production rules.

The core of the experimentation system is a small web-based workflow engine allowing several subjects to work together on a process instance execution (cf. requirement 3). It is written in Java using Apache Tomcat and runs on a central server. The subjects connect to that workflow engine using a standard web browser.

The workflow engine controls the execution of process instances. Each subject is assigned to a resource role[11]. When an activity becomes executable, it is delegated in first-come, first-served order to the next free subject with the corresponding role. The functions of that activity together with the values of the input parameters of the basic functions[12] are displayed in the web browser on the subject's screen (see Figure 7.2). The subject has to enter the computed values into special text fields. By clicking a button, the computed values are sent to the workflow engine for further processing. At XOR splits, the workflow engine automatically routes by evaluating the boolean constraint expressions for the different branches.

During execution, the following data is logged:

- start and end time of each activity and each process instance,
- correct or incorrect activity execution[13] and
- correct or incorrect process instance execution[14].

7.4 EXPERIMENTAL EVALUATION

In order to test Hypothesis 7.1 (error probability), an experiment using the experimentation system described in Section 7.3 was conducted.

11 Consequently, one needs at least as many subjects as resource roles in the executed process instances.

12 Basic functions are functions for which the values of its input parameters are not computed by other functions of the same activity.

13 The correctness of an activity execution is assessed based on the values of its input parameters. So, if the values of the input parameters are incorrect—caused by an earlier activity—but the output value of the function is correctly computed based on these input values, the activity execution is assessed as correct.

14 A process instance execution is assessed as incorrect if at least one of its activities was executed incorrectly.

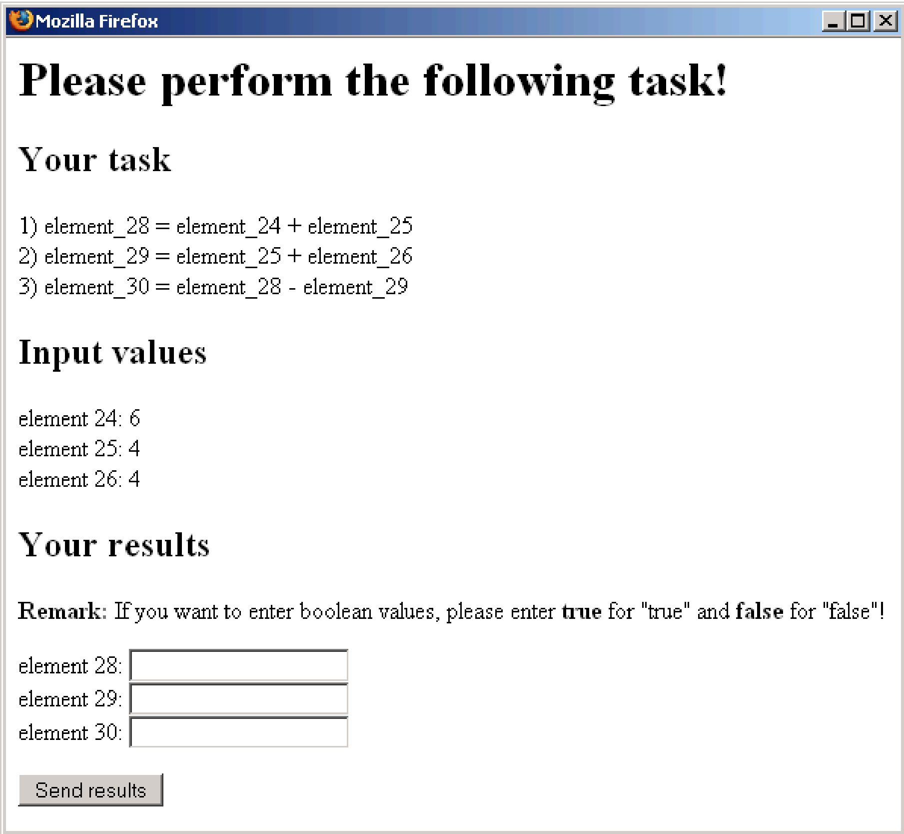


Figure 7.2: Screenshot of a subject's web browser.

In Subsection 7.4.1, the experiment design is explained. The results are presented in Subsection 7.4.2. In Subsection 7.4.3, an alternative error probability model which better explains the results of the experiment is proposed. The section closes with the validity evaluation of the experiment (Subsection 7.4.4).

7.4.1 *Experiment Design*

OBJECT For this experiment, the PDM depicted in Figure 7.3, which is presented as an example in [129, 168], was applied. It was used in the abstract fashion described in Section 7.3. So, only the structural properties—and consequently the process metric values—remained unchanged. The used production rules are shown in Figure 7.4.

Based on the PDM, the three different process model alternatives (Figure 7.5) already proposed in [129, 168] were used. The respective partition into activities is shown in Figure 7.6.

MEASUREMENT INSTRUMENTATION A set of ten process instances was created, which was used for all process model alternatives. All these process instances were executable from the start of the experiment and were processed in the same order for all alternatives. The instances had different values for its basic

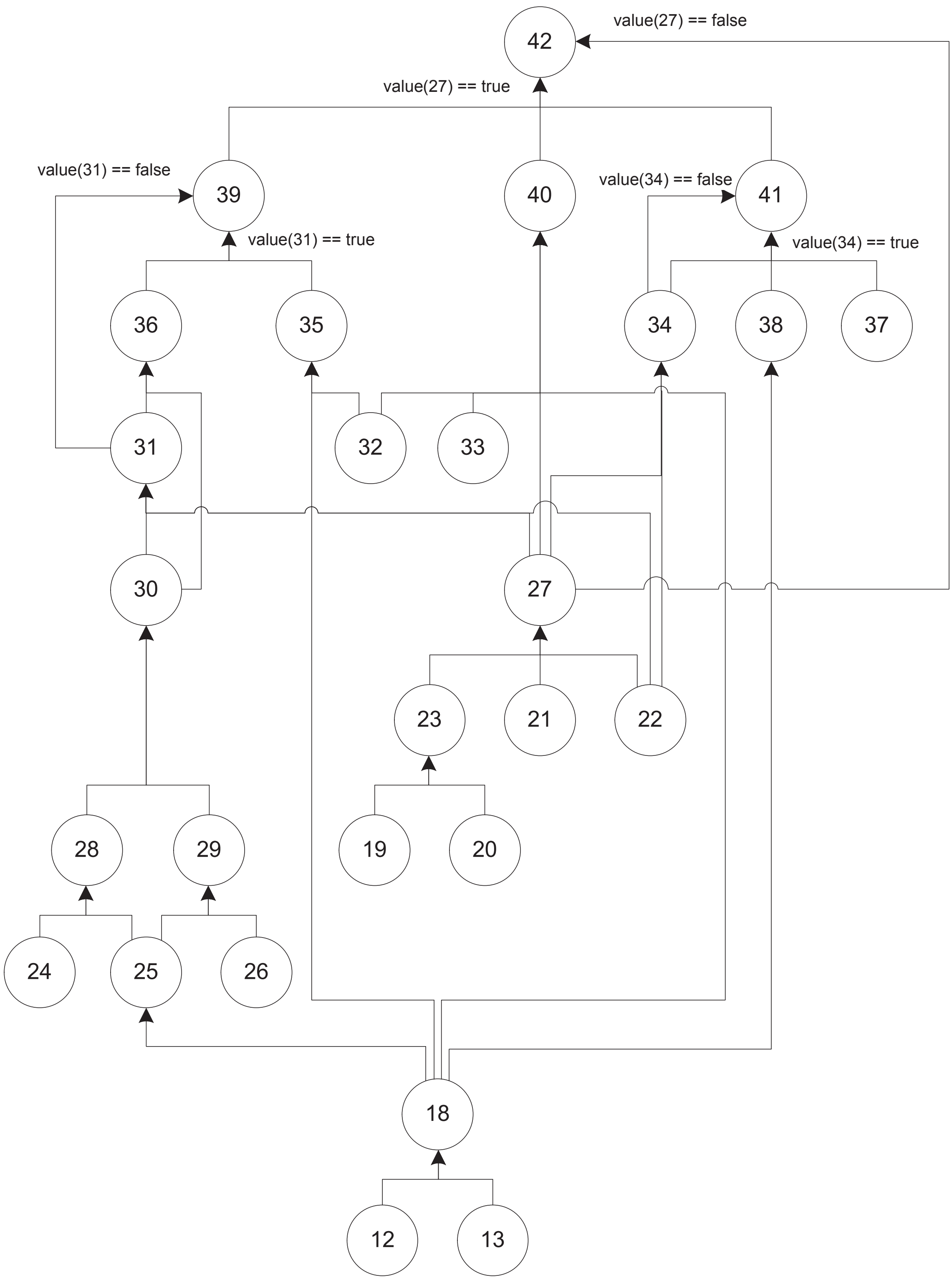


Figure 7.3: PDM used in the experiment [168, p. 422].

data elements[15]. If they were correctly executed, the first and last instances of the set were routed directly from activity C to G at the XOR split—the others had to take the branch with all the other activities.

15 Basic data elements are data elements whose values are not computed by any production rule. Instead, their values have to be given for each process instance before the execution.

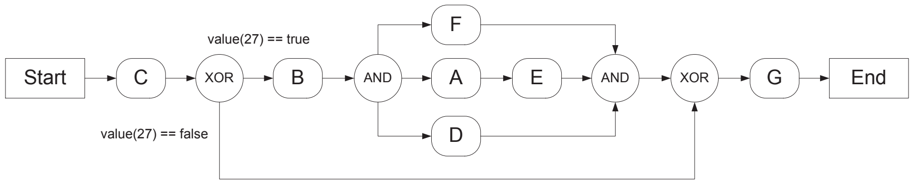
$$
\begin{aligned}
v_i(18) &:= v_i(12) + v_i(13)\\
v_i(23) &:= v_i(19) + v_i(20)\\
v_i(25) &:= v_i(18) - 1\\
v_b(27) &:= v_i(23) > [v_i(21) + v_i(22)]\\
v_i(28) &:= v_i(24) + v_i(25)\\
v_i(29) &:= v_i(25) + v_i(26)\\
v_i(30) &:= v_i(28) - v_i(29)\\
v_b(31) &:= v_b(27) \wedge [v_i(22) \neq v_i(30)]\\
v_b(34) &:= v_b(27) \wedge [v_i(22) \neq 0]\\
v_i(35) &:= v_i(18) + v_i(32)\\
v_b(36) &:= v_b(31) \wedge [v_i(30) > 20]\\
v_i(38) &:= v_i(18) + 1\\
v_b(39) &:= \begin{cases} v_b(31) & \text{for } v_b(31) = \text{FALSE}\\ v_b(36) \vee [v_i(35) \neq 0] & \text{for } v_b(31) = \text{TRUE} \end{cases}\\
v_b(40) &:= n_b(27) \wedge \{[v_i(18) + v_i(32)] > v_i(33)\}\\
v_b(41) &:= \begin{cases} v_b(34) & \text{for } v_b(34) = \text{FALSE}\\ v_b(34) \wedge [v_i(37) > v_i(38)] & \text{for } v_b(34) = \text{TRUE} \end{cases}\\
v_b(42) &:= \begin{cases} \neg v_b(27) & \text{for } v_b(27) = \text{FALSE}\\ [v_b(39) \wedge v_b(40)] \vee v_b(41) & \text{for } v_b(27) = \text{TRUE} \end{cases}
\end{aligned}
$$

Figure 7.4: Production rules used in the experiment. $v_i(18)$ stands for the integer variable representing data element 18, $v_b(27)$ for the boolean variable representing data element 27.

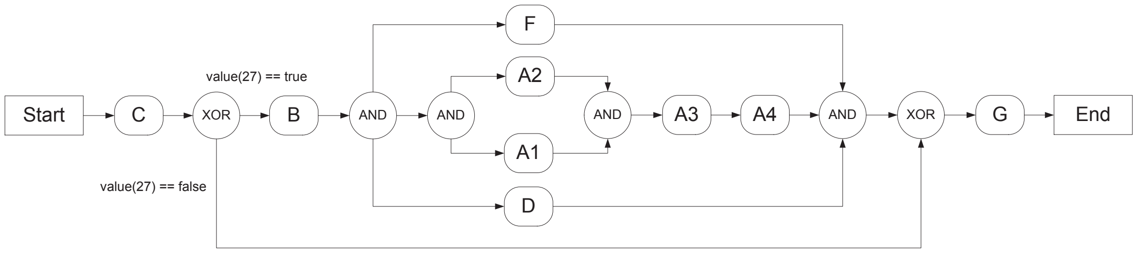
For each process model alternative, several teams were used, which each processed the same set of process instances which was mentioned above. Each team got exactly as many subjects as there are activities in its process model alternative. As the subjects executing activity AE in alternative 3 have much more work than other subjects, two types of teams were used for this alternative to analyze the effect of this possible bottleneck: The first got the "normal" six subjects (number of activities in alternative 3)—the second got seven subjects (two for the resource role of activity AE).

The web-based experimentation system presented in Section 7.3 was used for the experiment. The subjects accessed the system via web browser from different desktop computers (one computer per subject) which were located in a single computer laboratory. As they did not know their other team members, they could only "communicate"[16] via the experimentation system.

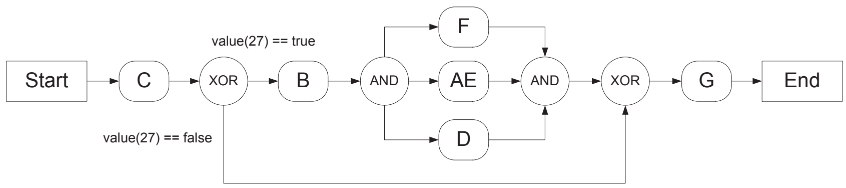
16 As explained in Section 7.3, the subjects could only exchange variable values via the web-based experimentation system. Communication via text messsages was impossible.



(a) Process model alternative 1.



(b) Process model alternative 2.



(c) Process model alternative 3.

Figure 7.5: Three different process model alternatives used in experiment [168, pp. 423–424].

Table 7.1: Metric values for the three process model alternatives.

	cp	ch	ρ
alternative 1	0.714	0.183	3.9
alternative 2	0.611	0.105	5.8
alternative 3	0.8	0.114	7.0

FACTOR The process model granularity of the three different process model alternatives acted as factor in the experiment.

ALTERNATIVES/LEVELS The process metric values for process model coupling (cp), cohesion (ch) and granularity (ρ) of the process model alternatives are listed in Table 7.1. So, there were three levels of the factor process model granularity: 3.9, 5.8 and 7.0.

According to the heuristic, alternative 1 should be preferred as it has the smallest value of ρ. Following Hypothesis 7.1, it should also have the smallest error probability.

RESPONSE VARIABLES The number of incorrectly executed process instances and the number of incorrectly executed activities were used as response variables.

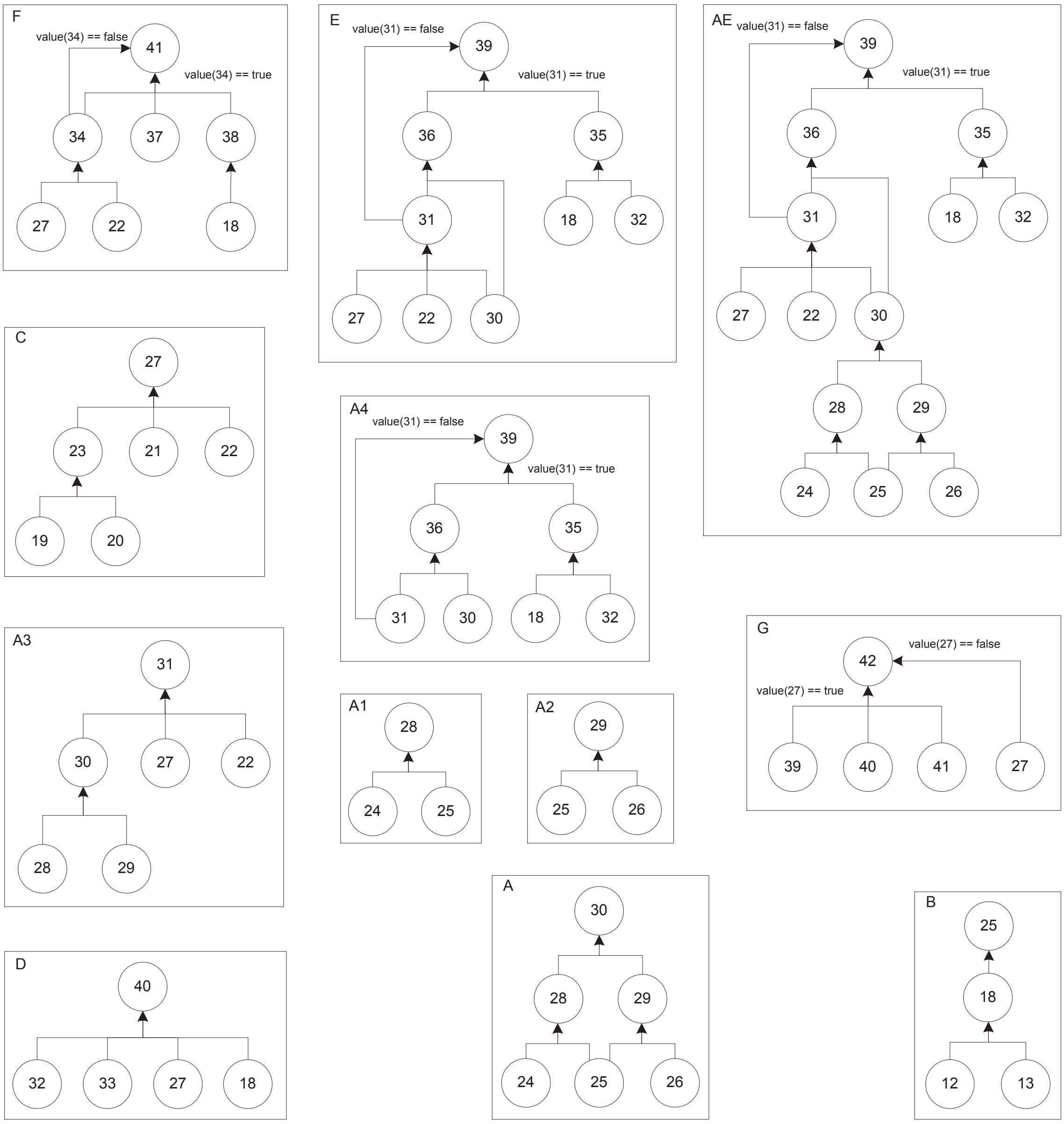


Figure 7.6: Partitioning of the PDM in smaller activities [168, pp. 423–425].

SUBJECTS 165 Business Engineering undergraduate students of the *Universität Karlsruhe (TH)* participated in the experiment. Participation was voluntary—participating students got bonus points for the admission to their final exam. They had no special training in the area of workflows but had the necessary mathematical knowledge for the used abstract functions (cf. Section 7.3). The subjects were randomly assigned to the resource roles within the different teams for the different process model alternatives. Finally, there were six teams for alternative 1, alternative 3 with six subjects and alternative 3 with seven subjects, respectively, as well as five teams for alternative 2.

An overview of the experiment's design is given in Table 7.2.

Table 7.2: Design of the experiment.

process model alternative	# subjects per team	# teams	# subjects
alternative 1	7	6	42
alternative 2	9	5	45
alternative 3a)	6	6	36
alternative 3b)	7	6	42
total number of subjects			165

7.4.2 *Results*

The number of incorrect process instances (over all teams) for the different process model alternatives are shown in Table 7.3.

First, it was checked whether there is a significant difference between alternative 3 with six and seven subjects. For that purpose, Pearson's chi-square test [117, pp. 643–648] was used. The null-hypothesis that the numbers belong to the same distribution could not be rejected on the $\alpha = 0.05$ level. Consequently, both cases were mixed together for the further analysis (row "sum alternative 3" in Table 7.3).

Afterwards, the actual hypothesis was looked at. As one can see in Table 7.3, alternative 1, which should be the best process model alternative according to Hypothesis 7.1, has the highest ratio of incorrect process instances closely followed by alternative 3, which should be the worst. Again, a chi-square test was done to test the alternatives for significant differences. Only for the pair alternative 1 and 2, the null-hypothesis (no difference) could be rejected ($p \approx 0.030$). So, the results of the experiment do not support Hypothesis 7.1.

Next, an analysis on the activity level was conducted. The results of pairwise chi-square tests are shown in Table 7.4. Looking at Table 7.3, one sees that the error probabilities of activities A–AE have exactly the opposite order than predicted by Hypothesis 7.1—even though, there is only a significant difference between alternatives 1 and 3. This was a motivation to further search for alternative factors of influence.

In a next step, the possible influence of the number of data elements and production rules (see Table 7.5) on the error probability of activities (see last row of Table 7.3) was analyzed and depicted in Figure 7.7. For that purpose, both Spearman's rank correlation coefficient (see Section C.2) and Pearson's product-moment correlation coefficient (see Section C.1) were computed. For the number of data elements, one gets 0.95 (Spearman) and 0.78 (Pearson) respectively—as well as 0.97 (Spearman) and 0.85 (Pearson) respectively for the number of production rules. So, roughly speaking, larger activities are more error-prone.

These results are interpreted as follows: Hypothesis 7.1, that process model granularity (a global process model property) influences the error probability

Table 7.3: Error statistics for the different process model alternatives (alternative 3 with six and seven subjects, respectively).

	# incorrect process instances	# incorrect activities C	# incorrect activities B	# incorrect activities F	# incorrect activities D	# incorrect activities G	# incorrect activities A	# incorrect activities E	# incorrect activities A1	# incorrect activities A2	# incorrect activities A3	# incorrect activities A4	# incorrect activities AE	# process instances with at least one of A–AE incorrect
alternative 1	29/60 48.3%	9/60 15.0%	1/41 2.4%	3/41 7.3%	1/41 2.4%	0/60 0.0%	5/41 12.2%	14/41 34.1%	-	-	-	-	-	18/41 43.9%
alternative 2	14/50 28.0%	0/50 0.0%	0/40 0.0%	2/40 5.0%	0/40 0.0%	0/50 0.0%	-	-	0/40 0.0%	1/40 2.5%	2/40 5.0%	10/40 25.0%	-	13/40 32.5%
alternative 3, 6 subjects	26/60 43.3%	3/60 5.0%	3/47 6.4%	9/47 19.1%	1/47 2.1%	7/60 11.7%	-	-	-	-	-	-	9/47 19.1%	9/47 19.1%
alternative 3, 7 subjects	24/60 40.0%	0/60 0.0%	2/48 4.2%	9/48 18.8%	1/48 2.1%	4/60 6.7%	-	-	-	-	-	-	15/48 31.3%	15/48 31.3%
sum alternative 3	50/120 41.7%	3/120 2.5%	5/95 5.3%	18/95 18.9%	2/95 2.1%	11/120 9.2%	-	-	-	-	-	-	24/95 25.3%	24/95 25.3%
sum		12/230 5.2%	6/176 3.4%	23/176 13.1%	3/176 1.7%	11/230 4.8%	5/41 12.2%	14/41 34.1%	0/40 0.0%	1/40 2.5%	2/40 5.0%	10/40 25.0%	24/95 25.3%	

Table 7.4: Results of chi-square tests for error statistics on activity level ($\alpha = 0.05$). For cells marked with "+", the null-hypothesis (no difference) was rejected.

	activity C	activity B	activity F	activity D	activity G	activities A–AE
# alternative 1 vs. 2	+	-	-	-	-	-
# alternative 1 vs. 3	+	-	-	-	+	+
# alternative 2 vs. 3	-	-	+	-	+	-

Table 7.5: Number of data elements and production rules per activity.

	activity C	activity B	activity F	activity D	activity G	activity A	activity E	activity A1	activity A2	activity A3	activity A4	activity AE
# data elements	6	4	7	5	5	6	9	3	3	6	7	14
# production rules	2	2	4	1	2	3	5	1	1	2	4	8

during process instance execution, might not be true. Instead, activity size seems to have a big influence on the error probability of an activity. From a subject's point of view, the remaining process model is some kind of "black box". He/she only sees its own activity with the contained production rules. This fact motivates the following alternative error probability model.

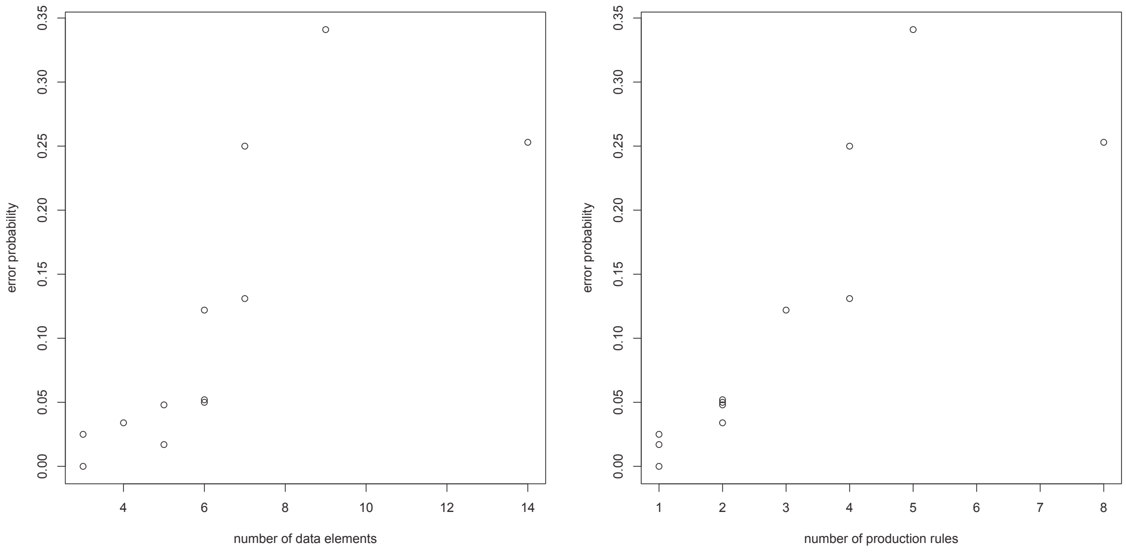
7.4.3 *Alternative Error Probability Model*

If the probabilities p_i that activity i is executed erroneously for a process instance are stochastically independent, then the probability P_{err} that the process instance is executed erroneously is

$$P_{err} = 1 - \prod_i (1 - p_i) \quad . \tag{7.7}$$

If one further assumes, for the sake of simplicity, that all error probabilities p_i of the n activities of a process model are equal with value p, one gets

$$P_{err} = 1 - (1 - p)^n \quad . \tag{7.8}$$



(a) Error probability/number of data elements plot.

(b) Error probability/number of production rules plot.

Figure 7.7: Possible influences on error probability.

Comparing the error probabilities P_{err_A} and P_{err_B} of two alternative process models, one gets the following theorem.

Theorem 7.1 *Given are two alternative process models A and B with n_A and n_B activities, respectively, and the activity error probability p_A and p_B, respectively.*
Then, process model A is more error-prone than process model B ($P_{err_A} > P_{err_B}$) if

1. $p_A > 1-(1-p_B)^{\frac{n_B}{n_A}}$ *or*
2. $n_A > n_B \cdot \frac{\ln(1-p_B)}{\ln(1-p_A)}$.

Proof. Let $p_A, p_B \in (0,1)$ and $n_A, n_B \in \mathbb{N}\backslash\{0\}$.

Regarding 1.)

$$\begin{aligned}
& & P_{err_A} &> P_{err_B} & & \\
&\overset{(7.8)}{\Leftrightarrow} & 1-(1-p_A)^{n_A} &> 1-(1-p_B)^{n_B} & & |-1 \\
&\Leftrightarrow & -(1-p_A)^{n_A} &> -(1-p_B)^{n_B} & & |\cdot(-1) \\
&\Leftrightarrow & \underbrace{(1-p_A)^{n_A}}_{0<\cdot<1} &< \underbrace{(1-p_B)^{n_B}}_{0<\cdot<1} & & |(\cdot)^{\frac{1}{n_A}} \\
&\Leftrightarrow & 1-p_A &< (1-p_B)^{\frac{n_B}{n_A}} & & |-1 \\
&\Leftrightarrow & -p_A &< -1+(1-p_B)^{\frac{n_B}{n_A}} & & |\cdot(-1) \\
&\Leftrightarrow & p_A &> 1-(1-p_B)^{\frac{n_B}{n_A}} & &
\end{aligned}$$

Regarding 2.)

$$\begin{array}{rrcll}
 & P_{err_A} & > & P_{err_B} & \\
\overset{(7.8)}{\Leftrightarrow} & 1-(1-p_A)^{n_A} & > & 1-(1-p_B)^{n_B} & |-1 \\
\Leftrightarrow & -(1-p_A)^{n_A} & > & -(1-p_B)^{n_B} & |\cdot(-1) \\
\Leftrightarrow & \underbrace{(1-p_A)^{n_A}}_{>0} & < & \underbrace{(1-p_B)^{n_B}}_{>0} & |\ln(\cdot) \\
\Leftrightarrow & n_A \cdot \ln(1-p_A) & < & n_B \cdot \ln(1-p_B) & |\div \underbrace{\ln(\underbrace{1-p_A}_{0<\cdot<1})}_{<0} \\
\Leftrightarrow & n_A & > & n_B \cdot \frac{\ln(1-p_B)}{\ln(1-p_A)} &
\end{array}$$

■

If one applies Theorem 7.1 on the more special case that one process model alternative has larger and more error-prone but less activities than the other one, one gets the following corollary.

Corollary 7.1 *Given are two alternative process models A and B. Alternative B has larger and more error-prone ($p_A < p_B$) but less activities ($n_A > n_B$) than alternative A. Then, process model A is more error-prone than process model B ($P_{err_A} > P_{err_B}$) if*

1. $1-(1-p_B)^{\frac{n_B}{n_A}} < p_A < p_B$ *or*
2. $n_A > n_B \cdot \frac{\ln(1-p_B)}{\ln(1-p_A)}$.

Proof. Let $p_A < p_B$ and $n_A > n_B$ with $p_A, p_B \in (0,1)$ and $n_A, n_B \in \mathbb{N}\backslash\{0\}$.

Regarding 1.) The proposition follows from case 1 of Theorem 7.1 together with the precondition $p_A < p_B$.

Regarding 2.) According to case 2 of Theorem 7.1,

$$n_A > n_B \cdot \frac{\ln(1-p_B)}{\ln(1-p_A)} \tag{7.9}$$

holds.

As

$$\begin{array}{rl}
 & p_A, p_B \in (0,1) \text{ and } p_A < p_B \\
\Rightarrow & 0 < 1-p_B < 1-p_A < 1 \\
\Rightarrow & \ln(1-p_B) < \ln(1-p_A) < 0 \\
\Rightarrow & \frac{\ln(1-p_B)}{\ln(1-p_A)} > 1 \quad ,
\end{array}$$

the precondition $n_A > n_B$ is already contained in (7.9). ■

Let us now look at the following example for Corollary 7.1: Alternative B has larger and more error-prone but less activities than alternative A. So, while alternative B has $n_B = n$ activities with error probability $p_B = 0.075$, alternative A has $n_A = 2n$ activities with error probability $p_A = 0.05$. As one can easily check using case 2 of Corollary 7.1, alternative B is less error-prone for all values of n.

Generally, one finds many parameters for this error probability model so that the process model with the larger and more error-prone but less activities has a smaller error probability than the alternative process model.

These findings on the error probability model are consistent with the interpretation of the results of the experiment made above. Hypothesis 7.1 could be wrong. Instead of process model granularity, the size (and consequently the error probability) and the number of activities in a process model could be the main reasons for different error probabilities of alternative process models.

7.4.4 *Validity Evaluation*

Finally, the necessary validity evaluation (see paragraph *Validity Evaluation* in Subsection B.3.3) of the experiment has to be carried out.

INTERNAL VALIDITY Internal validity refers to the fact that the effects observed in the experiment are not caused by a factor which one has no control of or has not measured (see Definition B.9).

Looking at the threats to internal validity mentioned in Subsection B.3.3, one can make the following statements:

- History: During the experiment which lasted less than one hour, no events occurred which were able to strongly influence the subjects.
- Maturation: The experiment was so short that factors as, for example, fatigue, boredom or hunger had no big influence.
- Instrumentation: There was no subjective influence of a human observer of the experiment on the assessment whether a given answer was correct or not.
- Mortality: As all subjects finished the experiment, this threat played no role.
- Selection: As the subjects were randomly assigned to the resource roles, teams and process model alternatives, possible personal differences should have been balanced.

EXTERNAL VALIDITY External validity refers to the extent to which the results of an experiment can be generalized out of the scope of the study (see Definition B.10).

Looking at the threats to external validity mentioned in Subsection B.3.3, one can make the following statements:

- Population validity: The subjects were students with the necessary mathematical knowledge for the used abstract functions—yet had no special training in the area of BPM. As no domain knowledge was needed and even knowledge about process modeling was unimportant, it is believed here to be most likely that other subjects—including professionals in BPM—would show the same effects. For process models with more realistic activities, things could be quite different.
- Ecological validity: The process models used in the experiment had only abstract production rules. So, no special domain knowledge was necessary for the subjects. It is not clear, whether process models with more realistic activities—requiring domain knowledge—would show similar results.
- Temporal validity: An influence of the time of the experiment (as long as the subjects are not tired) is hardly imaginable.

7.5 CONCLUSION

In this chapter, a short introduction into the process model granularity heuristic of Vanderfeesten *et al.* was given. This heuristic is aimed at giving a possible solution to the process model granularity problem (finding the proper size of activities in a process model) during the design phase. According to a hypothesis postulated by Vanderfesten *et al.*, process models with smaller process model granularity metric values (high cohesion and low coupling) are less error-prone during process instance execution.

An experimentation system consisting of a small web-based workflow system for analyzing the hypothesis of Vanderfeesten *et al.* was presented. It uses abstract production rules. Consequently, subjects do not need any domain knowledge. Furthermore, it makes different experiments comparable.

The results of an experiment involving 165 students using the experimentation system were reported. They do not support the hypothesis of Vanderfeesten *et al.*

Instead, an alternative error probability model was suggested which is able to explain the results. According to this model, the error probability of a process model depends on the size (number of data elements and number of production rules) of its activities (larger activities have a higher error probability) and the number of its activities.

The findings of the conducted experiment and the proposed alternative error probability model should be further examined in future work. Furthermore, it should also be tested whether process models with more realistic activities—even if they require some domain knowledge—show the same results related to the hypothesis of Vanderfeesten *et al.*

8 CONCLUSION AND OUTLOOK

Subsection 8.1 provides a conclusion of the results of this thesis. Possible future work in the area of process measurement is presented in Subsection 8.2.

8.1 CONCLUSION

This thesis presented a theoretical framework for process measurement (Chapter 3). It was used to search for open research questions in this field of study. Afterwards, some of the identified questions were dealt with in the thesis (Chapters 4–7).

- **Chapter 3 (Process Measurement)**

  At the beginning of the thesis, an overview of publications on process measurement was given. Many proposed process model metrics are adapted from software metrics and are claimed to measure process model complexity, quality and/or performance. It could be observed that there are no concrete definitions of process model complexity and process model quality in the literature. Often, both terms are even used as synonyms.

  Thus, a discussion of these terms followed. It could be shown that there does not exist a single formal definition of complexity. Instead, numerous aspects of complexity were identified and are analyzed in different research communities. Consequently, it is problematic to say that a process model metric measures the complexity of a process model.

  This discussion resulted in the introduced theoretical framework for process measurement, in which the existing work can be integrated and which can help to identify open research questions leading to new research directions in process measurement.

  For this, the more well-established concepts from software measurement were adopted for process measurement: The result was a prediction system measurement approach, which is based on measurement and prediction systems. The measurement approach consists of process model metrics measuring (structural) internal attributes and process model quality measures measuring external process model attributes. Through this, a concrete definition of process model complexity can be avoided. Nevertheless, process model complexity, quality and performance fit into this measurement approach.

  Furthermore, the necessity for a proper validation of measurement and prediction systems was emphasized. Reliability and validity were identified as important requirements for metrics and measures. Yet, both constructs have not received the necessary attention in process measurement literature so far.

- **Chapter 4 (Analysis of Process Model Metric Properties)**

  In this chapter, an approach for reducing the experimental effort for the validation of prediction systems was introduced.

  Its main idea is to add an additional analysis step before the selection of the prediction system which shall be validated. In this preceding step, the behavior as well as important properties of process model metrics which are part of the potential prediction systems which shall be validated are first analyzed. Through this, unfavorable properties of process model metrics (e. g., insufficient dispersion of metric values or strong correlation with other process model metrics) can be identified before the high effort for an experimental validation of the corresponding prediction system occurs.

  The approach distinguishes between general properties which hold for all process models because of their definition and process model collection specific properties which are only true for the examined process model collection.

  The approach was tested with 33 EPC process model metrics and 515 process models from the SAP Reference Model. During this test, some interesting properties could be found.

  As general properties, mathematical boundaries for the value pairs of some "size metrics" could be identified. Furthermore, it could be shown, that $\Delta(G) \approx \frac{1}{S_N(G)}$ holds.

  Even more process model collection specific properties could be discovered for the SAP Reference Model:

  - The metrics CFC and JC have some few extreme outliers. It is very unlikely that there is a linear dependency between one of these metrics and a process model quality measure. The existence of a threshold over which a process model has some undesirable properties (e. g., high error probability) is much more likely.
  - 94.6% of all process models have a *cyclicity* metric (CYC) value of 0—meaning that they do not contain any (directed) cycle.
  - There are many linear or at least monotonic correlations between the examined process model metrics. This was also confirmed by the result of a PCA. There are three major clusters for the metrics: The first one consists of almost all "size metrics". The second cluster comprises the metrics Δ, Π, Ξ, CP and CC and is clearly separated from the first one. The third (not so cohesive) cluster contains the remaining metrics. Consequently, some metrics do not provide much additional information compared to others.
  - The density metrics Δ and D have quite different behaviors. So, they do not measure the same concept.

- On the other hand, the metrics Δ and CP, which both measure some sort of ratio between existing arcs and possible arcs, are highly correlated. So, it seems that metric CP, which has a much more complicated computation rule, has little additional benefit compared to metric Δ.

- **Chapter 5 (Visualization and Clustering of Process Model Collections)**

  In this chapter, an approach for visualization and clustering of high-dimensional process model metric data of process model collections was proposed.

  First, different visualization techniques were examined for their suitability for visualizing many high-dimensional data points. Next, basic clustering methods were presented.

  The approach comprises

  1. a compact heatmap visualization of the metric data,
  2. a 3D scatter plot visualization of the outcome of a PCA of the data and
  3. a clustered heatmap visualization where the metric data is clustered for (structurally) similar process models within a process model collection.

  The approach was successfully applied to the same EPC process model metrics and process models as in Chapter 4.

  It could be demonstrated that the visualization of 33 process model metric values for 515 process models using heatmaps is possible and still clear for a human observer. Furthermore, the findings on the correlations between process model metrics and on their value ranges which could be gained visually are also consistent with the statistical results of Chapter 4.

  Using the results of a PCA of the process model metric data, it was possible to visualize the data within the three-dimensional coordinate system induced by the first three components of the PCA (comprising more than 71% of the total data variance).

  In contrast, the three clusters of structurally similar process models which were found were not that "spectacular".

- **Chapter 6 (Measuring Structural Process Model Understandability)**

  In this chapter, an approach for measuring structural process model understandability was proposed and several hypotheses about effects which have to be considered during measurement were postulated.

  First, the importance of measuring process model understandability was motivated. It was distinguished between structural and semantic process model understandability—only structural process model understandability was further considered. Next, an overview of existing measures for structural process model understandability was given. Looking at these measures, serious doubts about their validity arised.

  Based on a framework for evaluating modeling technique understanding, concrete and detailed definitions for measuring structural process model understandability were given. Using these definitions, hypotheses about effects of measuring structural process model understandability were formulated which have to be considered in the measuring process. Finally, two experiments (experiment 1: process model with five tasks, 18 subjects; experiment 2: process model with 12 tasks, 178 subjects) were conducted in order to examine these hypotheses. The results of these experiments are quite consistent.

They support the hypothesis that different aspects of structural process model understandability are of varying difficulty (only *exclusiveness* and *repetition* are quite similar in case of the second and larger process model). Thus, all different aspects have to be measured in order to get a feeling of the "overall structural process model understandability".

Furthermore, the hypothesis that asking only a small part of the set of possible questions for one aspect can cause values to differ substantially from the real value was strongly confirmed. Consequently, the coverage rate of asked questions should not be too small. With respect to the larger process model of experiment 2, a coverage rate of 0.25 resulted in less than 1% outliers (higher or lower than 95% confidence interval) for all four aspects. Finally, the asked questions should be selected randomly in order to minimize the risk of choosing particularly easy or difficult questions.

In both experiments, only the aspect *order* was normally distributed. This aspect also had the lowest values of all examined aspects—what is not directly intuitive. Arguably, *concurrency* and *exclusiveness* are more complicated matters than *order*.

- **Chapter 7 (Effects of Process Model Granularity)**

  In this chapter, a short introduction into the process model granularity heuristic of Vanderfeesten *et al.* was given. This heuristic is aimed at giving a possible solution to the process model granularity problem (finding the proper size of activities in a process model) during the design phase. According to a hypothesis postulated by Vanderfesten *et al.*, process models with smaller process model granularity metric values (high cohesion and low coupling) are less error-prone during process instance execution.

  An experimentation system consisting of a small web-based workflow system for analyzing the hypothesis of Vanderfeesten *et al.* was presented. It uses abstract production rules. Consequently, subjects do not need any domain knowledge. Furthermore, it makes different experiments comparable.

  The results of an experiment involving 165 students using the experimentation system were reported. They do not support the hypothesis of Vanderfeesten *et al.*

  Instead, an alternative error probability model was suggested which is able to explain the results. According to this model, the error probability of a process model depends on the size (number of data elements and number of production rules) of its activities (larger activities have a higher error probability) and the number of its activities.

8.2 OUTLOOK

Within the thesis, the following open research questions were identified:

- **Chapter 4 (Analysis of Process Model Metric Properties)**

  As future work, it should be examined whether the identified process model collection specific properties also hold for other collections of process models. Furthermore, the approach could also be applied to other process modeling languages than EPC.

  The results of this chapter may be helpful for planning future validation experiments for prediction systems. Maybe, it can contribute to decrease the lack of validation in this way.

- **Chapter 5 (Visualization and Clustering of Process Model Collections)**

  As a result of the clustering, three clusters of structurally similar process models were found—yet, they were not that "spectacular". It should be examined in future work whether other process model collections have more interesting clusterings.

- **Chapter 6 (Measuring Structural Process Model Understandability)**

  In both experiments, only the aspect *order* was normally distributed. This aspect also had the lowest values of all examined aspects—what is not directly intuitive. Arguably, *concurrency* and *exclusiveness* are more complicated matters than *order*. This fact should be further examined in future work.

  Another future issue is the selection of suitable coverage rates which minimize the measuring effort *and* the differences from the real structural process model understandability value. It should be investigated whether the ideal coverage rate is indicated relative or absolute to the process model size and whether it depends on other (structural) process model properties.

  Furthermore, it should also be examined whether other aspects of structural process model understandability exist.

- **Chapter 7 (Effects of Process Model Granularity)**

  The findings of the conducted experiment and the proposed alternative error probability model should be further examined in future work. Furthermore, it should also be tested whether process models with more realistic activities—even if they require some domain knowledge—show the same results related to the hypothesis of Vanderfeesten *et al.*

A MEASUREMENT FUNDAMENTALS

For Torgerson, the principal objectives of science are the description of empirical phenomena and the establishment of laws and theories which are able to explain observed phenomena and even predict future behavior. Measurement is an essential tool for this process. [153, p. 1]

Also in practice, measurement is important to analyze, compare and eventually optimize things. DeMarco, for example, states: "You can't control what you can't measure." [37, p. 3]

In this thesis, the measurement of properties of processes plays an important role. Chapter 3 deals with process measurement in general. In Chapter 4, proposed process model metrics are presented and their (statistical) properties are analyzed. The measurement of structural process model understandability is the topic of Chapter 6.

This chapter introduces the necessary theoretical fundamentals.

A.1 DEFINITIONS

According to Roberts, a "major difference between a 'well-developed' science such as physics and some of the less 'well-developed' sciences such as psychology or sociology is the degree to which things are measured" [130, p. 1].

Furthermore, he states that even though measurement in physics is usually based on powerful and well-established theories, many practicing physicists take measurement for granted. He points out that putting measurement on a firm foundation is not considered as a terribly important activity in the modern-day physical sciences. [130, p. 3]

In the social sciences, on the other hand, great effort has been put into the establishment of scientific approaches of measuring also non-physical properties such as the intelligence of a person. This area of research is called measurement theory.

If one tries to find a definition of measurement, one comes across several statements which partially differ in detail. In 1940, Campbell wrote [44, p. 340]: "Measurement in its widest sense may be defined as the assignment of numerals to things so as to represent facts or conventions about them." Paraphrasing this definition of Campbell, Stevens formulated in 1946 [148, p. 677]: "[M]easurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules." In 1958 [153, p. 14], Torgerson criticized Stevens' definition by pointing to the fact that not objects itself are measured but certain properties of objects. For him "[m]easurement of a property [...] involves the assignment of numbers to systems to represent that property".

The following definition tries to summarize the statements above.

Definition A.1 (Measurement) *Measurement of an object's (including living beings) or an event's property is the assignment of a (numerical) value according to a special rule in such a way that this value represents the magnitude of that property compared to the magnitudes of that property for other objects or events.*

Looking at this definition, one notices that "measurement has something to do with assigning numbers that correspond to or represent or 'preserve' certain observed relations" [130, p. 50]. Roberts illustrates this with an example [130, pp. 50–51]:

If A is a set of objects and the binary relation $H(a_1, a_2)$ holds if one judges a_1 to be heavier than a_2, one wants to assign a real number $\mu(a)$ to each $a \in A$ such that $\forall a_1, a_2 \in A$

$$H(a_1, a_2) \Leftrightarrow \mu(a_1) > \mu(a_2) \tag{A.1}$$

holds.

Furthermore, the "measure" shall be "additive" in the following manner: If one "combines" two objects a_1 and a_2 (binary operation $\bullet$), the function μ on A shall also "preserve" the binary operation $\bullet$ such that $\forall a_1, a_2 \in A$

$$\mu(a_1 \bullet a_2) \Leftrightarrow \mu(a_1) + \mu(a_2) \tag{A.2}$$

holds.

In order to express these requirements in a formal mathematical way, one needs the following definitions.

First, the term "relational system" has to be defined [130, p. 51].

Definition A.2 (Relational system) *A relational system $\mathcal{A}$ is an ordered $(1 + p + q)$-tuple $\mathcal{A} = (A, R_1, R_2, \ldots, R_p, \bullet_1, \bullet_2, \ldots, \bullet_q)$ where A is a set, $R_1, R_2, \ldots, R_p$ are (not necessarily binary) relations on A and $\bullet_1, \bullet_2, \ldots, \bullet_q$ are (binary) operations on A.*

The type *of the relational system is a sequence $(r_1, r_2, \ldots, r_p; q)$ of length $p + 1$ where $r_i = m$ if R_i is an m-ary relation.*

For the purpose of measurement, two different relational systems are necessary: an *empirical relational system* and a *formal relational system*.

The empirical relational system [130, p. 51] represents that part of reality which one wants to measure. In the example above, the empirical relational system is $\mathcal{A} = (A, H, \bullet)$ of type $(2; 1)$ with binary relation H and (binary) operation $\bullet$.

The formal relational system [180, p. 40] represents the system with which the "magnitude" of the measured property is expressed. In the example above, the formal relational system is $\mathcal{B} = (\mathbb{R}, >, +)$ of type $(2; 1)$. If the set $\mathbb{R}$ of real numbers is part of the formal relational system, the term *numerical relational system* is used [130, p. 51].

The function μ which maps an object from an empirical relational system to an object of a formal relational system (or a numerical value of a numerical relational system) is called *measure* [180, p. 40].

Definition A.3 (Measure) *A measure* μ *is a mapping* $\mu : A \mapsto B$ *from an empirical relational system* $\mathcal{A}$ *into a formal relational system* $\mathcal{B}$ *which yields for ever empirical object* $a \in A$ *a formal object (measurement value)* $\mu(a) \in B$.

If one wants the measure to "preserve" the properties of the empirical relational system, the "mapping may not be arbitrary" [180, p. 40]. Finally, this leads to the definition of a *scale* [130, pp. 51–52, 54] [180, pp. 40–41].

Definition A.4 (Scale) *Let* $\mathcal{A} = (A, R_1, R_2, \dots, R_p, \bullet_1, \bullet_2, \dots, \bullet_q)$ *be an empirical relational system,* $\mathcal{B} = (B, R_1^{'}, R_2^{'}, \dots, R_p^{'}, \bullet_1^{'}, \bullet_2^{'}, \dots, \bullet_q^{'})$ *a formal relational system of the same type as* $\mathcal{A}$ *and* μ *a measure from* $\mathcal{A}$ *into* $\mathcal{B}$.

The function $\mu : A \mapsto B$ *is called a* homomorphism *from* $\mathcal{A}$ *into* $\mathcal{B}$ *if* $\forall a_1, a_2, \dots, a_{r_i} \in A, \forall i \in \{1, \dots, p\}$

$$R_i(a_1, a_2, \dots, a_{r_i}) \Leftrightarrow R_i^{'}(\mu(a_1), \mu(a_2), \dots, \mu(a_{r_i})) \tag{A.3}$$

and $\forall a_1, a_2 \in A, \forall i \in \{1, \dots, p\}$

$$\mu(a_1 \bullet_i a_2) = \mu(a_1) \bullet_i^{'} \mu(a_2) \tag{A.4}$$

hold.

Then, the triple $(\mathcal{A}, \mathcal{B}, \mu)$ *is a* scale.

So, the fundamental problem of measurement is to find a scale for each property one wants to measure. Especially for non-physical properties, it is hard to find scales in such a way that it is generally accepted that they "preserve" the original properties of the empirical relational system.

Consequently, the term "scale" is only rarely used. Instead, one falls back to the term "measure" (Definition A.3) whose requirements are not as strict as for "scale".

Even though the term "measure" would be correct, the term "metric" is (more) often used in process measurement (e. g., [6, 19, 21–24, 54, 97–100, 129, 165–168]). Zuse already realized that for the area of software measurement [180, pp. 28–29].

Mathematically spoken, the term "metric" is incorrect in the context of process measurement as it is used in mathematics to describe a kind of "distance measure" between vectors of a more-dimensional space (see the following definition [170]).

Definition A.5 (Metric (mathematics)) *Let* X *be an arbitrary set. A function* $d : X \times X \mapsto \mathbb{R}$ *is called a metric on* X *if for all* $x, y, z \in X$ *the following three conditions are fulfilled:*

$$d(x, y) = 0 \Leftrightarrow x = y \qquad \textit{(identity of indiscernibles)} \tag{A.5}$$

$$d(x, y) = d(y, x) \qquad \textit{(symmetry)} \tag{A.6}$$

$$d(x, z) \leqslant d(x, y) + d(y, z) \qquad \textit{(triangle inequality)} \tag{A.7}$$

Corollary A.1 *Let* d *be a metric on set* X. *Then,*

$$d(x, y) \geqslant 0 \qquad \textit{(non-negativity)} \tag{A.8}$$

holds for all $x, y \in X$.

Proof. Let d be a metric on set X. For all $x, y \in X$, it holds:

$$2 \cdot d(x,y) = d(x,y) + d(x,y) \overset{(A.6)}{=} d(x,y) + d(y,x) \overset{(A.7)}{\geqslant} d(x,x) \overset{(A.5)}{=} 0 \Rightarrow d(x,y) \geqslant 0$$

■

In order to be consistent with the majority of publications in this area, the term "metric" is used in this thesis for a measure of an internal (structural) attribute (see Subsection 3.4.2) of a process model (cf. the metrics in Chapter 4, 5 and 7). For external attributes (e. g., structural understandability in Chapter 6), the term "measure" is used.

A.2 HIERARCHY OF SCALE TYPES

In [148], Stevens proposes a hierarchical classification system for (a subset of) scales which is based on the definition of admissible transformations [130, p. 58].

Definition A.6 (Admissible transformation) *Let* $(\mathcal{A}, \mathcal{B}, \mu)$ *be a scale,* A *the set underlying* $\mathcal{A}$ *and* B *the set underlaying* $\mathcal{B}$*. A function* $\Phi : \mu(A) \mapsto B$ *is called an* admissible transformation *if and only if* $(\mathcal{A}, \mathcal{B}, \Phi \circ \mu)$ *is also a scale.*

Stevens' five scale types are induced by five corresponding sets of admissible transformations (see Table A.1). From top to bottom of the table, the requirements for these transformations become harder. Thereby, the set of admissible transformations of a scale type is a subset of all sets of admissible transformations of the scale types which are above in Table A.1. Consequently, scales of a special scale type also have all scale types "higher" in the table. Furthermore, each scale type has a set of basic empirical operations which can be used for comparing the empirical property which is measured.

A *nominal scale* has the scale type with the weakest requirements. On such a scale, only the (in)equality of the measured property of two or more objects can be determined. According to Stevens [148, p. 678], two types of nominal assignment can be distinguished: (1) numbering each object with a distinct number for identification (e. g., shirt numbers in team sports) or (2) numbering each object which is member of a class with the same number (e. g., sex—0 for male and 1 for female). Actually, the first case is a special case of the second one where each class has exactly one member.

Using an *ordinal scale*, objects can be ordered according to the size of their measured property. An example are the rank numbers in sports league tables. The smaller the number, the better the team played in that season. Yet, there is no information about the "distances" between the different ranks. So, the question "Is the 'distance' between the teams on rank 1 and 2 as large as between rank 2 and 3?" cannot be answered with this scale type.

An *interval scale* provides this information about "distances". Examples are temperature measured in Celsius or Fahrenheit. Here, differences between scale

Table A.1: Hierarchy of different scale types [148, p. 678] [130, p. 64].

scale type	admissible transformations	basic empirical operations	examples
nominal	any one-to-one Φ	determination of • equality	• shirt numbers in team sports • sex
ordinal	any strictly monotone increasing Φ	determination of • equality • greater or less	rank numbers in sports league tables
interval	$\Phi(x) = \alpha x + \beta \quad , \alpha > 0$	determination of • equality • greater or less • equality of intervals or differences	temperature in Celsius or Fahrenheit
ratio	$\Phi(x) = \alpha x \quad , \alpha > 0$	determination of • equality • greater or less • equality of intervals or differences • equality of ratios	absolute temperature in Kelvin
absolute	$\Phi(x) = x$	determination of • equality • greater or less • equality of intervals or differences • equality of ratios	counting

values can be compared. So, one can say, for example, that the temperature difference between 10 °C and 20 °C is as large as between 20 °C and 30 °C.

A *ratio scale* has a defined and meaningful zero point. So, also ratios of scale values can be compared. An example is absolute temperature measured in Kelvin. Here, the zero point is defined as the lowest physically possible temperature.

For an *absolute scale*, finally, only the identity function is allowed as admissible transformation. Counting is an example for this scale type.

These scale types do not only allow to classify scales—they also imply which statistical operations can legitimately be applied to empirical data [148, p. 677].

In order to discuss this in more detail, the following definition of *meaningfulness* [130, p. 59] is necessary.

Definition A.7 (Meaningfulness) *A statement involving (numerical) scales is* meaningful *if and only if its truth or falsity is unchanged under admissible transformations of all the scales in question.*

According to Stevens, a statistical operation is appropriate for a scale type if it is invariant (meaningful) under the admissible transformations of this scale type [148, p. 678].

Table A.2 shows a fraction of the meaningful statistical operations of the different scale types. Only those statistical operations are listed which are used in this thesis (in Chapter 4). These are

- median [117, pp. 19–21], quantiles [117, pp. 25–26] and mean [117, pp. 16–17] as *measures of location*,

- range [117, p. 22], interquartile range [156, p. 55], median absolute deviation[1], standard deviation [117, p. 22, 35–36] and coefficient of variation [117, pp. 33–34] as *measures of dispersion* as well as

- Spearman's rank correlation coefficient (see Section C.2) and Pearson's product-moment correlation coefficient (see Section C.1) as *measures of correlation*.

The assignment of the statistical operations to the scale types according to meaningfulness shall be illustrated using three examples given by Stevens [148, p. 678].

The scale value at the median of a distribution maintains its position under all transformations which preserve order. So, the statistical operation *median* is meaningful for all scale types except nominal.

In contrast, a scale value at the mean of a distribution remains at the mean position only for linear transformations. Consequently, computing the *mean* is only meaningful for values measured at least on an interval scale.

The ratio expressed by the *coefficient of variation* only allows the admissible transformations of at least a ratio scale.

1 median of the set of absolute values of the differences between the values and the median of these values

Table A.2: Overview of different scale types and a fraction of their meaningful statistical operations [148, p. 678] [134, p. 55].

scale type	measures of location	measures of dispersion	measures of correlation
nominal			
ordinal	• median • quantiles		• Spearman's rank correlation coefficient
interval	• median • quantiles • mean	• range • interquartile range • median absolute deviation • standard deviation	• Spearman's rank correlation coefficient • Pearson's product-moment correlation coefficient
ratio	• median • quantiles • mean	• range • interquartile range • median absolute deviation • standard deviation • coefficient of variation	• Spearman's rank correlation coefficient • Pearson's product-moment correlation coefficient
absolute	• median • quantiles • mean	• range • interquartile range • median absolute deviation • standard deviation • coefficient of variation	• Spearman's rank correlation coefficient • Pearson's product-moment correlation coefficient

A.3 MEASUREMENT OF NON-PHYSICAL PROPERTIES

While defining generally accepted measures for physical properties (e. g., length, mass, temperature, etc.) is relatively easy in physics, the same purpose is much harder for non-physical properties as, for example, intelligence, religiosity, prejudice, etc. in psychology and sociology.

In Subsection A.3.1, the general procedure for finding adequate measures is introduced. Subsection A.3.2 presents necessary properties which these measures have to fulfill.

A.3.1 *Conceptualization, Operationalization and Measurement*

This subsection is based on parts of a textbook by Babbie [5, pp. 118–140]. There, one can also find further details.

The fundamental problem when measuring non-physical properties (e. g., the magnitude of prejudices of a person) is that these properties themselves do not exist in reality. Instead, these terms represent mental images in the brains of humans summarizing certain related observations and experiences which these persons have made during their lives. The terms are then used to communicate the corresponding mental constructs.

To illustrate this fact, one can think of the term "prejudice" as an example. Presumably, everybody has made certain of the following observations and experiences during his/her life or has—at least—heard about them:

- People saying nasty things about minority groups.
- People believing that women are inferior to men and that they should stay at home, care for their children and do the housework instead of working in a job.
- People beating foreigners.

Even if these observations and experiences differ in detail, they are somehow related. Consequently, people get some mental image about the underlying abstract construct in their brains and associate it with a corresponding term which they learn by others (e. g., their parents). The technical term for those mental images is *conception* [5, pp. 120, 122]

Definition A.8 (Conception) *A conception is a person's* subjective *mental image of a non-physical property based on his/her observations and experiences.*

As everybody has made different observations and experiences, conceptions are individual and differ in detail. Communication between two persons is only possible if their mental images (conceptions) behind a term overlap. For research purposes about such a term, however, one needs an exact definition. [5, p. 121]

The process of defining the exact meaning of a term used during a research study is called *conceptualization* [5, pp. 120, 122] with a *concept* [5, pp. 120, 122] as its result.

Definition A.9 (Conceptualization) *Conceptualization is the process through which the exact meaning of a non-physical property is specified for a specific study.*

Definition A.10 (Concept) *A concept is a construct derived by mutual agreement from mental images (conceptions) during conceptualization. It specifies the meaning of the non-physical property to be measured in the course of a specific study.*

Thereby, it is important to understand that different researchers could come to different "definitions" (concepts) for their studies for the same term.

During conceptualization, one can possibly identify different aspects of a concept. The technical term is *dimension* [5, p. 123].

Definition A.11 (Dimension) *A dimension is a specifiable aspect of a concept.*

Once again, think of the example "prejudice". Possible dimensions of the concept "prejudice" are prejudices with regard to race, gender, religion and sexual orientation [5, p. 125].

Conceptions and concepts are mental constructs—not existing in reality themselves [5, p. 122]. Thus, quantifiable "representations" of these mental constructs have to be found in reality during conceptualization. These are called *indicators* [5, p. 123].

Definition A.12 (Indicator) *An indicator is an observable sign of the presence or absence of the concept which is to be measured.*

Coming back to the "prejudice" example, one could create a questionnaire consisting of statements such as "Women are inferior to men.", "There should be no same-sex marriages." and "Foreigners take away our jobs." and ask the subjects to specify whether they support these statements or not.

The last step for finding a measure for a non-physical property is called *operationalization* [5, pp. 125, 132].

Definition A.13 (Operationalization) *Operationalization is the process of defining an exact research procedure (operations) for measuring a non-physical property based on a concept for that property and indicators.*

Operationalization provides a link between a concept as a mental construct not existing in reality and observable and quantifiable things in reality (indicators) [5, pp. 122, 132].

The operationalization for the "prejudice" example could look like this: After having decided which dimensions of the concept "prejudice" one wants to measure, a questionnaire with different statements (indicators) as described above is created. The subjects whose magnitudes of prejudices are to be measured specify which statements they support. Afterwards, the sum of supported statements is computed for every subject. This value is used as measure of the magnitude of a person's prejudices.

The conceptualization and operationalization have to be undertaken at the beginning of any study design before the study itself (e. g., using a questionnaire) is conducted and analyzed. Otherwise, the results would not be interpretable because important terms of the inquiry would not have been defined exactly. [5, p. 126]

In order to summarize the procedure, one can state that the purpose of conceptualization and operationalization is the creation of an *exact "working definition"* of a non-physical property *in the context of a given study*. As the definition is absolutely specific and unambiguous, others can interpret the results—even if they disagree with this definition. [5, pp. 125–126]

So, the inquiry is traceable and repeatable—an essential requisite of a scientific approach.

A.3.2 *Criteria of Measurement Quality*

When measuring non-physical properties as described in the previous subsection, there are two important criteria of measurement quality: reliability and validity.

Reliability

Babbie gives the following definition for reliability [5, p. 141].

Definition A.14 (Reliability) *Reliability is that quality of a measurement method that suggests that the same data would have been collected each time in repeated observations of the same phenomenon.*

The two subsequent examples shall help to explain the term. For easier understandability, the examples are taken from measuring physical properties.

Imagine, for example, that one is interested in a person's weight. One possible technique could be to ask two different people to estimate the person's weight. Presumably, these estimates would vary a lot. As an alternative, one could use a bathroom scale for measuring the person's weight. If the person steps on the scale twice, it would likely report the same weight in both cases. So, the second method is much more reliable than the first one. [5, p. 141]

Kan gives another good example [69, pp. 70–71]: If one wants to measure the body height of a person, one gets more reliable values if one uses a precise operationalization which specifies the time of the day to take the measurement, the scale to use, who takes the measurement, whether the measurement should be taken barefooted and so on.

Frequent threats to reliability are:

- impact of observer's subjectivity is higher than that of observed situation itself [5, pp. 141–142]
- influence of interviewer on answers given by respondents [5, p. 142]
- imprecise operationalization of measuring method (cf. body height example)

Validity

Validity is defined by Babbie [5, p. 143] as follows:

Definition A.15 (Validity) *Validity refers to the extent to which a measure adequately reflects the real meaning of the concept under consideration.*

As a concept is only a construct derived by mutual agreement (see Definition A.10), there is no actual "real meaning". Consequently, the ultimate validity of a measure can never be proven. It may only be agreed to its relative validity on the basis of four "aspects" of validity: face, criterion-related, construct and content validity. [5, p. 143]

Babbie gives the following definitions and examples [5, pp. 144–145]:

Definition A.16 (Face validity) *Face validity is that quality of an indicator that makes it seem a reasonable measure of some variable.*

Even though one might discuss whether there exist better indicators or not, it seems to make sense without a lot of explanation that the frequency of church attendance is some indication of a person's religiosity—the measure is valid "on its face".

Definition A.17 (Criterion-related validity) *Criterion-related validity (also called predictive validity) is the degree to which a measure relates to some external criterion.*

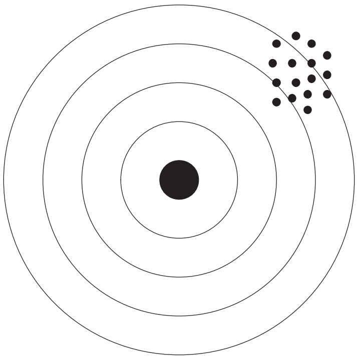
The criterion-related validity of a written driver's test is determined by the relationship between the obtained test score and the subsequent driving ability.

Definition A.18 (Construct validity) *Construct validity is the degree to which a measure relates to other variables as expected within a system of theoretical relationships.*

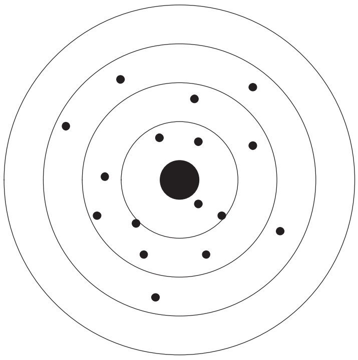
Imagine one wants to study the influence of marital satisfaction on the probability of cheating the spouse. For that purpose, a measure for marital satisfaction is developed whose validity has to be assessed. Furthermore, the hypothesis that marital satisfaction decreases the probability of cheating the spouse is made. If the measure relates to this expectation, that constitutes evidence for its construct validity.

Definition A.19 (Content validity) *Content validity is the degree to which a measure covers the range of meanings included within a concept.*

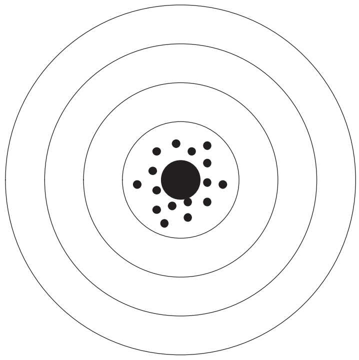
A test which wants to measure mathematical ability cannot be limited to addition. It also needs to cover subtraction, multiplication, division, etc. Otherwise, it has a lack of content validity.



(a) Reliable but not valid.



(b) Valid but not reliable.



(c) Reliable and valid.

Figure A.1: A graphical analogy of the difference between reliability and validity [5, p. 145].

Graphical Analogy

Babbie gives a good graphical analogy of the difference between reliability and validity (see Figure A.1) [5, p. 145]: He proposes to think of measurement as analogous to repeatedly shooting at the bull's-eye on a target. Then, reliability looks like a tight pattern of shots somewhere on the target (see Figure A.1a)—validity is a set of shots arranged around the bull's-eye (see A.1b). A failure of validity reveals itself as shots which are systematically off the mark (see Figure A.1a)—a failure of reliability as randomly distributed shots around the target (see Figure A.1b). Only reliable and valid metrics (see Figure A.1c) are useful.

B

BASICS OF EMPIRICAL RESEARCH

In this thesis, controlled experiments are used twice for examining effects of measuring structural process model understandability (Chapter 6) and effects of process model granularity (Chapter 7). This chapter gives the necessary theoretical background to the general need of empirical research and conducting controlled experiments in particular. It is mainly based on textbooks by Juristo and Moreno [68] as well as Wohlin *et al.* [178].

B.1 MOTIVATION

The area of BPM has developed and proposed a wide range of modeling languages (see Section 2.3), tools, methods and process model metrics. In many publications, additional hypotheses about the alleged superiority of a special modeling approach, a tool, etc. are being made. Often, the authors present some motivation why the postulated hypothesis shall be correct—yet, a stringent validation is often missing (see Section 3.6).

Juristo and Moreno emphasize the importance of validation in scientific research [68, p. 24]: "For a body of knowledge to be considered scientific, its truth and validity must be proven. A particular item of knowledge is considered to be scientifically valid if it has been checked against reality. [. . .] Scientific research is the antithesis of opinion. Ideally, researchers do not opine, they explain objective results. Their studies are not based on subjective factors, like emotions, opinions or tastes. Scientific investigations are objective studies, based on observations of or experimentation with the real world and its measurable changes."

As most aspects of BPM cannot be mathematically modeled, a mathematical proof of a hypothesis is often not possible[1].

Furthermore, one has to state that BPM is a human-intensive discipline. So, it "can be considered as a social process in that the artifacts (methods/tools/paradigms) to be used are affected by the experience, knowledge and capability of the user"[2] [68, p. 26].

These facts show the need for empirical studies which "have traditionally been used in social sciences and psychology, where we are unable to state any laws of nature, as in physics" [178, p. 5].

1 There are exceptions: Some properties of process models (e. g., soundness, reachability, existence or absence of deadlocks) are mathematically provable.

2 Originally, Juristo and Moreno write about software engineering. Yet, their statement is transferable to BPM.

B.2 EMPIRICAL APPROACHES

There are several empirical approaches—each with its special advantages and disadvantages. Consequently, one has to select the adequate approach with regard to the question which is to be examined.

In this section, three approaches (surveys, case studies and controlled experiments) are presented in more detail. At the end of the section, these approaches are compared according to their properties.

B.2.1 *Surveys*

Surveys are often done in retrospect, after having used a (new) technique for some time in order to get "a snapshot of the situation to capture the current status" [178, p. 10].

They consist of a set of questions which a group of persons—representatives from the population to be studied—is asked to answer. Afterwards, it is tried to generalize the results to the whole population.

Using surveys, one can collect data about *subjective* experiences, opinions, attitudes, etc. Yet, it is not possible to get *objective* values as, for example, the number of errors or the execution time of a process.

The questions of a survey can be asked either using (online) questionnaires or by conducting interviews. The advantages of (online) questionnaires are that the researcher is not needed while the respondents are answering the questions and the time for answering can be chosen within a specific period of time. Interviews (by telephone or face-to-face), on the other side, also have some advantages: Typically, the response rates are higher for interviews than for (online) questionnaires. Furthermore, the number of "do not know" and "no answer" are lower as the interviewer can answer possible questions of the respondents. [178, pp. 8, 10–12]

B.2.2 *Case Studies*

Case studies are conducted to investigate an attribute or phenomenon during a specific period of time. They are *observational* studies done by observing an on-going project or activity.

Due to this, they are especially suitable for industrial evaluations as these investigations can be conducted as side actions of (commercial) projects running anyway.

A big disadvantage is the lack of control: The exact circumstances of the project are not arbitrarily manipulable. This fact makes it also difficult to compare the effects of, for example, different techniques or tools as it is hard to conduct case studies in several quite similar projects which mainly differ only in this desired factor. [178, pp. 7–8, 12–14]

B.2.3 *Controlled Experiments*

Controlled experiments are *controlled* studies—compared to case studies which are "only" *observed* studies [178, p. 8]. They are used for analyzing the outcomes of different treatments while all other factors remain unchanged, i. e., the objective is to manipulate one or more variables and control all other variables at fixed levels, measure the effect of the manipulation and finally analyze these results with statistical methods.

Controlled experiments are usually done in a laboratory environment which provides a high level of control. Specific experimentation systems are often created for an experiment in order to produce the desired circumstances. The participating subjects are randomly assigned to the different tested treatments. [178, pp. 9, 14–15]

B.2.4 *Comparison of the Approaches*

Each approach has its advantages and disadvantages. So, the choice of the empirical approach "depends on the prerequisites for the investigation, the purpose of it, available resources and how we would like to analyze the collected data" [178, p. 17].

The presented approaches can be compared using the following three factors [120, p. 19] [178, pp. 16–17].

- **Level of control**

  The main goals of empirical approaches are to gain insights into dependencies, influencing factors and to examine proposed hypotheses. So, it is necessary that "interesting" variables (i. e., variables which are potential influencing factors) are well adjustable to desired values and that the whole investigation is controllable by the experimenter according to his/her wishes.

- **Level of replication**

  A very important requirement of empirical approaches is the possibility of replication—either by the same experimenter or by other researchers. Only this way, the confidence in the observed results can grow.

  A necessary prerequisite for replication is that the investigation can be repeated under the same conditions. That means that all important conditions of the first run are determinable and re-adjustable for further runs.

- **Investigation costs**

  The investigation costs of an empirical approach are comprised of the required time for preparation, execution and final evaluation of the investigation and the required financial resources. These financial costs are mainly caused by labor costs of the persons involved in the investigation. The labor

Table B.1: Comparison of the empirical approaches [120, p. 19][178, p. 16].

factor	survey	case study	controlled experiment
level of control	high	low	high
level of replication	high	low	high
investigation costs	low	medium	high

time and costs of the experimenter (for preparation and evaluation) are normally negligible compared to those of the participants of the investigation. So, the main factor for the costs is the size of the investigation (i. e., the number of participants and the amount of time they have to spend).

An overview of the evaluation of the approaches according to the above factors is given in Table B.1—a detailed evaluation of each approach in the following listing.

- **Surveys**

  Surveys have a high level of control as the asked questions and response options are definable in advance. Yet, "only" subjective values (e. g., personal experiences, opinions and attitudes, etc.) are measurable.

  A survey can be repeated with exactly the same set of questions and response options—either with the same group of subjects or another. So, this approach has a high level of replication.

  Finally, the investigation costs are quite low even for large groups of subjects: The effort for preparation is almost independent from the number of subjects—the final evaluation of the given answers can be automated when using online questionnaires. The subjects can choose when to answer the questions within a specific period of time. The investigation is independent from running projects in a company or expensive experimental setups.

- **Case studies**

  Case studies have a low level of control. They are always connected to—and depending from—a running project. The development of the project over time is not predictable in advance. So, it is hardly possible to get exactly the desired constellation.

  Consequently, also the level of replication is low as it is almost impossible to run the case study under the same circumstances in another project.

  The investigation costs are medium (between those of surveys and controlled experiments).

- **Controlled experiments**

  For controlled experiments, special experimental designs with exactly the desired circumstances can be used. So, they have a high level of control.

  As these experimental designs allow to run an experiment several times under equal circumstances (with different groups of subjects), this approach has also a high level of replication.

  A big drawback are the relative high investigation costs for controlled experiments. They are caused by the laborious preparation for such an experiment and mainly by the expenditure of time of the subjects. Participation in an experiment may not only be very time-intensive for each participant (e. g., for experiments which simulate longer processes)—it also prevents him/her from productive work at the same time in case of using employees of a company as subjects.

As high levels of control and replication are essential for the empirical analyses in this thesis, controlled experiments were selected among the presented approaches. The high investigation costs of a controlled experiment—at least the monetary costs—could be minimized by using students as subjects.

B.3 CONTROLLED EXPERIMENTS

In this section, controlled experiments, which have been chosen as empirical approach in this thesis, are presented in more detail.

The section starts with the main idea of controlled experiments (Subsection B.3.1). Afterwards, Subsection B.3.2 introduces the basic terminology. Finally, a description of the experiment process is given (Subsection B.3.3).

B.3.1 *About Controlled Experiments*

According to Wohlin *et al.* [178, p. 31], the starting point of a controlled experiment is the idea of an existing relationship between a cause and an effect. This relationship is expressible as a theory or—even more formally—it can be formulated as a hypothesis.

In order to test this assumed relationship, one can conduct a controlled experiment. It has the advantage of relative high control about the exact experimental conditions.

In the design of the experiment, one uses several different values (so-called *levels*) of the presumed influencing factor(s). During its execution, one observes the outcome (so-called *response variable(s)*) of these different conditions.

If the controlled experiment is properly set up, conclusions about the assumed relationship and/or hypothesis can be drawn after the analysis of the experiment.

Figure B.1 illustrates the general configuration of a controlled experiment. The used technical terms are explained in Subsection B.3.2.

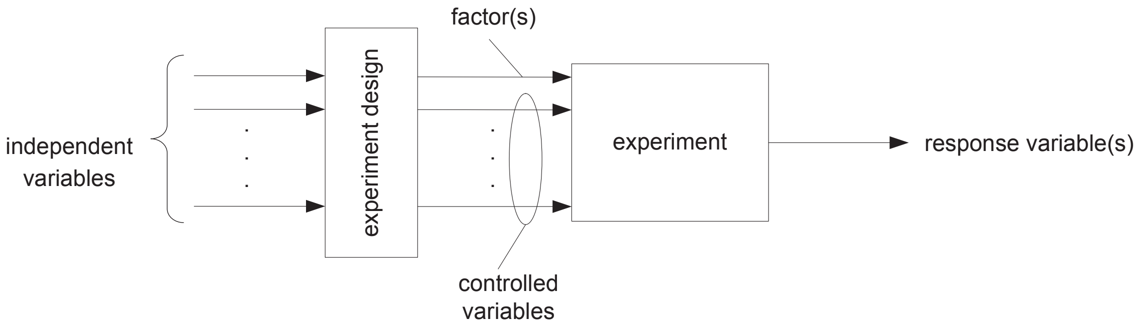


Figure B.1: Illustration of a controlled experiment (based on [178, p. 34]).

B.3.2 *Basic Terminology*

In this subsection, the terminology is introduced which is necessary for the remainder of this section as well as for describing the experiments conducted in Chapter 6 and 7.

Definition B.1 (Subject) *A subject is a person within a controlled experiment who is exposed to the assumed influencing factor (e. g., method, technique, tool, workflow design) to be examined [68, p. 58] [178, p. 34].*

Contrary to natural sciences like, for example, chemistry or physics, subjects have a much higher influence on the outcome of experiments in BPM. Consequently, the proper selection of subjects for a controlled experiment in BPM has to be addressed (see Subsection B.3.3).[3] [68, pp. 58–59]

Definition B.2 (Object) *An object is a thing (e. g., a process model) which is used (e. g., executed or analyzed) or created within a controlled experiment [178, p. 34].*

Definition B.3 (Measurement instrumentation) *The measurement instrumentation of a controlled experiment is how the desired data is collected and possibly examined [178, p. 63].*

In case of a questionnaire, that comprises the selection and arrangement of the questions as well as its later manual or automatic examination (cf. Chapter 6).

As in Chapter 7, measurement instrumentation can also mean the construction of a computer-based experimentation environment (e. g., the experimentation system described in Section 7.3).

The goal of a controlled experiment is the examination of a possible relationship between a cause (*independent variable*) and an effect (*response variable*).

Definition B.4 (Independent variable) *An independent variable is a variable whose changing value is possibly influencing the outcome of a controlled experiment [178, p. 33].*

There are two types of independent variables: *factors* and *controlled variables*.

---

3 Originally, Juristo and Moreno write about software engineering. Yet, their statement is transferable to BPM.

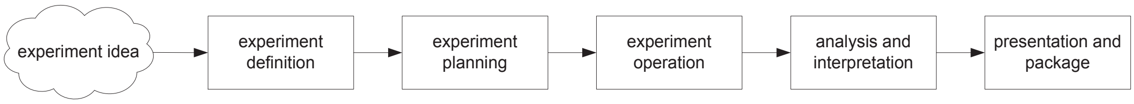


Figure B.2: Overview of the experiment process [178, p. 36].

Definition B.5 (Factor) *A factor is an independent variable whose value is changed throughout the experiment in order to examine its influence on the experiment's outcome [68, p. 60] [178, p. 33].*

Definition B.6 (Controlled variable) *A controlled variable (also called* parameter*) is an independent variable whose value is kept constant throughout the experiment as its possible influence on the experiment's outcome shall be masked [68, pp. 59–60] [178, p. 33].*

Definition B.7 (Response variable) *A response variable (also called* dependent variable*) measures the outcome of a controlled experiment—with it the possible influence of the independent variables [68, p. 59] [178, p. 33].*

The terms "independent variable", "factor", "controlled variable" and "response variable" as well as their relations are depicted in Figure B.1.

During a controlled experiment, several different values of a factor (so-called *alternatives* or *levels*) are used in order to examine the factor's possible influence.

Definition B.8 (Alternative/level) *An alternative or level (sometimes also called* treatment*) is a particular value of a factor used during a controlled experiment [68, pp. 60–61] [178, p. 33].*

B.3.3 *Experiment Process*

According to Wohlin *et al.*, the "starting point for an experiment is insight, and the idea that an experiment would be a possible way of evaluating whatever we are interested in. In other words, we have to realize that an experiment is appropriate for the question we are going to investigate." [178, p. 35]

An experiment is very time- and resource-intensive. Furthermore, many (common) mistakes can be made which may make the results useless. So, after the decision for an experiment, it is necessary to carefully plan its realization. Wohlin *et al.* propose an experiment process for how to perform experiments (see Figure B.2) which is explained in the rest of this subsection.

Definition

The first phase of the experiment process is the definition phase [178, pp. 37, 41–46]. In this phase, the foundation of the experiment is determined. All further steps are then based on this foundation.

Wholin *et al.* propose a goal definition template published in [16, pp. 255–256] and [81, pp. 243–245] for that purpose. It comprises the aspects

- object of study,
- purpose,
- quality focus,
- perspective and
- context.

OBJECT OF STUDY The object of the study is the entity which is studied in the experiment. It can be, for example, an approach, a process definition language, a metric, a tool or a theory.

PURPOSE The purpose defines the intention of the experiment. A possible example is the evaluation and comparison of two different techniques.

QUALITY FOCUS The quality focus is the primary effect of interest in the experiment. Examples are effectiveness, cost, reliability and maintainability.

PERSPECTIVE The perspective defines the viewpoint from which the experiment's results are interpreted. Possible perspectives are, for example, project analyst, developer, user and researcher.

CONTEXT Finally, the context determines the "environment" in which the experiment is run. It contains which persons are involved in the experiment (subjects) and which artifacts (objects) are used.

Planning

The definition phase is followed by the planning phase [178, pp. 37–38, 47–74]. While the definition phase defines *why* the experiment is conducted, the planning phase determines *how* it is conducted [178, p. 47]. It can be divided into the steps

- context selection,
- hypothesis formulation,
- selection of variables,
- selection of subjects,
- experiment design,
- instrumentation and
- validity evaluation.

CONTEXT SELECTION In the context selection step [178, pp. 48–49], the "situation" of the experiment is determined.

In order to get the most general and realistic results, it would be best to run an experiment in connection with a large real world project ("on-line") with professionals involved. Because of the high costs, the additional risks for the project or simply a missing fitting project, one often has to switch to experiments with students dealing with smaller "toy problems" which are not linked with a real project ("off-line").

Summing up, one can characterize the context of an experiment according to the four dimensions

- off-line vs. on-line,
- student vs. professinal,
- toy vs. real problems and
- specific vs. general.

HYPOTHESIS FORMULATION An important element of a controlled experiment is hypothesis testing using statistical methods. In the hypothesis formulation step [178, pp. 49–50], the hypothesis which is to be tested during the experiment is formally stated.

If, for example, an experiment is intended to compare a new process modeling method with an existing one, the hypothesis could be that the average number of errors made by subjects using the new modeling method is smaller than that for the existing one.

The necessary statistical details about hypothesis testing are given in the text about the analysis and interpretation phase (step "Hypothesis Testing" on page 214).

SELECTION OF VARIABLES In this step [178, p. 51], the independent and response variables are selected. Normally, these variables can be identified by looking at the hypothesis formulated in the previous step. Thereby, the independent variables measure the possible influencing values—while the response variable(s) measure(s) the outcome. The identified independent variables have to be further divided into factors (different variable values are used during the experiment) and controlled variables (variable value is kept constant throughout the experiment).

SELECTION OF SUBJECTS The next step [178, pp. 51–52] consists of the selection of the subjects for the experiment. Here, only an abstract decision is made. Concrete individuals are not searched until the following operation phase.

The selected subjects can be seen as a sample of the overall population desired for the experiment. In order to generalize the results, the sample has to be representative for that population.

The number of selected subjects (sample size) impacts the quality of the generalization of the results—the larger the sample, the lower the generalization error. The desirable sample size depends on the properties of the used statistical hypothesis test method and the variability of the overall population.

EXPERIMENT DESIGN In this step [178, pp. 52–62], the experiment design is chosen.

A controlled experiment consists of applying several different levels of the factor(s) to the subjects for later statistical analysis. Thereby, the experiment design determines how this series of single tests is organized. A design consists of the used factor levels, the order of the single tests and the assignment of the subjects to these tests.

Depending on the number of factors in the experiment, one distinguishes between designs with one or at least two factors. As only experiments with one factor are used in this thesis, only information about these designs are given here. Details about designs with at least two factors can be found, for example, in [178, pp. 58–62] and [107].

Independent of the selected experiment design, there are two main principles: randomization and balancing.

Most statistical methods which are used for later analysis of the collected data require that the observations come from independent random variables. *Randomization* is used to meet this requirement. It is applied on the allocation of objects and subjects to the single tests as well as the test order. Furthermore, it is used to assure the representativeness of the selected subjects for the overall population. It works as it averages out other effects.

Balancing means the assignment of the same number of subjects to each factor level of the experiment. Even if it is not absolutely necessary, it simplifies and strengthens the statistical analysis of the data.

Experiment designs with one factor and two or more levels are normally used to compare, for example, several tools, modeling methods, etc. Using a *completely randomized design*, the subjects are randomly assigned to exactly one of the different factor levels. In the case of a *randomized complete block design*, each subject participates in every single test (every factor level). Only the order in which a subject is exposed to the factor levels is randomly determined.

INSTRUMENTATION In the instrumentation step [178, pp. 62–63], the three types of instruments (objects, guidelines and measurement instruments) are chosen.

Possible objects are, for example, process models used in the experiment or the textual description of a process which is to be modeled by the subjects.

Guidelines give the subjects the necessary information about what is required from them in the experiment and how to do this. If, for example, two methods are to be compared, the participants have to be provided with guidelines about these methods. Additionally, they also need training in the applied methods.

The measurement instruments are used for collecting data during the experiment. Possible examples are (online) questionnaires (as in Chapter 6) and computer-based experimentation environments (as in Chapter 7).

VALIDITY EVALUATION The last step in the planning phase is validity evaluation [178, pp. 63–73].

It is important to deal with the question of validity already in the planning phase in order to gain usable results. If possible problems are identified, it is still possible to adjust the experiment before it is actually executed.

In the literature, two important aspects of validity are distinguished.

Definition B.9 (Internal validity) *Internal validity refers to the extent to which one can accurately infer that the independent variable caused the effect observed on the response variable and that it is not a result of a factor which one has no control of or has not measured. [27, p. 217] [178, p. 64]*

Wohlin *et al.* explain [178, p. 65]: "Threats to internal validity concern issues that may indicate a causal relationship, although there is none. Factors that impact on the internal validity are how the subjects are selected and divided into different classes, how the subjects are treated and compensated during the experiment, if special events occur during the experiment etc. All these factors can make the experiment show a behavior that is not due to the treatment but to the disturbing factor."

Definition B.10 (External validity) *External validity refers to the extent to which the results of an experiment can be generalized out of the scope of the study—namely across variations in people, settings, treatments, outcomes and times [27, p. 247] [178, p. 65].*

Wohlin *et al.* illustrate [178, p. 65]: "Threats to external validity concern the ability to generalize experiment results outside the experiment setting. External validity is affected by the experiment design chosen, but also by the objects in the experiment and the subjects chosen. There are three main risks: having wrong participants as subjects, conducting the experiment in the wrong environment and performing it with a timing that affects the results."

In the subsequent lists, important threats to internal and external validity are presented in more detail.

Important threats to internal validity:

- **History**

  The history threat comprises events which occur between the beginning of an experiment and the measurement of the response variable and could possibly influence the subjects' behavior [27, pp. 219–220] [31, p. 51].

- **Maturation**

  The maturation threat refers to changes in the internal conditions (both biological and psychological processes) of a subject as a function of time as, for example, age, learning, fatigue, boredom and hunger. These changes could influence the subject's behavior. [27, pp. 220–221] [31, p. 52]

- **Instrumentation**

  This threat is caused by changes in the measurement of the response variable over time. Automatic and physical measurements normally show no or only small temporal changes. Yet, in cases in which a human observer measures and evaluates the response variable, this could really become a problem similar to that of maturation. [27, p. 221] [31, p. 52]

- **Testing**

  The testing threat refers to the fact that answers given to questions of a test by a subject—and thereby the subject's achieved score—can change (the score often increases) for a second run of the test. This is caused as the subject already has experience with the questions—even if he or she was not given the correct answers before the second run. [27, p. 221] [31, p. 52]

- **Regression**

  The regression artifact refers to the fact that extreme scores in a particular distribution will tend to move—or regress—toward the mean of the distribution as a function of repeated testing. In other words: The scores of the high groups may become lower and vice versa—without any change of the experimental settings itself. [27, pp. 222–223] [31, pp. 52–53]

- **Mortality**

  The mortality threat (called "attrition" in [27, pp. 223-224]) refers to possible problems caused by subjects dropping out of an experiment before it is finished. This can destroy the representativeness of the remaining subjects and lead to effects not caused by the different factor levels but by properties of a subject which increase the possibility of leaving the experiment early. [27, pp. 223-224] [31, p. 53]

- **Selection**

  The selection threat exists when different selection procedures are used for assigning the subjects to the different groups of the experiment. This could lead to inequalities between the groups according to properties as intelligence, age and previous knowledge which also influence the outcome besides the actual factor(s). The best method to prevent this threat is to randomly choose the subjects from the overall population and to randomly assign them to the different groups of the experiment. [27, p. 224] [31, p. 53]

Important threats to external validity:

- **Population validity**

  Population validity (called "interaction of selection and treatment" in [31, p. 73] and [178, p. 73]) refers to the ability to generalize the results from the subjects of an experiment (sample of the population) to the overall population one is interested in. [27, pp. 248–250] [31, p. 73] [178, p. 73]

Christensen explains the problem more detailed [27, p. 248]: For that sake, he distinguishes between two kinds of populations. The *target population* is the larger population one actually wants to generalize to (e. g., all persons with BPM knowledge). The *experimentally accessible population* is that population available to the researcher (e. g., students with BPM knowledge). The desired generalization now consists of two steps. In the first step, one generalizes from the sample of subjects to the experimentally accessible population. If the sample is randomly selected and large enough, this is uncritical. The second step requires to generalize from the experimentally accessible population to the target population. This seldom can be made with any degree of confidence as the experimentally accessible population is rarely representative of the target population. In the example given above, one does not know exactly how representative students with BPM knowledge are to the population of all persons with BPM knowledge including professionals.

- **Ecological validity**

  Ecological validity (called "interaction of setting and treatment" in [31, p. 74] and [178, p. 73]) refers to the generalizability of the results of an experiment across settings and environmental conditions. Examples are a laboratory situation compared to a real project situation or a toy problem compared to a real world process. [27, p. 251] [31, p. 74] [178, p. 73]

- **Temporal validity**

  Temporal validity (called "interaction of history and treatment" in [31, p. 74] and [178, p. 73]) refers to the extent to which the results of an experiment can be generalized across time. The subjects' behavior could differ, for example, between Mondays—returning from weekend—and Fridays—after a hard work week—or between the end of a labor-intensive project and the day returning from a recreative holiday. [27, pp. 251–252] [31, p. 74] [178, p. 73]

Operation

The next phase of the experiment process is operation [178, pp. 38, 75–80]. In this phase, the different treatments of the subjects are actually conducted. It consists of the three steps

- preparation,
- execution and
- data validation.

PREPARATION In the preparation step, concrete individuals have to be found as subjects according to the abstract decision about subjects which was made in the previous phase (step "Selection of Subjects" of phase "Planning" on page 209). In order to keep the participants motivated throughout the experiment, their participation should be voluntary—a small incentive is also possible. Before the experiment is started, the subjects have to be provided with the necessary information about the procedure of the experiment—without informing them about its actual goal as this could influence the experiment's outcome.

Furthermore, the measurement instrumentation must be ready for the experiment. Depending on the actual arrangement of the experiment, e. g., sufficient copies of a questionnaire or computer workplaces (including user accounts) have to be prepared.

EXECUTION In the following execution step, the actual experiment is conducted. Depending on its concrete arrangement, the experiment can take place in a short period of time with the experimenter being present all the time (e. g., paper questionnaire or computer experiment) or it is spread over a larger period of time (e. g., online questionnaire). Also, the actual data collection can vary between fully automated (e. g., online questionnaire or computer experiment) and fully manual (e. g., interview).

DATA VALIDATION The phase finishes with the data validation step. Here, the collected data is checked for reasonability. In case of a questionnaire, for example, it is checked whether the subjects have understood how to answer the questions (not whether the answers are correct!) and whether the answers are serious[4]. Otherwise, the answers of a subject are rejected.

Analysis and Interpretation

The operation phase is followed by the analysis and interpretation phase [178, pp. 38, 81–113]. After the actual execution of the experiment, the obtained data has to be analyzed in this phase with regard to the hypothesis in question and the results have to be interpreted. The phase consists of the steps

- descriptive statistics and
- hypothesis testing.

DESCRIPTIVE STATISTICS A good first step is to apply descriptive statistics (e. g., mean, variance, etc.) to the collected data [178, pp. 38, 82–88] and to use visualization techniques (e. g., box plots, scatter plots, histograms and pie charts) [178, pp. 38, 88–90] for getting a first impression about the outcomes of the experiment.

HYPOTHESIS TESTING The second step is hypothesis testing [117, pp. 483–496] [178, pp. 49–50, 92–95]. In this step, the "correctness" of the hypothesis formulated in the planning phase is examined using the collected data and statistical methods.

4 If a subject, for example, has chosen the same answer possibility for each question or the time spent for answering an online questionnaire is extremely short, it is very likely that this subject's answers are not serious.

Table B.2: Possible outcomes of hypothesis tests [117, p. 488] [156, p. 424].

		decision	
		do not reject H_0	reject H_0
reality	H_0 true	correct decision	type I error P(type I error) $= \alpha$
	H_0 false	type II error P(type II error) $= \beta$	correct decision

In contrast to the proof of a mathematical theorem, the "correctness" of a hypothesis cannot be proofed—it can only be shown to be true with a very large probability. For that sake, the following statistical "trick" is used.

One formulates two hypotheses—namely the null hypothesis and the alternative hypothesis. The *null hypothesis* H_0 is the opposite of the hypothesis which one tries to "proof" during the experiment. It represents the possible fact that the expected cause-effect relationship between independent variable(s) and response variable(s) does not exist. The *alternative hypothesis* H_A—on the other side—represents the fact that the expected relationship really exists. Initially, one assumes that the null hypothesis H_0 is true. Next, one tries to show that this assumption is very unlikely looking at the collected data. In the words of Panik [117, p. 486], "[...] the actual research objective is usually to *obtain support for the alternative hypothesis*. That is, the null hypothesis is the proposition that we wish, in a sense, to *disprove*."

Imagine the following example which is used in the remainder of this paragraph: In an experiment, one tries to show that a new technique causes less errors than an existing one. Let μ_e be the error probability of the existing technique and μ_n that of the new one. Then, the null hypothesis H_0 is that there is no difference between the error probabilities of the two techniques ($H_0 : \mu_n = \mu_e$). The alternative hypothesis H_A, in contrast, states that $H_A : \mu_n < \mu_e$.

The hypothesis testing approach, which was sketchily described above, can produce four different outcomes which are depicted in Table B.2.

Panik exposes [117, p. 487]: "Note that the second option has been expressed as *do not reject* H_0 rather than *accept* H_0. This is because the null hypothesis is regarded as valid unless the data dictates otherwise and thus, if the sample evidence does not support the rejection of the null hypothesis, it only means that *the data has not made its case*."

Besides two cases with a correct decision, there are also two erroneous outcomes—each with a special type of error:

- **Type I error**

  A type I error occurs if a true H_0 is rejected. In other words, a non-existing cause-effect relationship is expected to exist. The probability of this type of error is

  $$P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ true}) = \alpha \quad . \tag{B.1}$$

  α is called the *level of significance* of a hypothesis test.

- **Type II error**

  A type II error occurs if a false H_0 is not rejected. In other words, an existing cause-effect relationship is expected not to exist. The probability of this type of error is

  $$P(\text{type II error}) = P(\text{do not reject } H_0 | H_0 \text{ false}) = \beta \quad . \tag{B.2}$$

Both types of errors cannot be eliminated. Furthermore, they are related in that way that decreasing the probability of one error type increases the probability of the other one. Normally, one tries to decrease the probability of type I errors—minimizing the risk that a cause-effect relationship expected to exist is still not existing [117, p. 491]. Typical values of α are 0.05 or less.

At this point, one needs a decision rule which tells whether or not to reject the null hypothesis H_0 according to the collected data in the experiment.

For that purpose, one selects a *test statistic* T or a random variable whose sampling distribution is known under the assumption that H_0 is true. Now, one can compute the realization t of the test statistic T for the observed experiment data.

The range of possible values of the test statistic T is partitioned into two regions—the critical region (or region of rejection) $\mathcal{R}$ and the region of non-rejection $\overline{\mathcal{R}}$. The *critical region* consists of those sample realizations of T for which the null hypothesis H_0 is rejected. The boundary between the critical region and the region of non-rejection is called the *critical value* t_c. In order to compute t_c and subsequently $\mathcal{R}$, one can modify (B.1) into

$$P(\text{type I error}) = P(t \in \mathcal{R} | H_0 \text{ true}) = \alpha \quad . \tag{B.3}$$

So, the location and size of the critical region depend on the null hypothesis, the alternative hypothesis and the level of significance α.

In the above example, the so-called *t-test* [178, p. 99] can be used. In this case, the test statistic T is a Student's t-distributed t [117, pp. 357–361] which is mainly computed as the difference between the average error probabilities for both techniques multiplied by some correction factors (including the sample variances for both techniques). This test statistic is symmetric to the t value 0, its expected value and maximum are both 0. A possible example of such a test statistic T is depicted in Figure B.3. The critical region is at the lower tail of the distribution. The critical value t_c can be computed so that

$$P(\text{type I error}) = P(t \in \mathcal{R} | H_0 \text{ true}) = P(t \leqslant t_c | \mu_n = \mu_e) = \alpha \quad . \tag{B.4}$$

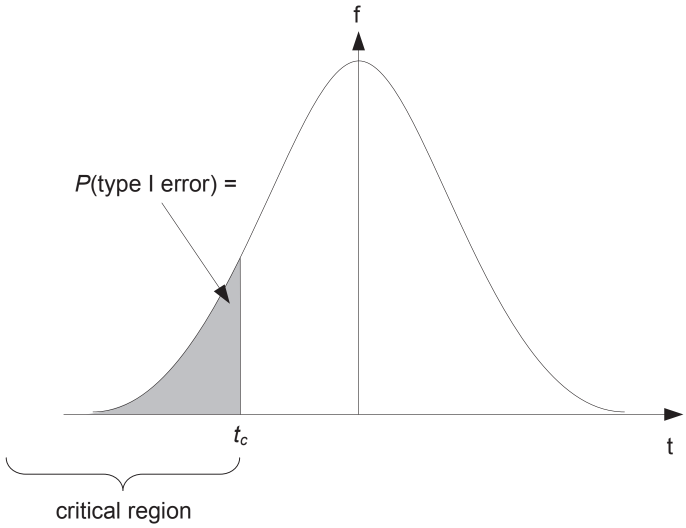


Figure B.3: Critical region $\mathcal{R} = \{t|t \leqslant t_c\}$ for a hypothesis test.

In the critical region, the average error probability for the new technique, which was measured during the experiment, is *significantly* lower than that for the existing technique on the significance level α. Consequently, the null hypothesis (equal error probabilities) is rejected.

Beside the significance level α, there also exists the so-called p-*value*. While α has to be chosen *before* an experiment is actually executed and influences the position of the critical value t_c, the p-value is computed *after* the experiment execution in the "opposite" direction: Here, the critical value t_c is chosen so that it is equal to the realization t of the test statistic T. Afterwards, p is computed as the probability of a type I error using that critical region.

The general hypothesis test procedure is summarized by Panik as [117, pp. 495–496]:

1. Formulate the null hypothesis H_0 (assumed true) and the alternative hypothesis H_A (it determines the location of the critical region $\mathcal{R}$).
2. Specify the level of significance α (it determines the size of $\mathcal{R}$).
3. Select an appropriate test statistic T whose sampling distribution is known under the assumption that H_0 is true.
4. Find $\mathcal{R}$. This involves determining t_c, the critical value of T.
5. Compute the value t of the test statistic (according to the measured experiment data).
6. If t is an element of $\mathcal{R}$, then reject H_0.

Finally, the results of the hypothesis test have to be interpreted: If the null hypothesis could be rejected, the assumed influence of the independent variable(s) on the response variable(s) is most likely. Considering the external validity of the experiment, general conclusions can be made for settings similar to those of the experiment. Furthermore, the results can cause decisions about future applications of the analyzed tool, method, etc. in (commercial) projects and give ideas for additional experiments. [178, pp. 38, 112–113]

No specific hypothesis tests are explained here. The used tests are presented within the thesis at the place where they are used.

Presentation and Package

The last phase of the experiment process is presentation and package [178, pp. 39, 115–118].

The intention of this step is to document the outline and findings of the experiment. So, it is also possible to inform others about the results. This can be done, for example, as an article for a conference or a journal. Wohlin *et al.* propose a template for such documents [178, pp. 116–118].

For enabling the assessment and interpretability of the experiment, all necessary information about its motivation, goals, design and results must be provided. This is also necessary for a possible replication.

C

MEASURING CORRELATIONS

Throughout this thesis, there is a need for examining the dependency of observed value pairs (x_i, y_i) (e. g., values of two process model metrics for a collection of process models as in Chapter 4).

There is a large variety of possible dependencies. Figure C.1 shows some examples of observable value pairs depicted as scatter plots.

Knowing an exact functional dependency of the form $y_i = f(x_i)$ or $x_i = g(y_i)$ would be best. Yet, this is often impossible—especially, due to measuring errors which obscure the underlying dependency.

Here, the concept of *correlation* is a good alternative. Correlation indicates the strength and direction of linear—or at least monotonic—dependencies. Yet, it must be clearly stated that correlation says *nothing* about the existence—or absence—of causality between the two considered quantities.

If one finds a high correlation between X and Y, there are at least three different possible types of causality behind it (see Figure C.2):

- X influences Y.
- Y influences X.
- Neither X nor Y influences the other variable. Instead, an additional variable (a so-called *lurking variable*) Z influences both X and Y.

Watkins *et al.* give an example for the last case [171, p. 142]: In a sample of elementary school pupils, there is a strong correlation between shoe size and scores on a test of mathematical abilities. Nevertheless, neither mathematical skills cause feet to get bigger nor vice versa. In fact, the lurking variable age influences both foot size and mathematical knowledge.

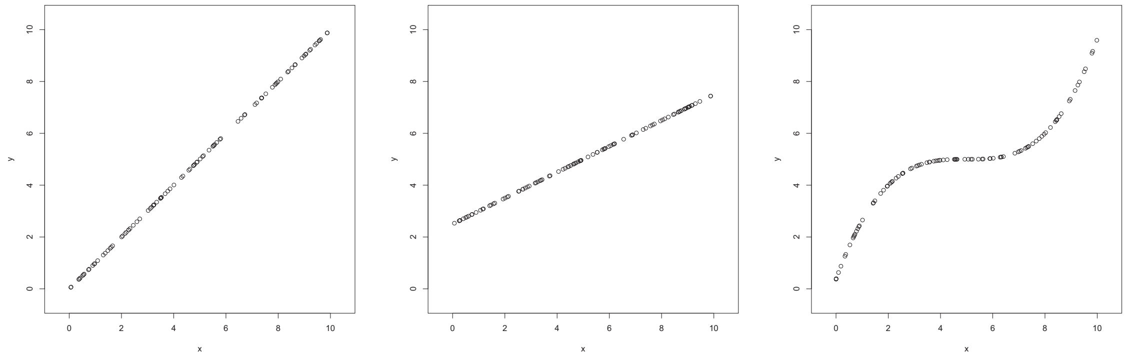
In the remaining chapter, two often used measures of correlation—Pearson's product-moment correlation coefficient for measuring linear dependencies (Section C.1) and Spearman's rank correlation coefficient for measuring monotonic dependencies (Section C.2)—are introduced.

C.1 PEARSON'S PRODUCT-MOMENT CORRELATION COEFFICIENT

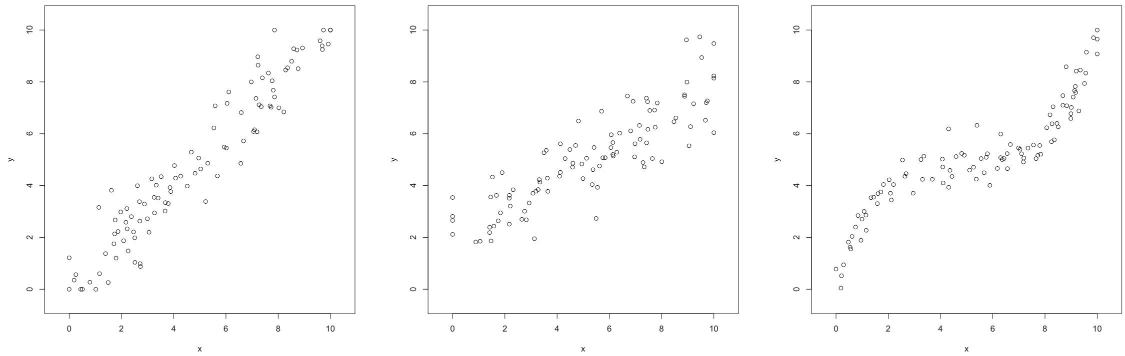
The first presented measure of correlation, Pearson's product-moment correlation coefficient, measures the strength and direction of linear dependencies.

Details about its historical development and different ways of interpretation can be found in [131].

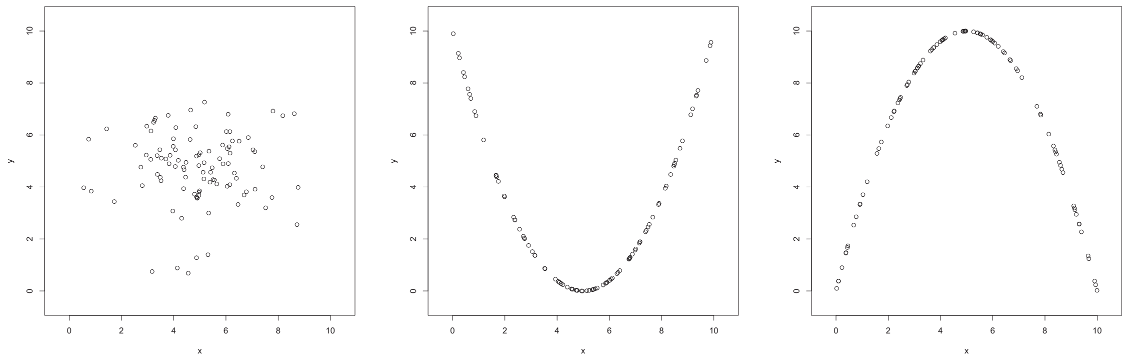
The basic "idea" behind the measure is (sample) covariance as defined in Definition C.1 [122].



(a) 1.000/1.000 (b) 1.000/1.000 (c) 0.918/1.000



(d) 0.957/0.951 (e) 0.868/0.885 (f) 0.906/0.945



(g) −0.010/−0.062 (h) −0.060/0.032 (i) −0.026/0.040

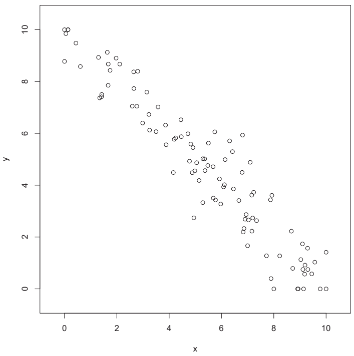
Figure C.1: Pearson's correlation/Spearman's rank correlation for different sets of value pairs (Part 1 of 2).

Definition C.1 ((Sample) covariance) *The (sample) covariance* cov *of a set of observed value pairs* (x_i, y_i), $i \in \{1, \dots, n\}$ *is defined as*

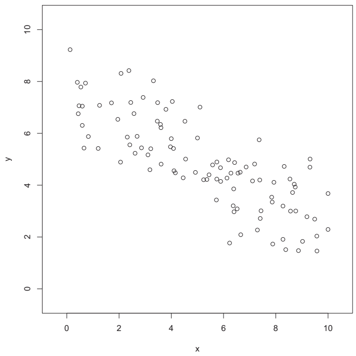
$$\text{cov} := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \tag{C.1}$$

with $\overline{x}$ *and* $\overline{y}$ *as means of the* x_i *and* y_i *respectively.*

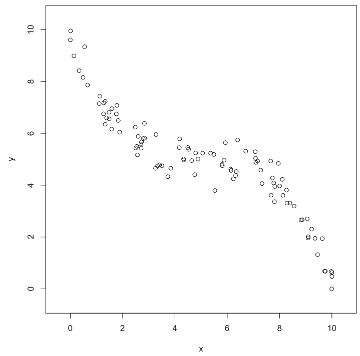
For value pairs with a positive linear dependency (the larger the x value, the larger the y value), the (sample) covariance has a positive value—for pairs with a



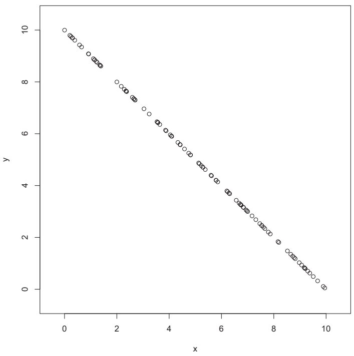
(j) −0.948/−0.938



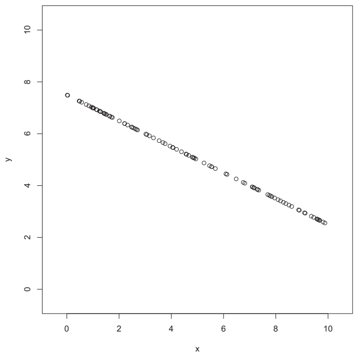
(k) −0.804/−0.826



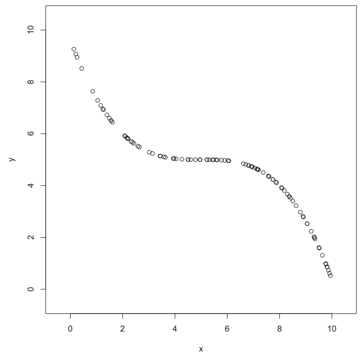
(l) −0.901/−0.927



(m) −1.000/−1.000



(n) −1.000/−1.000

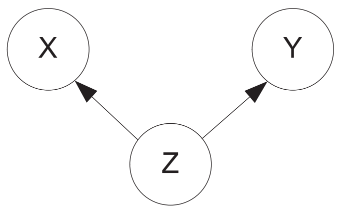


(o) −0.914/−1.000

Figure C.1: Pearson's correlation/Spearman's rank correlation for different sets of value pairs (Part 2 of 2).

(a) X influences Y.

(b) Y influences X.



(c) Lurking variable Z influences X and Y.

Figure C.2: Three different types of causality.

negative dependency (the larger the x value, the smaller the y value) a negative one.

Yet, covariance has a critical disadvantage: the covariance values of different sets of value pairs are not comparable. If one measures, for example, the same set of value pairs with two different units of measurement (e. g., $x'_i := 2x_i$ and $y'_i := 2y_i$), one gets two different values for the covariance. As the underlying dependencies are the same in both cases, this effect is unsatisfying.

This problem can be solved by dividing the covariance by the product of the standard deviations of the x_i and y_i—resulting in the definition of Pearson's (sample) product-moment correlation coefficient [109].

Definition C.2 (Pearson's (sample) product-moment correlation coefficient) *Pearson's (sample) product-moment correlation coefficient* r *of a set of observed value pairs* (x_i, y_i), $i \in \{1, \ldots, n\}$ *is defined as*

$$r := \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{C.2}$$

with $\overline{x}$ *and* $\overline{y}$ *means of the* x_i *and* y_i *respectively.*

As this measure has the following desired properties, its values are now comparable for different sets of observed value pairs.

Theorem C.1 *For the Pearson's (sample) product-moment correlation coefficient* r, *the properties*

1. $-1 \leqslant r \leqslant 1$ *and*
2. $|r| = 1 \Leftrightarrow$ *There is a perfect linear dependency between the* x_i *and* y_i

hold.

Proof. Let $a_i, b_i \in \mathbb{R}$, $i \in \{1, \ldots, n\}$. Then,

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leqslant \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 \tag{C.3}$$

holds with equality if and only if there is a $c \in \mathbb{R}$ such that $a_i = r b_i$ for all $i \in \{1, \ldots, n\}$. (Cauchy-Bunyakovsky-Schwarz inequality) [38, Theorem 2.1, p. 5]

REGARDING 1.) Let $a_i := x_i - \overline{x}$ and $b_i := y_i - \overline{y}$. From (C.3) follows

$$\left(\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})\right)^2 \leqslant \sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2 \quad . \tag{C.4}$$

By extracting the square root, one gets

$$\left|\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})\right| \leqslant \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2} \quad . \tag{C.5}$$

What can be transformed into

$$\frac{|\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})|}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} = |r| \leqslant 1 \quad . \tag{C.6}$$

REGARDING 2.)

There is a perfect linear dependency between the x_i and y_i.

$$\Leftrightarrow \quad y_i = mx_i + t \quad, m \in \mathbb{R} \setminus \{0\}, t \in \mathbb{R}$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(mx_i + t) = \frac{m}{n}\sum_{i=1}^{n} x_i + t = m\overline{x} + t$$

$$\Leftrightarrow \quad \begin{aligned} x_i - \overline{x} &= \frac{1}{m} \cdot m \cdot (x_i - \overline{x}) \\ &= \frac{1}{m}(mx_i - m\overline{x}) \\ &= \frac{1}{m}[(mx_i + t) - (m\overline{x} + t)] \\ &= \frac{1}{m}(y_i - \overline{y}) \end{aligned}$$

$$\Leftrightarrow \quad \exists c \in \mathbb{R} \setminus \{0\}^1 : x_i - \overline{x} = a_i = cb_i = c(y_i - \overline{y})$$

$$\Leftrightarrow \quad \frac{\left|\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})\right|}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} = |r| = 1$$

■

On the basis of Figure C.1, the general behavior of Pearson's product-moment correlation coefficient will be explained:

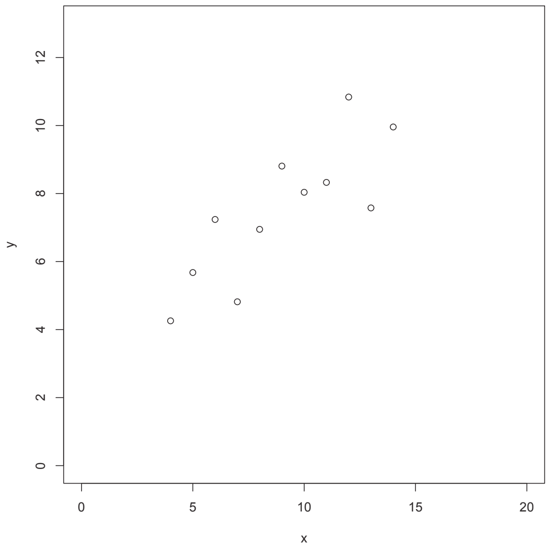
- For positive dependencies (the larger the x_i, the larger the y_i), Pearson's product-moment correlation coefficient is positive (see Figure C.1a, C.1b, C.1d and C.1e)—for negative dependencies (the larger the x_i, the smaller the y_i), it is negative (see Figure C.1j, C.1k, C.1m and C.1n).
- For perfect linear dependencies, the absolute value of Pearson's product-moment correlation coefficient is 1 (see Figure C.1a, C.1b, C.1m and C.1n). Different slopes of the lines connecting the value pairs do not have any influence (e. g., Figure C.1a compared to Figure C.1b).
- The less perfect the linear dependency is, the smaller the absolute value of Pearson's product-moment correlation coefficient (see Figure C.1d, C.1e, C.1j and C.1k).
- If the value pairs are nearly randomly distributed (no linear dependency at all), Pearson's product-moment correlation coefficient is approximately 0 (see Figure C.1g).

For determining Pearson's product-moment correlation coefficient, the observed value pairs must be measured on at least an interval scale (see Section A.2). Otherwise, the necessary computation of means and standard deviations would be impossible or at least meaningless.

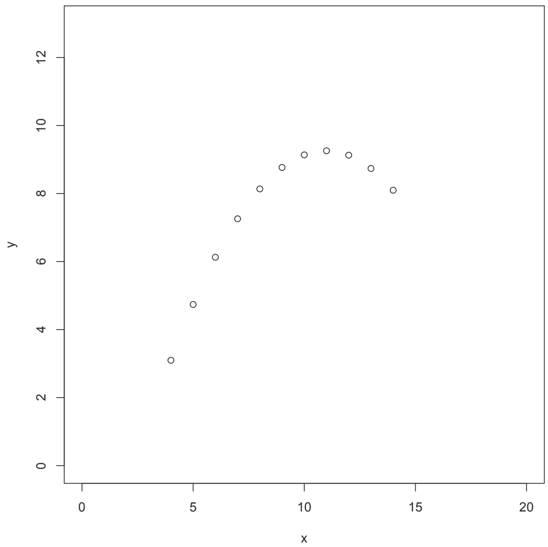
Finally, it must be stated that although the existence of a linear dependency implies a high Pearson's product-moment correlation coefficient, the opposite direction does not have to be true. Anscombe's quartet (see Figure C.3), four data sets constructed and published by Anscombe in [3], is a good counter-example.

The four data sets all have approximately the same relatively high Pearson's product-moment correlation coefficient of 0.816 (Figure C.3a–C.3c) and 0.817

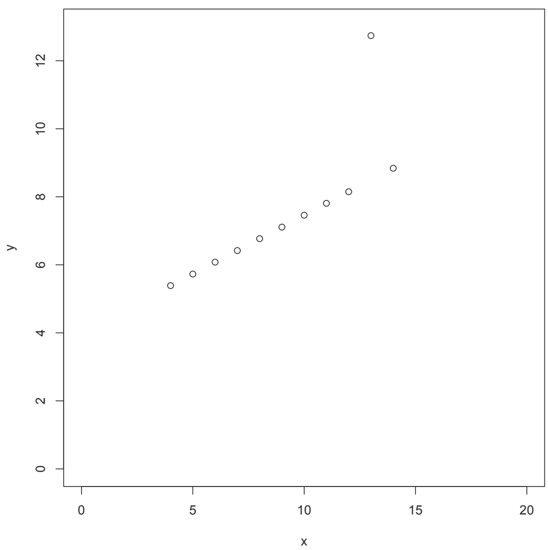
1 If $c = 0$, there would be a division by zero in the definition of r.



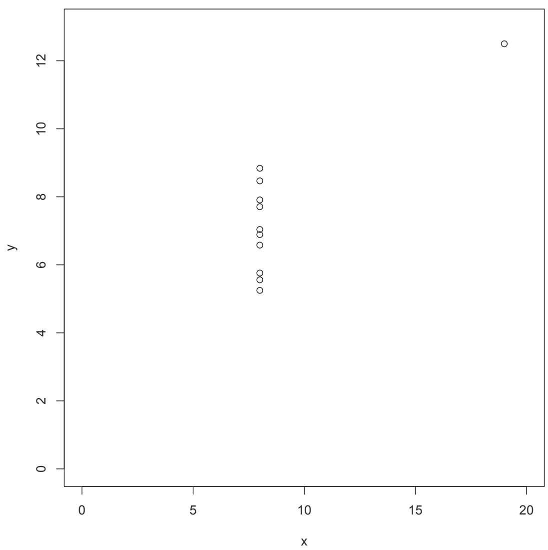
(a) Data set 1.



(b) Data set 2.



(c) Data set 3.



(d) Data set 4.

Figure C.3: Anscombe's quartet: Four data sets with Pearson's product-moment correlation coefficient of 0.816 (Figure C.3a–C.3c) and 0.817 (Figure C.3d), respectively [3].

(Figure C.3d). Nevertheless, it can be easily seen that some of them are not linearly dependent.

So, when having found a high correlation coefficient, visually verifying the actual existence of a linear dependency by depicting the values in a scatter plot is a good idea.

C.2 SPEARMAN'S RANK CORRELATION COEFFICIENT

For value pairs with a perfect monotonic—but not linear—dependency (as in Figure C.1c and C.10), Pearson's product-moment correlation coefficient does not reach 1 or -1, respectively.

For values which are not measured on at least an interval scale (see Section A.2), Pearson's product-moment correlation coefficient is not applicable[2].

For both cases, there is another measure of dependency—Spearman's rank correlation coefficient—which uses the concept of ranks and only needs values on at least an ordinal scale (see Section A.2).

Definition C.3 (Ranking order) *Let $\{x_1, \dots, x_n\}$ be a set of observations on an ordinal scale. A permutation $\delta : \{1, \dots, n\} \mapsto \{1, \dots, n\}$ constitutes a ranking order of the x_i if the condition*

$$x_{\delta^{-1}(1)} \leqslant x_{\delta^{-1}(2)} \leqslant \dots \leqslant x_{\delta^{-1}(n-1)} \leqslant x_{\delta^{-1}(n)} \tag{C.7}$$

is fulfilled.

If there are ties in the x_i (i. e., there are several x_i with the same value), multiple valid ranking orders exist.

In a next step, the rank of an observation can be defined [144, p. 5].

Definition C.4 (Rank) *Let $\{x_1, \dots, x_n\}$ be a set of observations on an ordinal scale and $\delta : \{1, \dots, n\} \mapsto \{1, \dots, n\}$ be a permutation constituting a ranking order of the x_i. The function $\mathrm{rank} : \{x_i\} \mapsto \mathbb{R}_+$ defined as*

$$\mathrm{rank}(x_i) := \begin{cases} \delta(i) & \textit{if } \nexists j : x_i = x_j, i \neq j \\ \frac{\sum_{j:x_i=x_j} \delta(j)}{|\{x_j | x_i = x_j\}|} & \textit{otherwise} \end{cases} \tag{C.8}$$

assigns each observation x_i its rank.

Theorem C.2 *The means of the ranks of the x_i ($\overline{\mathrm{rank}_x}$) and y_i ($\overline{\mathrm{rank}_y}$) of observed value pairs (x_i, y_i), $i \in \{1, \dots, n\}$ are*

$$\overline{\mathrm{rank}_x} = \overline{\mathrm{rank}_y} = \frac{n+1}{2} \quad . \tag{C.9}$$

Proof. It holds

$$\overline{\mathrm{rank}_x} = \frac{\sum_{i=1}^{n} \mathrm{rank}(x_i)}{n} = \frac{\sum_{i=1}^{n} \delta_x(i)}{n} = \frac{\sum_{i=1}^{n} i}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2} \tag{C.10}$$

and

$$\overline{\mathrm{rank}_y} = \frac{\sum_{i=1}^{n} \mathrm{rank}(y_i)}{n} = \frac{\sum_{i=1}^{n} \delta_y(i)}{n} = \frac{\sum_{i=1}^{n} i}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2} \quad . \tag{C.11}$$

■

2 For values on an ordinal scale (see Section A.2), Pearson's product-moment correlation coefficient would be computable. Yet, the resulting coefficient would be meaningless as there are no equal distances on that scale what would be necessary for measuring a linear dependency between value pairs.

The definition of Spearman's rank correlation coefficient is based on Pearson's product-moment correlation coefficient. Instead of using the x_i and y_i directly, they are replaced by their ranks—resulting in Definition C.5 [110].

Definition C.5 (Spearman's rank correlation coefficient) *Spearman's rank correlation coefficient* r_s *of a set of observed value pairs* (x_i, y_i), $i \in \{1, \ldots, n\}$ *is defined as*

$$r_s := \frac{\sum_{i=1}^{n} \left[\mathrm{rank}(x_i) - \overline{\mathrm{rank}_x}\right] \left[\mathrm{rank}(y_i) - \overline{\mathrm{rank}_y}\right]}{\sqrt{\sum_{i=1}^{n} \left[\mathrm{rank}(x_i) - \overline{\mathrm{rank}_x}\right]^2 \sum_{i=1}^{n} \left[\mathrm{rank}(y_i) - \overline{\mathrm{rank}_y}\right]^2}} \tag{C.12}$$

$$= \frac{\sum_{i=1}^{n} \left[\mathrm{rank}(x_i) - \frac{n+1}{2}\right] \left[\mathrm{rank}(y_i) - \frac{n+1}{2}\right]}{\sqrt{\sum_{i=1}^{n} \left[\mathrm{rank}(x_i) - \frac{n+1}{2}\right]^2 \sum_{i=1}^{n} \left[\mathrm{rank}(y_i) - \frac{n+1}{2}\right]^2}} . \tag{C.13}$$

EXAMPLE Using the values of Table C.1, the computation of the ranks and Spearman's rank correlation coefficient shall be demonstrated.

For the determination of the ranks, valid ranking orders for the x_i and y_i have to be found. As one can easily see, both the x_i and y_i have ties. Consequently, all three x values of 2 get the rank 3 (as defined in the second alternative of equation C.8) and the two y values of 8.7 get rank 6.5. The computation of the other values' ranks is "intuitive".

According to Theorem C.2, the means of the ranks are $\overline{\mathrm{rank}_x} = \overline{\mathrm{rank}_y} = \frac{8+1}{2} = 4.5$.

Inserting the computed ranks and their means in (C.12) results in

$$\sum_{i=1}^{n} \left[\mathrm{rank}(x_i) - \overline{\mathrm{rank}_x}\right]^2$$
$$= 12.25 + 2.25 + 2.25 + 2.25 + 0.25 + 2.25 + 6.25 + 12.25 = 40$$

as well as

$$\sum_{i=1}^{n} \left[\mathrm{rank}(y_i) - \overline{\mathrm{rank}_y}\right]^2$$
$$= 2.25 + 0.25 + 4 + 12.25 + 6.25 + 4 + 0.25 + 12.25 = 41.5$$

and finally in a Spearman's rank correlation coefficient r_s of

$$r_s = \frac{5.25 - 0.75 - 3 - 5.25 - 1.25 + 3 - 1.25 - 12.25}{\sqrt{40 \cdot 41.5}} = -\frac{15.5}{\sqrt{1660}} \approx -0.380 .$$

As already written above, Spearman's rank correlation coefficient is also applicable for data measured on an ordinal scale. The reason is that the values are replaced by their ranks. These ranks are subsequent increasing numbers with distance 1 (except for observed values with ties). Consequently, the ranks, which are used in the formula of Pearson's product-moment correlation coefficient to compute Spearman's rank correlation coefficient, are on an interval scale instead of an ordinal scale as the original values.

Table C.1: Example for computation of Spearman's rank correlation coefficient.

x_i	1.5	2	2	2	4.8	4.9	5	10
y_i	2.5	3.2	8.7	13	1.4	8.7	3	1
$\mathrm{rank}(x_i)$	1	3	3	3	5	6	7	8
$\mathrm{rank}(y_i)$	3	5	6.5	8	2	6.5	4	1
$\mathrm{rank}(x_i) - \overline{\mathrm{rank}_x}$	−3.5	−1.5	−1.5	−1.5	0.5	1.5	2.5	3.5
$\mathrm{rank}(y_i) - \overline{\mathrm{rank}_y}$	−1.5	0.5	2	3.5	−2.5	2	−0.5	−3.5
$[\mathrm{rank}(x_i) - \overline{\mathrm{rank}_x}][\mathrm{rank}(y_i) - \overline{\mathrm{rank}_y}]$	5.25	−0.75	−3	−5.25	−1.25	3	−1.25	−12.25
$[\mathrm{rank}(x_i) - \overline{\mathrm{rank}_x}]^2$	12.25	2.25	2.25	2.25	0.25	2.25	6.25	12.25
$[\mathrm{rank}(y_i) - \overline{\mathrm{rank}_y}]^2$	2.25	0.25	4	12.25	6.25	4	0.25	12.25

For value pairs with a perfect monotonic—but not linear—dependency (as in Figure C.1c and C.1o), their ranks have a perfect linear dependency. Because of that, Spearman's rank correlation coefficient is 1 or −1, respectively. Pearson's product-moment correlation coefficient only reaches values with an absolute value smaller than 1 for these cases as the value pairs itself are not linearly dependent.

Value pairs with a Pearson's product-moment correlation coefficient of 1 or −1 (perfect linear dependency) also have a Spearman's rank correlation coefficient of 1 or −1, respectively (perfect monotonic dependency)—as can be seen in Figure C.1a, C.1b, C.1m and C.1n.

As one can see in Figure C.1h and C.1i, there are cases for which both Pearson's product-moment and Spearman's rank correlation coefficient have very small values even though the x_i and y_i are strongly dependent. Because of this, it is always useful to additionally depict the observed value pairs using a scatter plot in order to visually search for dependencies not "detectable" by the correlation coefficients (neither linear nor monotonic).

BIBLIOGRAPHY

[1] BPM: Not just workflow anymore. Survey, AIIM, 2007. http://www.xerox.com/downloads/usa/en/x/XGS_article_bpm_report1.pdf (last accessed on 2011-12-01).

[2] D. F. Andrews. Plots of high-dimensional data. *Biometrics*, 28(1):125–136, 1972.

[3] F. J. Anscombe. Graphs in statistical analysis. *The American Statistician*, 27(1):17–21, 1973.

[4] W. Ross Ashby. Some peculiarities of complex systems. *Cybernetic Medicine*, 9(2):1–8, 1973.

[5] Earl Babbie. *The Practice of Social Research*. Thomson/Wadsworth, 10th edition, 2004.

[6] S. Balasubramanian and Mayank Gupta. Structural metrics for goal based business process design and evaluation. *Business Process Management Journal*, 11(6):680–694, 2005. doi:10.1108/14637150510630855.

[7] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. doi:10.1126/science.286.5439.509.

[8] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272(1-2):173–187, 1999. doi:10.1016/S0378-4371(99)00291-5.

[9] Victor R. Basili, Gianluigi Caldiera, and H. Dieter Rombach. Goal question metric paradigm. In John J. Marciniak, editor, *Encyclopedia of Software Engineering*, volume 1, pages 528–532. Wiley, 1994.

[10] Jörg Becker and Dieter Kahn. The process in focus. In Jörg Becker, Martin Kugeler, and Michael Rosemann, editors, *Process Management: A Guide for the Design of Business Processes*, pages 1–12. Springer, 2003.

[11] Claude Berge. *Graphs and Hypergraphs*, volume 6 of *North-Holland Mathematical Library*. North-Holland, 2nd edition, 1976.

[12] P. Berkhin. A survey of clustering data mining techniques. In Jacob Kogan, Charles Nicholas, and Marc Teboulle, editors, *Grouping Multidimensional Data: Recent Advances in Clustering*, pages 25–71. Springer, 2006. doi:10.1007/3-540-28349-8_2.

[13] Thorsten Blecker and Wolfgang Kersten, editors. *Complexity Management in Supply Chains: Concepts, Tools and Methods*, volume 2 of *Operations and Technology Management*. Erich Schmidt, 2006.

[14] Vincent D. Blondel and John N. Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 36(9):1249–1274, 2000. doi:10.1016/S0005-1098(00)00050-9.

[15] Berndt Brehmer and Dietrich Dörner. Experiments with computer-simulated mircoworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, 9(2–3):171–184, 1993. doi:10.1016/0747-5632(93)90005-D.

[16] Lionel C. Briand, Christiane M. Differding, and H. Dieter Rombach. Practical guidelines for measurement-based process improvement. *Sofware Process: Improvement and Practice*, 2(4):253–280, 1996. doi:10.1002/(SICI)1099-1670(199612)2:4<253::AID-SPIP53>3.0.CO;2-G.

[17] Axel Buchner. Basic topics and approaches to the study of complex problem solving. In Peter A. Frensch and Joachim Funke, editors, *Complex Problem Solving: The European Perspective*, pages 27–63. Erlbaum, 1995.

[18] Elwood S. Buffa and Rakesh K. Sarin. *Modern Production/Operations Management*. Wiley Series in Production/Operations Management. Wiley, 8th edition, 1987.

[19] J. Cardoso, J. Mendling, G. Neumann, and H. A. Reijers. A discourse on complexity of process models. In Johann Eder and Schahram Dustdar, editors, *Business Process Management Workshops: Proceedings of the BPM 2006 International Workshops*, volume 4103 of *Lecture Notes in Computer Science*, pages 117–128, 2006. doi:10.1007/11837862_13.

[20] Jorge Cardoso. About the data-flow complexity of web processes. In *Proceedings of the 6th International Workshop on Business Process Modeling, Development, and Support (BPMDS'05)*, pages 67–74, 2005.

[21] Jorge Cardoso. How to measure the control-flow complexity of web processes and workflows. In Layna Fischer, editor, *Workflow Handbook 2005*, pages 199–212. Future Strategies Inc., 2005.

[22] Jorge Cardoso. Process control-flow complexity metric: An empirical validation. In *Proceedings of the 2006 IEEE International Conference on Services Computing (SCC 2006)*, pages 167–173, 2006. doi:10.1109/SCC.2006.82.

[23] Jorge Cardoso. Business process quality metrics: Log-based complexity of workflow patterns. In Robert Meersman and Zahir Tari, editors, *On the Move to Meaningful Internet Systems 2007: Proceedings of the OTM Confederated International Conferences CoopIS, DOA, ODBASE, GADA, and IS 2007 (Part I)*, volume 4803 of *Lecture Notes in Computer Science*, pages 427–434, 2007. doi:10.1007/978-3-540-76848-7_30.

[24] Jorge Cardoso. Complexity analysis of BPEL web processes. *Software Process: Improvement and Practice*, 12(1):35–49, 2007. `doi:10.1002/spip.302`.

[25] Lewis Carroll. *Through the Looking-glass and What Alice Found There*. Macmillan, 1887.

[26] Shyam R. Chidamber and Chris F. Kemerer. A metrics suite for object oriented design. *IEEE Transactions on Software Engineering*, 20(6):476–493, 1994. `doi:10.1109/32.295895`.

[27] Larry B. Christensen. *Experimental Methodology*. Pearson, 10th edition, 2007.

[28] BPM ist für viele Neuland. *Computerwoche*, (9):26, 2010. – In German.

[29] Prozessintelligenz dient als Frühindikator für Finanzen. *Computer Zeitung*, 39(13):1, 2009. – In German.

[30] Prozessmanagement amortisiert sich ein Jahr nach Einführung. *Computer Zeitung*, 39(23):11, 2009. – In German.

[31] Thomas D. Cook and Donald T. Campbell. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Houghton Mifflin, 1979.

[32] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, 2nd edition, 2001.

[33] Thomas Curran, Gerhard Keller, and Andrew Ladd. *SAP R/3 Business Blueprint: Understanding the Business Process Reference Model*. Prentice Hall PTR, 1998.

[34] David P. Darcy, Chris F. Kemerer, Sandra A. Slaughter, and James E. Tomayko. The structural complexity of software: An experimental test. *IEEE Transactions on Software Engineering*, 31(11):982–995, 2005. `doi:10.1109/TSE.2005.130`.

[35] Thomas H. Davenport. *Process Innovation: Reengineering Work through Information Technology*. Harvard Business School Press, 1993.

[36] David L. Davies and Donald W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2):224–227, 1979. `doi:10.1109/TPAMI.1979.4766909`.

[37] Tom DeMarco. *Controlling Software Projects: Management, Measurement & Estimation*. Yourdon Press, 1982.

[38] S. S. Dragomir. A survey on Cauchy-Bunyakovsky-Schwarz type discrete inequalities. *Journal of Inequalities in Pure and Applied Mathematics*, 4(3), 2003.

[39] Dietrich Dörner. *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations*. Basic Books, 1997.

[40] Dietrich Dörner and Alex J. Wearing. Complex problem solving: Toward a (computersimulated) theory. In Peter A. Frensch and Joachim Funke, editors, *Complex Problem Solving: The European Perspective*, pages 27–63. Erlbaum, 1995.

[41] Bruce Edmonds. Complexity and scientific modelling. *Foundations of Science*, 5(3):379–390, 2000. doi:10.1023/A:1011383422394.

[42] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, 1993.

[43] Norman E. Fenton and Shari Lawrence Pfleeger. *Software Metrics: A Rigorous and Practical Approach*. International Thomson Computer Press, 2nd edition, 1996.

[44] A. Ferguson, C.S . Myers, R. J. Bartlett, H. Banister, F. C. Bartlett, W. Brown, N. R. Campbell, K. J. W. Craik, J. Drever, J. Guild, R. A. Houstoun, J. O. Irwin, G. W. C. Kaye, S. J. F. Philpott, L. F. Richardson, J. H. Shaxby, T. Smith, R. H. Thouless, and W. S. Tucker. Quantitative estimates of sensory events: Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. *Advancement of Science*, 1(2):331–349, 1940.

[45] Peter A. Frensch and Joachim Funke. Definitions, traditions, and a general framework for understanding complex problem solving. In Peter A. Frensch and Joachim Funke, editors, *Complex Problem Solving: The European Perspective*, pages 3–25. Erlbaum, 1995.

[46] S. Friedmann. Graphische Darstellung der jährlichen Temperatur eines Ortes durch geschlossene Curven. *Mittheilungen der Kaiserlich-Königlichen Geographischen Gesellschaft*, 6:244–246, 1862. – In German.

[47] Joachim Funke. Solving complex problems: Exploration and control of complex systems. In Robert J. Sternberg and Peter A. Frensch, editors, *Complex Problem Solving: Principles and Mechanisms*, pages 185–222. Erlbaum, 1991.

[48] Joachim Funke. Experimental research on complex problem solving. In Peter A. Frensch and Joachim Funke, editors, *Complex Problem Solving: The European Perspective*, pages 243–268. Erlbaum, 1995.

[49] Félix García, Mario Piattini, Francisco Ruiz, Gerardo Canfora, and Corrado A. Visaggio. FMESP: Framework for the modeling and evaluation of software processes. *Journal of Systems Architecture*, 52(11):627–639, 2006. doi:10.1016/j.sysarc.2006.06.007.

[50] Micheal R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. A Series of Books in the Mathematical Sciences. Freeman, 1979.

[51] Andrew Gemino and Yair Wand. Evaluating modeling techniques based on models of learning. *Communications of the ACM*, 46(10):79–84, 2003. doi:10.1145/944217.944243.

[52] Diimitrios Georgakopoulos, Mark Hornick, and Amit Sheth. An overview of workflow management: From process modeling to workflow automation infrastructure. *Distributed and Parallel Databases*, 3(2):119–153, 1995. doi: 10.1007/BF01277643.

[53] Volker Gruhn and Ralf Laue. Adopting the cognitive complexity measure for business process models. In Yiyu Yao, Zhongzhi Shi, Yingxu Wang, and Witold Kinsner, editors, *Proceedings of the 5th IEEE International Conference on Cognitive Informatics (ICCI 2006)*, volume 1, pages 236–241, 2006. doi: 10.1109/COGINF.2006.365702.

[54] Volker Gruhn and Ralf Laue. Complexity metrics for business process models. In Witold Abramowicz and Heinrich C. Mayr, editors, *Business Information Systems: Proceedings of the 9th International Conference on Business Information Systems (BIS 2006)*, volume P-85 of *Lecture Notes in Informatics (LNI)*, pages 1–12, 2006.

[55] Maurice H. Halstead. *Elements of Software Science*, volume 2 of *Operating and Programming Systems Series*. Elsevier, 1977.

[56] Michael Hammer and James Champy. *Reengineering the Corporation: A Manifesto for Business Revolution*. HarperBusiness, 1993.

[57] Sallie Henry and Dennis Kafura. Software structure metrics based on information flow. *IEEE Transactions on Software Engineering*, 7(5):510–518, 1981.

[58] David Hollingsworth. The workflow reference model. Specification TC00-1003, Workflow Management Coalition, 1995. Issue 1.1, http://www.wfmc.org/standards/docs/tc003v11.pdf (last accessed on 2011-12-01).

[59] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 2nd edition, 2001.

[60] IEEE Computer Society. IEEE Standard 1061 - IEEE Standard for a software quality metrics methodology, 1998.

[61] Alfred Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985. doi:10.1007/BF01898350.

[62] Alfred Inselberg. Parallel coordinates: Visualization, exploration and classification of high-dimensional data. In Chun-houh Chen, Wolfgang Härdle, and Antony Unwin, editors, *Handbook of Data Visualization*, Springer Handbooks of Computational Statistics, pages 643–680. Springer, 2008. doi:10.1007/978-3-540-33037-0_25.

[63] M. H. Jansen-Vullers, P. A. M. Kleingeld, and H. Netjes. Quantifying the performance of workflows. *Information Systems Management*, 25(4):332–343, 2008. doi:10.1080/10580530802384589.

[64] M. H. Jansen-Vullers, M. W. N. C. Loosschilder, P. A. M. Kleingeld, and H. A. Reijers. Performance measures to evaluate the impact of best practices. In Barbara Pernici and Jon Atle Gulla, editors, *Proceedings of Workshops and Doctoral Consortium of the 19th International Conference on Advanced Information Engineering (CAiSE '07), Vol. 1: EMMSAD, BPMDS, BUSITAL*, pages 359–368, 2007.

[65] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, 2000.

[66] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2nd edition, 2002. doi:10.1007/b98835.

[67] Jae-Yoon Jung. Measuring entropy in business process models. In *Proceedings of the 3rd International Conference on Innovative Computing, Information and Control (ICICIC '08)*, page 246, 2008. doi:10.1109/ICICIC.2008.350.

[68] Natalia Juristo and Ana M. Moreno. *Basics of Software Engineering Experimentation*. Kluwer, 2001.

[69] Stephen H. Kan. *Metrics and Models in Software Quality Engineering*. Addison-Wesley, 2nd edition, 2002.

[70] G. Keller, M. Nüttgens, and A.-W. Scheer. Semantische Prozeßmodellierung auf der Grundlage "Ereignisgesteuerter Prozeßzetten (EPK)". Veröffentlichungen des Instituts für Wirtschaftsinformatik (IWi), Universität des Saarlandes , Heft 89, 1992. – In German.

[71] Gerhard Keller and Thomas Teufel. *SAP R/3 Process-oriented Implementation: Iterative Process Prototyping*. Addison-Wesley, 1998.

[72] Ryan K. L. Ko. A computer scientist's introductory guide to business process management (BPM). *Crossroads*, 15(4):11–18, 2009. doi:10.1145/1558897.1558901.

[73] Ryan K. L. Ko, Stephen S. G. Lee, and Eng Wah Lee. Business process management (BPM) standards: a survey. *Business Process Management Journal*, 15(5):744–791, 2009. doi:10.1108/14637150910987937.

[74] Antti M. Latva-Koivisto. Finding a complexity measure for business process models. Research report, Helsinki University of Technology, Systems Analysis Laboratory, 2001.

[75] Gang Soo Lee and Jung-Mo Yoon. An empirical study on the complexity metrics of petri nets. *Microelectronics and Reliability*, 32(3):323–329, 1992. doi:10.1016/0026-2714(92)90061-O.

[76] Hau L. Lee, V. Padmanabhan, and Seungjin Whang. Information distortion in a supply chain: The bullwhip effect. *Management Science*, 43(4):546–558, 1997.

[77] Frank Leymann and Dieter Roller. *Production Workflow: Concepts and Techniques*. Prentice Hall PTR, 2000.

[78] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Texts in Computer Science. Springer, 3rd edition, 2008. doi:10.1007/978-0-387-49820-1.

[79] Udo Lindemann, Maik Maurer, and Thomas Braun. *Structural Complexity Management: An Approach for the Field of Product Design*. Springer, 2009. doi:10.1007/978-3-540-87889-6.

[80] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

[81] Christopher M. Lott and H. Dieter Rombach. Repeatable software engineering experiments for comparing defect-detection techniques. *Empirical Software Engineering*, 1(3):241–277, 1996. doi:10.1007/BF00127447.

[82] Michael Marti. *Complexity Management: Optimizing Product Architecture of Industrial Products*. Gabler Edition Wissenschaft. Deutscher Universitäts-Verlag/GWV Fachverlage, 2007. doi:10.1007/978-3-8350-5435-6 – PhD thesis, Universität St. Gallen, 2007.

[83] Robert M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261(5560):459–467, 1976. doi:10.1038/261459a0.

[84] Richard E. Mayer. Models for understanding. *Review of Educational Research*, 59(1):43–64, 1989. doi:10.3102/00346543059001043.

[85] Georg Mayr. *Die Gesetzmäßigkeit im Gesellschaftsleben: Statistische Studien*, volume 23 of *Naturkräfte*. Oldenbourg, 1877. – In German.

[86] Thomas J. McCabe. A complexity measure. *IEEE Transactions on Software Engineering*, 2(4):308–320, 1976.

[87] Joachim Melcher, Jan Mendling, Hajo A. Reijers, and Detlef Seese. On measuring the understandability of process models (experimental results). Research report, Universität Karlsruhe (TH), Institut AIFB; Humboldt-Universität Berlin; Eindhoven University of Technology, 2009. urn:nbn:de:swb:90-119933, http://digbib.ubka.uni-karlsruhe.de/volltexte/1000011993 (last accessed on 2011-12-01).

[88] Joachim Melcher, Jan Mendling, Hajo A. Reijers, and Detlef Seese. On measuring the understandability of process models. In Stefanie Rinderle-Ma, Shazia Sadiq, and Frank Leymann, editors, *Business Process Management Workshops: Revised Papers of the BPM 2009 International Workshops*, volume 43 of *Lecture Notes in Business Information Processing*, pages 465–476, 2010. doi:10.1007/978-3-642-12186-9_44.

[89] Joachim Melcher and Detlef Seese. Process measurement: Insights from software measurement on measuring process complexity, quality and performance. Research report, Universität Karlsruhe (TH), Institut AIFB, 2008. urn:nbn:de:swb:90-92259, http://digbib.ubka.uni-karlsruhe.de/volltexte/1000009225 (last accessed on 2011-12-01).

[90] Joachim Melcher and Detlef Seese. Towards validating prediction systems for process understandability: Measuring process understandability. In Viorel Negru, Tudor Jebelean, Dana Petcu, and Daniela Zaharie, editors, *Proceedings of the 10th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2008)*, pages 564–571, 2008. doi:10.1109/SYNASC.2008.24.

[91] Joachim Melcher and Detlef Seese. Towards validating prediction systems for process understandability: Measuring process understandability (experimental results). Research report, Universität Karlsruhe (TH), Institut AIFB, 2008. urn:nbn:de:swb:90-92605, http://digbib.ubka.uni-karlsruhe.de/volltexte/1000009260 (last accessed on 2011-12-01).

[92] Joachim Melcher and Detlef Seese. Visualization and clustering of business process collections based on process metric values. Research report, Universität Karlsruhe (TH), Institut AIFB, 2008. urn:nbn:de:swb:90-98483, http://digbib.ubka.uni-karlsruhe.de/volltexte/1000009848 (last accessed on 2011-12-01).

[93] Joachim Melcher and Detlef Seese. Visualization and clustering of business process collections based on process metric values. In Viorel Negru, Tudor Jebelean, Dana Petcu, and Daniela Zaharie, editors, *Proceedings of the 10th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2008)*, pages 572–575, 2008. doi:10.1109/SYNASC.2008.37.

[94] Joachim Melcher and Detlef Seese. Empirical analysis of a proposed process granularity heuristic (experimental details). Research report, Universität Karlsruhe (TH), Institut AIFB, 2009. urn:nbn:de:swb:90-120163, http://digbib.ubka.uni-karlsruhe.de/volltexte/1000012016 (last accessed on 2011-12-01).

[95] Joachim Melcher and Detlef Seese. Empirical analysis of a proposed process granularity heuristic. In Stefanie Rinderle-Ma, Shazia Sadiq,

and Frank Leymann, editors, *Business Process Management Workshops: Revised Papers of the BPM 2009 International Workshops*, volume 43 of *Lecture Notes in Business Information Processing*, pages 513–524, 2010. doi: 10.1007/978-3-642-12186-9_48.

[96] J. Mendling, M. Moser, G. Neumann, H. M. W. Verbeek, B. F. van Dongen, and W. M. P. van der Aalst. A quantitative analysis of faulty EPCs in the SAP reference model. BPM Center Report BPM-06-08, 2006. http://wwwis.win.tue.nl/~wvdaalst/BPMcenter/reports/2006/BPM-06-08.pdf (last accessed on 2011-12-01).

[97] Jan Mendling. Testing density as a complexity metric for EPCs. Technical Report JM-2006-11-15, Vienna University of Economics and Business Administration, 2006. http://www.mendling.com/publications/TR06-density.pdf (last accessed on 2011-12-01).

[98] Jan Mendling. *Detection and Prediction of Errors in EPC Business Process Models*. PhD thesis, Vienna University of Economics and Business Administration, 2007.

[99] Jan Mendling. *Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness*, volume 6 of *Lecture Notes in Business Information Processing*. Springer, 2008. doi:10.1007/978-3-540-89224-3.

[100] Jan Mendling and Gustaf Neumann. Error metrics for business process models. In Johann Eder, Stein L. Tomassen, Andreas Opdahl, and Guttorm Sindre, editors, *Proceedings of the CAiSE '07 Forum at the 19th International Conference on Advanced Information Systems Engineering*, volume 247 of *CEUR Workshop Proceedings*, pages 53–56, 2007. urn:nbn:de:0074-247-5, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-247/FORUM_14.pdf (last accessed on 2011-12-01).

[101] Jan Mendling, Hajo A. Reijers, and Jorge Cardoso. What makes process models understandable? In Gustavo Alonso, Peter Dadam, and Michael Rosemann, editors, *Business Process Management: Proceedings of the 5th International Conference BPM 2007*, volume 4714 of *Lecture Notes in Computer Science*, pages 48–63, 2007. doi:10.1007/978-3-540-75183-0_4.

[102] Jan Mendling and Mark Strembeck. Influence factors of understanding business process models. In Witold Abramowicz and Dieter Fensel, editors, *Business Information Systems: Proceedings of the 11th International Conference BIS 2008*, volume 7 of *Lecture Notes in Business Information Processing*, pages 142–153, 2008. doi:10.1007/978-3-540-79396-0_13.

[103] Annual report 2009. Annual report, Microsoft, 2009. http://www.microsoft.com/investor/reports/ar09/downloads/MS_2009_AR.doc (last accessed on 2011-12-01).

[104] Stanley Milgram. The small-world problem. *Psychology Today*, 1(1):61–67, 1967.

[105] Glenn W. Milligan and Martha C. Cooper. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179, 1985. doi:10.1007/BF02294245.

[106] Frederick C. Mish, editor. *Merriam-Webster's Collegiate Dictionary*. Merriam-Webster, 10th edition, 2001.

[107] Douglas C. Montgomery. *Design and Analysis of Experiments*. Wiley, 6th edition, 2005.

[108] Sandro Morasca. Measuring attributes of concurrent software specifications in petri nets. In *Proceedings of the 6th IEEE International Software Metrics Symposium*, pages 100–110, 1999. doi:10.1109/METRIC.1999.809731.

[109] R. B. Nelsen. Pearson product-moment correlation coefficient. In M. Hazewinkel, editor, *Encyclopaedia of Mathematics*, volume Supplement III, page 301. Kluwer, 2001.

[110] R. B. Nelsen. Spearman rho metric. In M. Hazewinkel, editor, *Encyclopaedia of Mathematics*, volume Supplement III, pages 375–376. Kluwer, 2001.

[111] Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[112] Mark E. Nissen. Redesigning reengineering through measurement-driven inference. *MIS Quarterly*, 22(4):509–534, 1998.

[113] Business process maturity model (BPMM). Specification formal/2008-06-01, Object Management Group, 2008. Version 1.0, http://www.omg.org/spec/BPMM/1.0/PDF (last accessed on 2011-12-01).

[114] State of the business process management market 2008. White paper, Oracle, 2008.

[115] Joseph A. Orlicky, George W. Plossl, and Oliver W. Wight. Structuring the bill of material for MRP. *Production and Inventory Management*, 13(4):19–42, 1972.

[116] Martyn A. Ould. *Business Processes: Modelling and Analysis for Re-engineering and Improvement*. Wiley, 1995.

[117] Michael J. Panik. *Advanced Statistics from an Elementary Point of View*. Elsevier Academic Press, 2005.

[118] Heinz-Otto Peitgen, Hartmut Jürgens, and Dietmar Saupe. *Chaos and Fractals: New Frontiers of Science*. Springer, 2nd edition, 2004. doi:10.1007/b97624.

[119] Business Process Management: Status Quo und Marktentwicklung im Bereich BPM. Management summary, Pentadoc and Trovarit, 2010. – In German.

[120] Shari Lawrence Pfleeger. Design and analysis in software engineering: Part 1: The language of case studies and formal experiments. *ACM SIGSOFT Software Engineering Notes*, 19(4):16–20, 1994. doi:10.1145/190679.190680.

[121] Stephen G. Powell, Markus Schwaninger, and Chris Trimble. Measurement and control of business processes. *System Dynamics Review*, 17(1):63–91, 2001. doi:10.1002/sdr.206.

[122] A. V. Prokhorov. Covariance. In M. Hazewinkel, editor, *Encyclopaedia of Mathematics*, volume 2, page 448. Kluwer, 1988.

[123] Andy Pryke, Sanaz Mostaghim, and Alireza Nazemi. Heatmap visualization of population based multi objective algorithms. In Shigeru Obayashi, Kalyanmoy Deb, Carlo Poloni, Tomoyuki Hiroyasu, and Tadahiko Murata, editors, *Evolutionary Multi-Criterion Optimization: Proceedings of the 4th International Conference EMO 2007*, volume 4403 of *Lecture Notes in Computer Science*, pages 361–375, 2007. doi:10.1007/978-3-540-70928-2_29.

[124] Günter Radons and Reimund Neugebauer, editors. *Nonlinear Dynamics of Production Systems*. Wiley-VCH, 2004. doi:10.1002/3527602585.

[125] Jan Recker and Alexander Dreiling. Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education. In *Proceedings of the 18th Australasian Conference on Information Systems (ACIS 2007)*, pages 356–366, 2007.

[126] H. A. Reijers and J. Mendling. Modularity in process models: Review and effects. In Marlon Dumas, Manfred Reichert, and Ming-Chien Shan, editors, *Business Process Management: Proceedings of the 6th International Conference BPM 2008*, volume 5240 of *Lecture Notes in Computer Science*, pages 20–35, 2008. doi:10.1007/978-3-540-85758-7_5.

[127] Hajo A. Reijers. *Design and Control of Workflow Processes: Business Process Management for the Service Industry*, volume 2617 of *Lecture Notes in Computer Science*. Springer, 2003. doi:10.1007/3-540-36615-6.

[128] Hajo A. Reijers, Selma Limam, and Wil M. P. van der Aalst. Product-based workflow design. *Journal of Management Information Systems*, 20(1):229–262, 2003.

[129] Hajo A. Reijers and Irene T. P. Vanderfeesten. Cohesion and coupling metrics for workflow process design. In Jörg Desel, Barbara Pernici, and Mathias Weske, editors, *Business Process Management: Proceedings of the Second International Conference BPM 2004*, volume 3080 of *Lecture Notes in Computer Science*, pages 290–305, 2004. doi:10.1007/b98280.

[130] Fred S. Roberts. *Measurement Theory with Applications to Decisionmaking, Utility, and the Social Sciences*, volume 7 of *Encyclopedia of Mathematics and Its Applications*. Addison-Wesley, 1979.

[131] Joseph Lee Rodgers and W. Alan Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.

[132] Elvira Rolón, Francisco Ruiz, Félix García, and Mario Piattini. Applying software metrics to evaluate business process models. *CLEI Electronic Journal*, 9(1), 2006.

[133] Per Runeson. Using students as experiment subjects: An analysis on graduate and freshmen student data. In *Proceedings of the 7th International Conference on Empirical Assessment & Evaluation in Software Engineering (EASE '03)*, pages 95–102, 2003.

[134] Lothar Sachs and Jürgen Hedderich. *Angewandte Statistik: Methodensammlung mit R*. Springer, 12th edition, 2006. `doi:10.1007/978-3-540-32161-3` – In German.

[135] Laura Sanchez, Andrea Delgado, Francisco Ruiz, Félix García, and Mario Piattini. Measurement and maturity of business processes. In Jorge Cardoso and Wil van der Aalst, editors, *Handbook of Research on Business Process Modeling*, pages 532–556. Information Science Reference, 2009.

[136] Laura Sánchez González, Félix García Rubio, Francisco Ruiz González, and Mario Piattini Velthuis. Measurement in business processes: a systematic review. *Business Process Management Journal*, 16(1):114–134, 2010. `doi:10.1108/14637151011017976`.

[137] Clarity fosters innovation: Annual report 2009. Annual report, SAP, 2010. `http://www.sap.com/corporate-en/investors/reports/annualreport/2009/pdf/SAP_2009_Annual_Report.pdf` (last accessed on 2011-12-01).

[138] Kamyar Sarshar, Philipp Dominitzki, and Peter Loos. Comparing the control-flow of EPC and Petri net from the end-user perspective: Statistical results of a laboratory experiment. Working Paper 25, Universität Mainz, Information Systems & Management (ISYM), 2005. `urn:nbn:de:0006-0252`, `http://d-nb.info/975462385/34` (last accessed on 2011-12-01).

[139] Kamyar Sarshar and Peter Loos. Comparing the control-flow of EPC and Petri net from the end-user perspective. In W. M. P. van der Aalst, Boualem Benatallah, Fabio Casati, and Francisco Curbera, editors, *Business Process Management: Proceedings of the 3rd International Conference BPM 2005*, volume 3649 of *Lecture Notes in Computer Science*, pages 434–439, 2005. `doi:10.1007/11538394_36`.

[140] Sven Schnägelberger. BPM: mehr Hype als Realität? *Computerwoche*, (14):21, 2008. – In German.

[141] Detlef Seese. Complexity management. Lecture 1: Introduction. Lecture notes, Karlsruher Institut für Technologie, Institut AIFB, 2010.

[142] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4):591–611, 1965. `doi:10.1093/biomet/52.3-4.591`.

[143] Micheal Sipser. *Introduction to the Theory of Computation*. Cengage Learning, 2nd international edition, 2006.

[144] Peter J. Smith. *Into Statistics: A Guide to Understanding Statistical Concepts in Engineering and the Sciences*. Springer, 2nd edition, 1998.

[145] Eduardo D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, volume 6 of *Texts in Applied Mathematics*. Springer, 2nd edition, 1998.

[146] Dieter Spath and Anette Weisbecker, editors. *Business Process Management Tools 2008: Eine evaluierende Marktstudie zu aktuellen Werkzeugen*. Fraunhofer IRB Verlag, 2008. – In German.

[147] John D. Sterman. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, 35(3):321–339, 1989.

[148] S. S. Stevens. On the theory of scales of measurement. *Science*, 103(2684):677–680, 1946. `doi:10.1126/science.103.2684.677`.

[149] W. P. Stevens, G. J. Myers, and L. L. Constantine. Structured design. *IBM Systems Journal*, 13(2):115–139, 1974. `doi:10.1147/sj.132.0115`.

[150] David Stirzaker. *Elementary Probability*. Cambridge University Press, 2nd edition, 2003.

[151] Ramanath Subramanyam and M. S. Krishnan. Empirical analysis of CK metrics for object-oriented design complexity: Implications for software defects. *IEEE Transactions on Software Engineering*, 29(4):297–310, 2003. `doi:10.1109/TSE.2003.1191795`.

[152] H. N. V. Temperley. *Graph Theory and Applications*. Ellis Horwood Series in Mathematics and Its Applications. Ellis Horwood, 1981.

[153] Warren S. Torgerson. *Theory and Methods of Scaling*. Wiley, 1958.

[154] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[155] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society (2nd Series)*, 42:230–265, 1937. doi:10.1112/plms/s2-42.1.230.

[156] Graham Upton and Ian Cook. *Understanding Statistics*. Oxford University Press, 1996.

[157] W. M. P. van der Aalst. Designing workflows based on product structures. In K. Li, S. Olariu, Y. Pan, and I. Stojmenovic, editors, *Proceedings of the Ninth IASTED International Conference on Parallel and Distributed Computing Systems (PDCS 1997)*, pages 337–342, 1997.

[158] W. M. P. van der Aalst. On the automatic generation of workflow processes based on product structures. *Computers in Industry*, 39(2):97–111, 1999. doi:10.1016/S0166-3615(99)00007-X.

[159] W. M. P. van der Aalst, H. A. Reijers, and S. Limam. Product-driven workflow design. In Weiming Shen, Zongkai Lin, Jean-Paul Barthès, and Mohamed Kamel, editors, *Proceedings of the Sixth International Conference on Computer Supported Cooperative Work in Design*, pages 397–402, 2001. doi:10.1109/CSCWD.2001.942292.

[160] W. M. P. van der Aalst, H. A. Reijers, A. J. M. M. Weijters, B. F. van Dongen, A. K. Alves de Medeiros, M. Song, and H. M. W. Verbeek. Business process mining: An industrial application. *Information Systems*, 32(5):713–732, 2007. doi:10.1016/j.is.2006.05.003.

[161] W. M. P. van der Aalst, A. H. M. ter Hofstede, B. Kiepuszewski, and A. P. Barros. Workflow patterns. *Distributed and Parallel Databases*, 14(1):5–51, 2003. doi:10.1023/A:1022883727209.

[162] Wil van der Aalst and Kees van Hee. *Workflow Management: Models, Methods, and Systems*. Cooperative Information Systems. MIT Press, 2002.

[163] Wil M. P. van der Aalst, Arthur H. M. ter Hofstede, and Mathias Weske. Business process management: A survey. In Wil van der Aalst, Arthur ter Hofstede, and Mathias Weske, editors, *Business Process Management: Proceedings of the International Conference BPM 2003*, volume 2678 of *Lecture Notes in Computer Science*, pages 1–12, 2003. doi:10.1007/3-540-44895-0_1.

[164] Boudewijn van Dongen, Remco Dijkman, and Jan Mendling. Measuring similarity between business process models. In Zohra Bellahsène and Michel Léonard, editors, *Advanced Information Systems Engineering: Proceedings of the 20th International Conference CAiSE 2008*, volume 5074 of *Lecture Notes in Computer Science*, pages 450–464, 2008. doi:10.1007/978-3-540-69534-9_34.

[165] Irene Vanderfeesten, Jorge Cardoso, Jan Mendling, Hajo A. Reijers, and Wil van der Aalst. Quality metrics for business process models. In Layna Fischer, editor, *BPM and Workflow Handbook 2007*, pages 179–190. Future Strategies Inc., 2007.

[166] Irene Vanderfeesten, Jorge Cardoso, and Hajo A. Reijers. A weighted coupling metric for business process models. In Johann Eder, Stein L. Tomassen, Andreas L. Opdahl, and Guttorm Sindre, editors, *Proceedings of the CAiSE'07 Forum at the 19th International Conference on Advanced Information Systems Engineering*, pages 41–44, 2007.

[167] Irene Vanderfeesten, Hajo A. Reijers, Jan Mendling, Wil M. P. van der Aalst, and Jorge Cardoso. On a quest for good process models: The cross-connectivity metric. In Zohra Bellahsène and Michel Léonard, editors, *Advanced Information Systems Engineering: Proceedings of the 20th International Conference CAiSE 2008*, volume 5074 of *Lecture Notes in Computer Science*, pages 480–494, 2008. `doi:10.1007/978-3-540-69534-9_36`.

[168] Irene Vanderfeesten, Hajo A. Reijers, and Wil M. P. van der Aalst. Evaluating workflow process designs using cohesion and coupling metrics. *Computers in Industry*, 59(5):420–437, 2008. `doi:10.1016/j.compind.2007.12.007`.

[169] Juha Vesanto and Esa Alhoniemi. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3):586–600, 2000. `doi:10.1109/72.846731`.

[170] M. I. Voĭtsekhovskiĭ. Metric. In M. Hazewinkel, editor, *Encyclopaedia of Mathematics*, volume 6, pages 206–207. Kluwer, 1990.

[171] Ann E. Watkins, Richard L. Scheaffer, and George W. Cobb. *Statistics in Action: Understanding a World of Data*. Key Curriculum Press, 2004.

[172] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998. `doi:10.1038/30918`.

[173] Chris Weber. Complexity is in the eye of the beholder. Blog entry, 2008. `http://pragmaticprose.com/complexity-is-in-the-eye-of-the-beholder` (last accessed on 2011-12-01).

[174] Gerald M. Weinberg and Daniela Weinberg. *On the Design of Stable Systems*. Wiley Series on Systems Engineering and Analysis. Wiley, 1979.

[175] Mathias Weske. *Business Process Management: Concepts, Languages, Architectures*. Springer, 2007. `doi:10.1007/978-3-540-73522-9`.

[176] Terminology & glossary. Specification WFMC-TC-1011, Workflow Management Coalition, 1999. Issue 3.0, `http://www.wfmc.org/standards/docs/TC-1011_term_glossary_v3.pdf` (last accessed on 2011-12-01).

[177] Leland Wilkinson and Michael Friendly. The history of the cluster heat map. *The American Statistician*, 63(2):179–184, 2009. `doi:10.1198/tas.2009.0033`.

[178] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in Software Engineering: An Introduction*, volume 6 of *Kluwer International Series in Software Engineering*. Kluwer, 2000.

[179] Edward Yourdon and Larry L. Constantine. *Structured Design: Fundamentals of a Discipline of Computer Program and Systems Design*. Prentice-Hall, 1979.

[180] Horst Zuse. *Software Complexity: Measures and Methods*, volume 4 of *Programming Complex Systems*. Walter de Gruyter, 1991.

ISBN 978-3-86644-789-9