Issue 203

December 2016

Institut für Wirtschaftsinformatik

Towards the Development of Stylized Facts on the Understandability of Graphical Business Process Models

Constantin Houy, Peter Fettke, Peter Loos

Publications of the Institute for Information Systems (IWi)
at the German Research Center for Artificial Intelligence (DFKI)

Editor: Prof. Dr. Peter Loos

C. HOUY, P. FETTKE , P. LOOS

Towards the Development of Stylized Facts on the Understandability of Graphical Business Process Models[1]

Publications of the Institute for Information Systems (IWi)
Editor: Prof. Dr. Peter Loos

Issue 203

ISSN 1438-5678

Institute for Information Systems (IWi)
at the German Research Center for Artificial Intelligence (DFKI) GmbH
Stuhlsatzenhausweg 3, Geb. D3 2, 66123 Saarbrücken, Germany
Telefon: +49 (0) 6 81 / 85775 – 31 06, Fax: +49 (0) 6 81 / 85775 – 36 96
E-Mail: iwi@iwi.uni-sb.de, URL: http:/ / www.iwi.uni-sb.de/

December 2016

[1] This report presents the idea of developing Stylized Facts concerning the understandability of graphical business process models and the current state of progress of an on-going dissertation project which started in 2014. The presented idea and work is supposed to result in a major part of the doctoral dissertation of Constantin Houy, the first author of this report.

Abstract

The development of theory is one of the major tasks of every scientific discipline, and thus of Information Systems Research (ISR) as well as Business Informatics (BI). While different approaches can be used to develop theory in ISR and BI, there is one "dominant" way of IS theory development which has been described by GROVER and LYYTINEN in a recent article published in MISQ as the common "epistemic script". The authors criticize this epistemic script for promoting a quite restricted production of IS-related knowledge. Furthermore, GROVER and LYYTINEN, identify new potential ways of overcoming the common epistemic script and propose – among others – the concept of Stylized Facts (SF) as one potential way for innovative knowledge production in ISR and BI.

Against the background that we – the authors of this report – have been using Stylized Facts as a research approach for some years and can confirm the potential of this approach, the following report presents the idea and the current state of a promising comprehensive dissertation project (first author of this report) using Stylized Facts in ISR and BI which started in 2014. In the following, the idea of developing Stylized Facts regarding the understandability of graphical business process models is elaborated. Besides the presentation of an approach for a transparent development of SF, a comprehensive application example will illustrate the derivation of a SF regarding the relationships of the structuredness of business process models and the resulting model understandability.

Keywords: Stylized Facts, Model Understandability, Business Process Modeling, Qualitative Research, Quantitative Research, Meta-Analysis, State-of-the-Art, Review

 December 2016

Content

Figures

Tables

List of Abbreviations

AS	Aggregated Statement
AS_L	Aggregated Statement concerning Language Characteristics
AS_M	Aggregated Statement concerning Model Characteristics
AS_O	Aggregated Statement concerning other Characteristics
AS_P	Aggregated Statement concerning Personal Characteristics
BI	Business Informatics
CAIS	Communications of the Association for Information Systems
EPC	Event-driven Process Chain
Fx	Finding x
GA	Genetic Algorithm
GP	Genetic Programming
GS	Generalized Statement
GS_L	Generalized Statement concerning Language Characteristics
GS_M	Generalized Statement concerning Model Characteristics
GS_O	Generalized Statement concerning other Characteristics
GS_P	Generalized Statement concerning Personal Characteristics
IS	Information Systems
ISR	Information Systems Research
IT	Information Technology
LoE	Level of Evidence
MISQ	Management Information Systems Quarterly
Sx	Study x
Sx_Fy	Finding y within Study x
SF	Stylized Fact
TR	Technological Rule

1 Introduction

The development of theory is one of the major tasks of every scientific discipline, and thus of Information Systems Research (ISR) and Business Informatics (BI).[2] While different approaches can be used for the development of theory, e. g. qualitative methods for building initial theory models and quantitative methods for falsifying existing theoretical models, there seems to be one "dominant" way of developing theory in ISR and BI which has been described by GROVER and LYYTINEN in a recent MISQ article named "New State of Play in Information Systems Research: The Push to the Edges". They call this "dominant way of producing knowledge" in ISR the common "epistemic script" which "seeks to domesticate high-level reference theory in the form of mid-level abstractions involving generic and atheoretical information technology (IT) components. Enacting such epistemic scripts squeezes IS theory to the middle range, where abstract reference theory concepts are directly instantiated or slightly modified to the IS context". Against this background of a quite restricted way of producing IS knowledge and theory, the authors invite "individual scholars to be more open to practices that permit richer theorizing".[3]

While the concept of Stylized Facts (SF) has been discussed as an interesting approach supporting theory development in ISR and BI,[4] SF seem to offer particular potential in the context of the search for new and innovative ways to overcome the common "epistemic script" in ISR by institutionalizing a "data-driven, inductive research" approach.[5] GROVER and LYYTINEN name Stylized Facts as one interesting way of conducting data-driven research in order to re-establish and strengthen new ways of developing theory in ISR. However, so far only a few studies using the concept of Stylized Facts are known in ISR and BI research and SF are far from being an established research approach in this field.

---

[2] Cf. BICHLER ET AL. (2016), p. 292. For the delineation of different academic disciplines studying information systems, such as ISR, BI and others, as well as their own focus and theoretical backgrounds, see the contribution of FETTKE in this panel discussion (FETTKE (2016): Towards a Coherent View on Information Systems. In: BICHLER ET AL. (2016), pp. 296-301.)

[3] GROVER ET AL. (2015), p. 271.

[4] Cf. LOOS ET AL. (2011), cf. HOUY ET AL. (2015).

[5] Cf. GROVER ET AL. (2015), p. 285.

However, the authors of this report have contributed to the following studies using SF as a research approach in ISR and BI:

Nr.	Source	Topic
1.	HOUY ET AL. (2009) (in German)	First, more detailed description of the idea of using SF in ISR and BI for theory development and presentation of an application example focusing on EPC as a business process modeling language
2.	HOUY ET AL. (2011)(in German)	Description of the general potential of SF for theory development in ISR and BI, conceptual work focusing on methodological aspects
3.	LOOS ET AL. (2011)	Discussion panel regarding the potential of SF for ISR and BI theory development
4.	REITER ET AL. (2013)	Exemplary application of the SF approach in the context of ERP systems for the evaluation of existing theory
5.	HOUY ET AL. (2013)	Discussion of the general potential of SF for theory development in ISR and BI
6.	HOUY ET AL. (2015)	Comprehensive introduction of the potential of SF for theory development in ISR and BI and a more comprehensive application example using studies on EPCs as a business process modeling language

Table 1: *Overview "Stylized Facts in ISR and BI"*

Against that background, this report presents the idea and current state of a promising and comprehensive dissertation project using Stylized Facts in ISR and BI which started in 2014.[6] This research report presents the idea of developing Stylized Facts concerning the understandability of graphical business process models. After this introduction, the

[6] The presented idea and work is supposed to result in a major part of the doctoral dissertation of Constantin Houy, the first author of this report.

basic idea of this research endeavour as well as the underlying conceptualizations and the methodical approach are presented in section two. Furthermore, an overview of the current results will be given in this section. Section three presents a comprehensive example of the derivation process and results of one SF concerning the relationships between *process model structure* and *model understandability*. Section four shortly discusses the findings and current state of results before section five concludes this report.

2 Basic Idea and Current State of Research Progress

2.1 Preliminary Notes

We have defined the concept of Stylized Facts (SF) in a more detailed manner in a recent article published in CAIS as follows:[7]

> "*Stylized facts (SFs) constitute knowledge in the form of generalized and simplified statements describing interesting characteristics and relationships concerning empirically observable phenomena.*[8] *SFs can be conceptualized as interesting, sometimes counterintuitive, patterns in empirical data (empirical generalizations, accumulations of evidence) documented in different sources. An important characteristic of SFs is their focus on the most relevant aspects of observable phenomena by abstracting from details (stylization). Thus, SFs are broadly supported and simplified representations of complex relationships that are not necessarily valid in every situation and context.*[9] *SFs do not aim to represent causal relationships but rather interesting correlations that are observable in reality. Thus, reducing the complexity of real-world phenomena, SFs can – according to Stephan Zelewski – serve as "a 'seed of crystallization' for the construction and critical review of [.] models or theories".*[10] *Kaldor (1961) introduced the SF concept in the context of macroeconomic growth theory to compare the explanatory power of existing economic models and support the development of new theoretical models that should be able to explain empirically observable phenomena.*[11]"

---

[7] HOUY ET AL. (2015), p. 228.

[8] Cf. HEINE ET AL. (2005); HELFAT (2007).

[9] Cf. HEINE ET AL. (2007); HOUY ET AL. (2011); HOUY ET AL. (2013).

[10] ZELEWSKI in LOOS ET AL. (2011), p. 112.

[11] Cf. KALDOR (1961).

Furthermore, we have described a procedure model for the development of SF in ISR and BI which is visualized in the following figure.

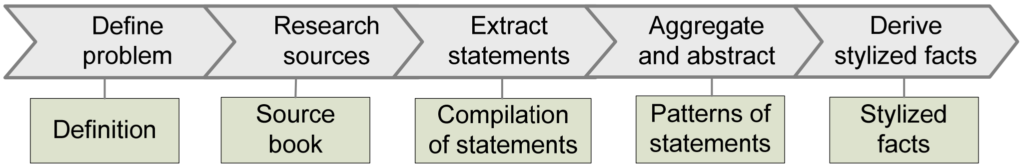


Figure 1: *Procedure model for the development of Stylized Facts*[12]

In the following, we describe the application of this procedure model in the dissertation project on the understandability of graphical business process models and give an overview of the current state of results.

2.2 Research Procedure and Overview of Results

In the following passage, the different phases of the above procedure model and its usage in the exemplary application context are described in more detail.

1. *Define problem*

To develop SF regarding business process model understandability, relevant literature sources containing knowledge on this topic are needed. Therefore, it is necessary to define the problem and to determine the relevant content. Relevant sources are those which contain statements regarding reliable relationships (potential causes and effects) in the context of perceiving, reading and understanding business process models. In the following, only literature sources stemming from academic publication outlets such as scientific conferences and journals were used. In order to assure inter-subjective confirmability and traceability of the literature selection procedure, a structured literature research process was performed which will be described in more detail in the following.

2. *Research sources*

In the context of the structured literature source research, the literature database SCOPUS has been used.[13] In order to find relevant sources concerning business process model understandability, it was first searched for appropriate sources treating business process models and business process modeling languages using the following terms:

---

[12] The procedure model is based on the contributions published by WEIßENBERGER ET AL. (2007) and HEINE ET AL. (2007) and was also used in HOUY ET AL. (2009); HOUY ET AL. (2011); HOUY ET AL. (2013); HOUY ET AL. (2015).

[13] http://www.scopus.com/

> *"process model*" OR "process descri*" OR "process diagram*" OR "business process*" OR bpmn OR epc OR "petri net*" OR "UML Activity" OR "UML collaboration" OR yawl*

Furthermore, the amount of retrieved sources was limited by selecting only those which particularly treat the topics "understandability", "comprehension", "making sense of models", "cognitive aspects" and "perception processes" using for the following search terms:

> *(understandab* OR comprehens* OR understanding OR comprehending OR "making sense" OR complexity OR cognitive OR perce*) AND "business process")*

The mentioned search terms have also been used in the context of an in-depth investigation of the theoretical foundations of business process model understandability research published in the proceedings of the ECIS 2014.[14] In this research, a total amount of 121 articles was identified using the above mentioned literature database. A deeper investigation of these 121 articles' relevance resulted in a reduced amount of 88 corresponding articles. The above search has been performed several times even after the ECIS article has been published in order to keep the amount of relevant articles up-to-date. Furthermore, the reference sections of identified articles have been used to find more relevant articles which could not be found by means of the database search ("backward search"). This literature research for the development of SF has been completed in May 2015, while newly published articles on the topic will, nevertheless, be considered and kept in mind when discussing the results. In total an amount of 101 journal articles, conference and workshop articles as well as relevant doctoral dissertations have been included in the process of developing SF on the topic "business process model understandability" in the presented research project. In the following step, relevant statements made in the investigated sources were extracted, which will be explained in more detail in the following section.

3. *Extract statements*

In the next step, the 101 contributions were analyzed and relevant statements concerning business process model understandability were extracted. In total, 1004 separate text passages were documented including the "context", the used "research method", the analyzed "independent variable / treatment" – if available – as well as the "conceptualization of understanding and understandability / dependent variable" – if available – of the underlying study. Each documented text passage was uniquely indexed by means a unique "study" number (e.g. *S54*) and a unique "finding" number (e.g. *F11*) and can,

[14] Cf. HOUY ET AL. (2014).

thus be identified and retrieved via this primary key (e.g. S54_F11). Moreover, further information was documented concerning each text passage: 1.) the according page number in the original source, 2.) the underlying research method on which the statement is based, 3.) the text passage's *Level of Evidence* (LoE). We differentiate between the following *Levels of Evidence* presented in table 1 which have been similarly introduced in FETTKE et al. (2010) in order to assess the validity and the available support of given statements:[15]

I	plausible statement without further justification
IIa	plausible statement backed up by conceptual consideration and argumentation (without empirical evidence or references)
IIb	plausible statement backed up by conceptual consideration, argumentation and one or more literature references
III	statement which is backed up by exemplary experience (e. g. by a single or a few known cases)
IV	statement which has held good in a variety of applications and cases

Table 2: *Levels of Evidence (LoE)*

Furthermore, it was documented whether the text passage contains so-called *technological rules* representing reliable means-end-relationships which can give hints for successful possibilities of action to improve business process model understandability.

Important conventions which were considered during the extraction of statements from the original sources and which proved to be useful are the following:

(1) Only those text passages were selected which contain relevant statements concerning business process model understandability. Passages containing a mere enumeration of influence factors on business process model understandability and not indicating whether a certain factor has a positive or a negative influence were not considered.

(2) As it is the goal of the presented research endeavor to develop basic, generalized and reliable statements (SF) concerning relevant influence factors on business process model understandability, no text passages were considered which merely compare different modeling languages such as Event-driven Process Chains (EPC) or Petri Nets, e.g. statements like "EPCs are easier to understand than Petri Nets".

[15] FETTKE ET AL. (2010), pp. 353-354.

(3) Relevant text passages were extracted from original sources and documented without any changes. Furthermore, they will be completely displayed in the appendix of the final documentation – and in our comprehensive example in section three – in order to assure a transparent and inter-subjectively comprehensible development process of SF.

(4) References displayed in extracted text passages are documented as in the original sources (original citation style) and were not reformatted.

(5) No additions were made to the extracted and documented text passages. Exceptions from this rule were short explanations concerning the meaning of abbreviations which are given in square brackets, e.g. „GP [“genetic programming”] and GA [“genetic algorithm”], in order to improve the readability of such passages. This was necessary because the meaning of several abbreviations is not always obvious.

(6) Sometimes, tables and figures were also documented, especially when presenting relevant information on business process model understandability in compressed form which would take significantly more space when described in textual representation.

Concerning the documentation of relevant content it has to be stated that in total 1004 classified and categorized text passages have been extracted (more than 122.000 words) which contain interesting statements about business process model understandability. In the following step, the relevant content was aggregated and particular details which are irrelevant for the development of SF were transparently eliminated (“abstraction”).

4. *Aggregate and Abstract*

In the next steps, the content of the developed collection of classified and categorized statements was analyzed. In this context a collection of simple (abstracted) statements, which are as “atomic” as possible, was developed. “Atomic” means that a statement should possibly only address one single issue in the context of business process model understandability. In this collection of aggregated statements (AS), each original documented finding (F) from the underlying study (S), e.g. S54_F11, is clearly assigned to the aggregated statement (AS) it is supporting. Sometimes, extracted text passages can support several different atomic statements concerning business process model understandability. Thus, multiple assignments of one finding in a study (Sx_Fy) to aggregated statements (AS) are possible.[16] In total, 373 different aggregated statements (AS) concerning several topics of business process model understandability were developed. The AS were assigned to different appropriate topical clusters. The following topical categories were found and will be used for the presentation of results:

[16] This classification system is inspired by the work of STRANGFELD (2012) who used Stylized Facts in the context of computer-based simulation and presented a comprehensive conceptualization of SF and an elaborated development approach.

(1) Aggregated statements concerning characteristics of process modeling languages (Abbreviation: *AS_L*, number: *115* statements),
(2) Aggregated statements concerning characteristics of process models (Abbreviation: *AS_M*, number: *179* statements),
(3) Aggregated statements concerning personal characteristics of model viewers or users (Abbreviation: *AS_P*, number: *80* statements) and
(4) Aggregated statements concerning other findings on process model understandability (Abbreviation: *AS_O*, number: *27* statements).

If you add the above numbers of AS, the result is 401. This number is larger than 373 because it was not always possible to assign each AS to exactly one topical cluster because more than one topic was addressed in particular statements. There are, e.g., statements on relationships between combinations of characteristics on the one side and process model understanding on the other side, such as the combination of particular process model characteristics (e.g. *model complexity*) and particular personal characteristics of the model viewer (e.g. *modeling experience*). Such statements were assigned to several categories, e.g. model-related (*AS_M*) and personal characteristics (*AS_P*).

The further consolidation and concentration of knowledge was performed in consideration of the developed topical clusters. Therefore, in each topical cluster different detailed sub-topics were inductively developed based on the available AS. In this context, the following sub-topic categories emerged:

(1) AS_L: AS concerning characteristics of process modeling languages:
   a. On the general influence of modeling languages,
   b. Primary notation and language constructs,
   c. Modeling paradigm and modeling languages,
   d. On the fit of tasks to be performed and modeling languages,
   e. Modeling languages and domain-specific content,
   f. On the combination of graphical elements and text ("dual coding") and
   g. Process model hierarchies and specific modeling languages.

(2) AS_M: AS concerning characteristics of process models:
   a. Model design (secondary notation),
   b. Model labels,
   c. Model complexity,
   d. Modularity and modularization,
   e. Views and perspectives on models,
   f. On the fit of tasks to be performed and the model purpose and
   g. Domain-specific issues and the content of process models.

(3) AS_P: AS concerning personal characteristics of model viewers or users:

a. A person's education and training in the field of process modeling,
b. A person's experience in the field of process modeling,
c. Reading strategies and techniques,
d. A person's familiarity with a process modeling language,
e. A person's knowledge of the domain addressed by a model,
f. A person's cognitive style, learning type and motivation, and
g. Other findings related to personal characteristics.

(4) AS_O: AS concerning other findings on process model understandability:

a. Effects of modeling guidelines,
b. Approaches for the measurement of model quality,
c. Additional textual context information,
d. Process mining,
e. Refactoring and automated model transformation and
f. Influence of the modeling process.

The introduction of these sub-topic categories supports a further-going consolidation and concentration of available knowledge on business process model understandability in a transparent and inter-subjectively accessible way when developing the SF.

5. Derive stylized facts

In the next step, all available aggregated statements in the sub-topic categories were further consolidated and concentrated by eliminating details from the different AS and further aggregating compatible statements. In total, 102 SF on process model understandability were elaborated. These will provide the basis for the development of specific theoretical models describing the observed relationships in each topical cluster or even in several sub-topics in a broader context. In the following section, an example of one developed SF and its support by the underlying material will be demonstrated.

3 A Stylized Fact on Structuredness and Understandability

3.1 Preliminary notes

In the following, the derivation of one SF will be presented. SF are based on aggregated statements (AS). Aggregated statements themselves are based on findings (F) of different studies (S). Hence, the development of a SF is an inductive process which is based on original findings. The following material will, nevertheless, be presented top-down in the following order: (1) SF → (2) AS → (3) F.

This supports a transparent and understandable access to the material. However, the material can also be read the other way around in order to follow the inductive process of developing the SF: (1) F → (2) AS → (3) SF.

Table 3 presents a legend with relevant abbreviations concerning the following content.

1. Basic Methods	**SU**	Survey
	LE / FE	Laboratory experiment / Field experiment
	CS	Case study
	SI	Simulation
	DO	Design-oriented research / Prototyping
	CA	Conceptual or argumentative analysis
	EI	Expert interview
2. Level of Evidence (LoE)	**I**	Plausible statement without further justification
	IIa	Plausible statement backed up by conceptual consideration and argumentation (without empirical evidence or references)
	IIb	Plausible statement backed up by conceptual consideration, argumentation and one or more literature references
	III	Statement which is backed up by exemplary experience (e. g. by a single or a few known cases)
	IV	Statement which has held good in a variety of applications and cases

Table 3: *Legend*

Furthermore, the text passages of the documented *findings* which were relevant for the SF development are each marked in ***bold and italics*** in the following tables.

3.2 SF_M: “Structuredness and Process Model Understandability”

The presented SF regards structuredness as a process model characteristic:

SF_M: „The more structured a process model is (“split connectors do match a corresponding join connector”) the easier the model will be understood. Accordingly, the less structured a process model is in comparison, the more difficult it is to understand.”

This SF is addressed by a total of 24 different studies considered in this project. In this context, four AS (AS_M_6, AS_M_14, AS_M_60, AS_M_91) were developed supporting the SF.[17] Furthermore, there is one statement not supporting the SF (S19_F6).

3.3 The Aggregated Statements supporting the Stylized Fact

AS_M_6		**"Process models which are well-structured – containing split connectors which do match a corresponding join connector – are easier to understand."**								
total # studies / total # references		ref. **LoE I** # ref. / # studies		ref. **LoE IIa** # ref. / # studies		ref. **LoE IIb** # ref. / # studies		ref. **LoE III** # ref. / # studies		ref. **LoE IV** # ref. / # studies
		S13_F3 (CA) S19_F3 (CA) S22_F3 (CA)		S30_F5 (CA) S58_F1 (CA)		S11_F1 (CA) S30_F11 (CA) S38_F9 (CA) S47_F4 (CA) S47_F5 (CA) S53_F9 (CA) S54_F11 (CA) S59_F2 (CA) S67_F2 (CA) S67_F6 (CA) S70_F3 (CA) S70_F5 (CA) S73_F14 (CA) S87_F5 (CA) S91_F6 (CA) S100_F2 (CA)		S13_F12 (EI)		S70_F9 (SU) S70_F12 (SU) S75_F4 (LE)
18	**25**	**3**	**3**	**2**	**2**	**16**	**13**	**1**	**1**	**3**	**2**

Table 4: *AS_M_6*

AS_M_14		**"Process models which are not well-structured – containing split connectors which do not match a corresponding join connector (typically measured as "gateway mismatch") – often contain deadlocks and are more difficult to understand."**									
total # studies / total # references		ref. **LoE I** # ref. / # studies		ref. **LoE IIa** # ref. / # studies		ref. **LoE IIb** # ref. / # studies		ref. **LoE III** # ref. / # studies		ref. **LoE IV** # ref. / # studies	
		S11_F4 (CA)		S9_F2 (CA)		S11_F1 (CA) S38_F9 (CA) S54_F11 (CA) S70_F1 (CA) S87_F5 (CA)		-		S40_F2 (LE) S40_F4 (LE) S40_F5 (LE) S40_F6 (LE) S40_F7 (LE) S41_F1 (LE) S41_F3 (LE) S41_F5 (LE) S70_F9 (SU) S70_F12 (SU) S75_F3 (LE) S75_F4 (LE)	
9	**19**	**1**	**1**	**1**	**1**	**5**	**5**	**-**	**-**	**12**	**4**

Table 5: *AS_M_14*

AS_M_60		**"Node duplication in process models ("controlled redundancy") can improve the structuredness of a model and can thus improve model understandability."**									
total # studies / total # references		ref. **LoE I** # ref. / # studies		ref. **LoE IIa** # ref. / # studies		ref. **LoE IIb** # ref. / # studies		ref. **LoE III** # ref. / # studies		ref. **LoE IV** # ref. / # studies	
		-		S1_F2 (CA)		S47_F5 (CA) S84_F5 (CA) S87_F7 (CA)		-		-	
4	**4**	**-**	**-**	**1**	**1**	**3**	**3**	**-**	**-**	**-**	**-**

Table 6: *AS_M_60*

[17] One study can support several aggregated statements. This is why adding up the numbers for "total # studies" concerning this SF does not equal 24.

AS_M_91		“Structuring process models can result in better understandability due to the decrease of diagrammatic complexity and, thus, cognitive load.”									
total # studies / total # references		ref. **LoE I** # ref. / # studies		ref. **LoE IIa** # ref. / # studies		ref. **LoE IIb** # ref. / # studies		ref. **LoE III** # ref. / # studies		ref. **LoE IV** # ref. / # studies	
		-		-		S45_F11 (CA) S47_F4 (CA) S47_F5 (CA)		-		-	
2	**3**	**-**	**-**	**-**	**-**	**3**	**2**	**-**	**-**	**-**	**-**

Table 7: *AS_M_91*

3.4 The Findings not supporting the Stylized Fact

S19_F6:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S19	MENDLING ET AL. (2008)	investigation of influence factors on process model understandability	laboratory experiment (n=42, p. 147)	1. personal factors 2. model factors 3. content-related factors	1. correctly answering questions about the model content per person (PSCORE) / per model (MSCORE) 2. time needed to answer questions

code	statements / findings	ref.	meth.	LoE	TR
F6	“Figure 4 gives an overview of the MSCORE that the different participants achieved per model. The mean percentage was 70% across the models. The model with the lowest MSCORE had on average 60% correct answers. This model had loops and parallel execution paths. From the variables mentioned in **M1** and **M2** only SEPARABILITY had a significant correlation according to Spearman with mscore of 0.886 (p=0.019). This strongly confirms the hypothetical impact direction of **M2**. The other variables showed a direction of correlation as expected, but without a sufficient significance. ***As an exception, structuredness had zero correlation in our sample.***“	p. 149	LE	IV	no

Table 8: *S19*

3.5 The Findings supporting the Aggregated Statements

3.5.1 Aggregated Statement “*AS_M_6*”

Level of Evidence I:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S13	MENDLING ET AL. (2007)	investigation of the influence of personal and model characteristics on process model understandability	field experiment (n=73, p. 52) and expert interview (n=12, p. 60)	1. personal characteristics of model readers 2. model characteristics	1. perceived ease of understanding 2. correctly answering questions about the model (reg. order, concurrency, exclusiveness, repetition) 3. relative perceived understandability (ranking of models)

code	statements / findings	ref.	meth.	LoE	TR
F3	“In particular, we expect that the perceived difficulty of a process model (PERCEIVED) would be negatively connected with the score as an operationalization of actual understandability. The same positive connection is assumed with THEORY and PRACTICE while the count metrics #NODE, etc., and the DIAMETER of the process model (i.e. the longest path) should be related to a lower understandability. The precise formulae for calculating these and the following metrics are presented in [8]. ***The SEQUENTIALITY, i.e. the degree to which the model is constructed of task sequences, is expected to be positively connected with understandability. The same is expected for*** SEPARABILITY, which relates to the degree of articulation points in a model (i.e. nodes whose deletion separates the process model into multiple components), and ***STRUCTUREDNESS, which relates to how far a process model is built by nesting blocks of matching join and split routing elements***. Both CONNECTIVITY and DENSITY relate arcs to nodes: the former by dividing #arcs by #nodes, the latter by dividing #arcs to the maximally possible number of arcs. The TOKEN SPLIT metric captures how many new tokens can be introduced by AND- and OR-splits. It should be negatively connected with understandability. The AVERAGE and MAXIMUM CONNECTOR DEGREE refer to the number of input and output arcs of a routing element, which are expected to be negatively connected with SCORE. The same expectation is there for potential routing elements’ MISMATCH, also calculated on the basis of their degree and summed up per routing element; for DEPTH related to the nesting of structured blocks; for the CONTROL FLOW COMPLEXITY metric as the number of choices that can be made at splits in the model; and for CONNECTOR HETEROGENEITY as the degree to which routing elements of different types appear in a model.”	p. 53f.	CA	I	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S19	MENDLING ET AL. (2008)	investigation of influence factors on process model understandability	laboratory experiment (n=42, p. 147)	1. personal factors 2. model factors 3. content-related factors	1. correctly answering questions about the model content per person (PSCORE) / per model (MSCORE) 2. time needed to answer questions

code	statements / findings	ref.	meth.	LoE	TR
F3	"Before conducting the statistical analysis we make hypothetical connections between the different variables explicit. In particular, we identify hypotheses related to personal factors, model factors, and content factors: **P1** A higher PSCORE of participants should be connected with higher values in THEORY, DURATION, INTENSITY, and TIME. **M1** A higher MSCORE of models should be associated with lower values in SIZE, DIAMETER, TOKEN SPLIT, and HETEROGENEITY since these metrics might indicate that the model is easier to comprehend. ***M2 A higher MSCORE of models should be connected with higher values in STRUCTUREDNESS, SEPARABILITY, and SOUND since these metrics might be associated with models that are easier to comprehend.*** **C1** A higher sum of CORRECTANSWER should be connected with abstract labels (value of 0 in TEXT), basically our questions refer to structural properties of the model. **C2** A CORRECTANSWER (value of 1) should be connected with a lower value in TEXTLENGTH, since it becomes harder to match the elements mentioned in the question with the elements in the graphical model."	p. 148	CA	I	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S22	VANDERFEESTEN ET AL. (2008)	introduction and investigation of the significance of the cross-connectivity metric for process understandability	design-oriented, empirical evaluation using the SAP reference model and survey data (n=73, p. 489)	model characteristics influencing "cross-connectivity"	correctly answering questions about the model content (SCORE)

code	statements / findings	ref.	meth.	LoE	TR
F3	"***Intuitively, one may expect that a block-structure will positively affect model comprehension.***"	p. 486	CA	I	no

Level of Evidence IIa:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S30	LASSEN ET AL. (2009)	introduction and investigation of three process model complexity metrics	design-oriented, introduction of metrics and comparison of metrics using a application study (survey with 262 complex models (p. 621))	1. extended Cardoso metric (ECaM) 2. extended cyclomatic metric (ECyM) 3. new structuredness metric (SM)	perceived ease of understanding (n. e.)

code	statements / findings	ref.	meth.	LoE	TR
F5	"Metrics such as the Cyclomatic metric only focus on the resulting behavior and ignore the complexity of the model itself. There may be two different models that have the same state space where one model is compact and simple while the other one is large and difficult. ***The addition of an implicit place (i.e., a place that does not affect the behavior) may make the net more complex because it becomes bigger. However, in some cases, such a place can also make the net simpler because of symmetry reasons.***"	p. 614	CA	IIa	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S58	SCHALLES ET AL. (2011)	investigation of factors influencing the usability of modeling languages with a focus on model interpretation	survey and testing of potentially causal relationships using structure equation modeling (n=57, p. 791)	1. visual properties of the modeling language 2. language complexity	1. learnability 2. memorability 3. effectiveness 4. perceptibility 5. efficiency 6. user satisfaction

code	statements / findings	ref.	meth.	LoE	TR
F1	"In general, graphical modelling languages aim to support the expression of relevant aspects of real world domains such as business processes or application system structures [1]. ***For accurate human interpretation it is important that a model reproduces the knowledge contained in a clearly arranged and well-structured manner.***"	p. 787	CA	IIa	no

Level of Evidence IIb:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S11	GRUHN ET AL. (2006)	investigation of software complexity metrics and their adoption in business process modeling	conceptual analysis and discussion	1. factors influencing the control flow complexity of process models 2. complexity metrics	n. e.

code	statements / findings	ref.	meth.	LoE	TR
F1	“The easiest complexity measurement for software is the ”lines of code” (LOC) count which represents the program size. While for assembler programs a line of code is usually the same as an instruction statement, for programs written in a modern programming language, the LOC count usually refers to the number of executable statements (ignoring comments, line breaks etc.) [9]. For BPMs, the number of activities in the model can be regarded as an equivalent to the number of executable statements in a piece of software. For this reason, the ”number of activities” is a simple, easy to understand measure for the size of a BPM. ***However, the ”number of activities” metric does not take into account the structure of the model: A BPM with 50 activities may be written using a well-structured control flow which is easy to understand or in an unstructured way which makes understanding very hard.***“	p. 3	CA	IIb	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S30	LASSEN ET AL. (2009)	introduction and investigation of three process model complexity metrics	design-oriented, introduction of metrics and comparison of metrics using a application study (survey with 262 complex models (p. 621))	1. extended Cardoso metric (ECaM) 2. extended cyclomatic metric (ECyM) 3. new structuredness metric (SM)	perceived ease of understanding (n. e.)

code	statements / findings	ref.	meth.	LoE	TR
F11	“Much empirical work has been done by Mendling et al. [31,30], to learn what makes a model understandable. They operationalize understandability by introducing three categories of factors that they feel are important in understanding a model: personal (beyond psychological and intellectual); structural (model characteristics); and textual (description in the model). Besides characterizing understandability they do a web survey to test a number of hypothesis on the three categories of understandability. Among their findings they saw that higher knowledge of theory of concurrency and daily work with models lead to better understanding of models. ***Also, that the larger the score the participants of the web survey got wrt. a particular model was positively correlated with the structuredness and soundness of the model, regardless of their prior knowledge of the theory of concurrency. Their experiments show that there is a connection between the degree of structuredness in a process model and the understandability of it, and thereby also to lower complexity of the process model.***“	p. 624	CA	IIb	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S38	MENDLING ET AL. (2010)	investigation of existing research on the relationship of model structure on the one hand and error probability and understanding on the other hand	literature review and synthesis of research results into modeling guidelines, survey of experts concerning a ranking of the guidelines concerning their importance	model characteristics concerning structure and label style	n. e., “degree to which a process model can be easily understood” (p. 130)

code	statements / findings	ref.	meth.	LoE	TR
F9	“***G4: Model as structured as possible.*** A process model is structured if every split connector matches a respective join connector of the same type. Structured models can be seen as formulas with balanced brackets, i.e., every opening bracket has a corresponding closing bracket of the same type. ***Unstructured models are not only more likely to include errors [44], people also tend to understand them less easily [31].***”	p. 130	CA	IIb	yes

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S47	LA ROSA ET AL. (2011)	introduction and investigation of patterns for visual process models in order to decrease model complexity based on the “abstract syntax”	design-oriented, review of literature, prototypes and languages, conceptual analysis, introduction of patterns, tool and language survey (n=11) and usability evaluation survey (n=9, p. 625)	complexity reduction mechanisms (introduced modeling patterns) concerning abstract syntax	the usability evaluation: 1. perceived usefulness 2. perceived ease of use

code	statements / findings	ref.	meth.	LoE	TR
F4	“**Pattern 1 (Block-Structuring)**: Description: This pattern refers to methods to structure a process model in blocks. In a block-structured process model, each split element has a corresponding join element of the same type, and split-join pairs are properly nested [74].	p. 616	CA	IIb	yes

	Purpose: To improve understandability and maintenance through a simpler process model structure. […] ***Metrics: Increases structuredness of a process model.*** ***Rationale: Structured models are easier to understand [80], [81] and less error-prone [76], [69] than unstructured models.*** Realization: The problem of structuring process models has been extensively analyzed in the literature both from an empirical and from a theoretical point of view. Lau and Mendling [69] report the results of a study showing that structured models are less error-prone than unstructured equivalent models. Mendling et al. [81] propose seven guidelines to model easily-understandable process models. ***One of these guidelines is to model processes as structured as possible, which was ranked by a pool of practitioners as the guideline with the highest relative potential for improving process model understandability.*** Kiepuszewski et al. [56] provide a first attempt to classifying unstructured process models that can be transformed to structured equivalents, and show that structured models are less expressive than unstructured ones, thus unstructured model fragments cannot always be replaced with structured fragments that are behavior-equivalent. […] Finally, Weber et al. [117] propose a set of refactoring mechanisms for process models wherein they devise (but do not operationalize) a mechanism to replace a process fragment with a trace equivalent fragment having simpler structure.“				
F5	“**Pattern 2 (Duplication)**: Description: Duplication (aka Cloning) introduces controlled redundancy in a process model by repeating model elements. Two model elements are duplicated if they point to the same conceptual definition. ***Purpose: To improve understandability and maintenance through a simpler process model structure. Often required to block-structure an unstructured process model.*** […] Metrics: Despite increasing model size, this pattern typically also increases structuredness. Rationale: Less cluttered and more structured process models are easier to comprehend [80], [81] and less error-prone [76], [69]. Realization: Process modeling languages generally provide the possibility of creating duplicate model elements. […] In the literature, duplication is used to block-structure process models. For instance, the block-structuring approach in [90] uses unfolding techniques from Petri net theory to construct an occurrence net [37]. In an occurrence net, each XOR-join is unfolded by repeating the subsequent net. The result is a structured, but often much bigger model.”	p. 616f.	CA	IIb	yes

Nr.	**study**	**context**	**overall method of the study**	**treatment / independent variable (IV)**	**conceptualization of understanding / dependent variable (DV)**			
S53	REIJERS ET AL. (2011a)	investigation of the influence of syntax highlighting approaches on understandability of business process models	design-oriented, introduction of a concept of syntax highlighting for workflow nets (p. 342ff.) + laboratory experiment for evaluating the approach (n=103, p. 345)	usage of syntax highlighting	1. correctly answering questions about the model content (accuracy) 2. time needed to understand the model (understanding speed)			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F10	“It is arguable that the effect of highlighting on performance of both experts and novices might have been stronger if the models had been more complex. It is well known from prior research that more complex models are more difficult to understand [39, 42]. Several metrics have been proposed to measure different dimensions of complexity of a process model, e.g. in [1, 8, 9, 36, 39, 45, 48, 49, 64]. ***The models we used in the experiment are fairly structured such that a split operator most often has a direct join counterpart. Such structured models are rather easy to understand for experts. The highlighting effect might have been more effective also for experts if the models had been less structured.*** The reader may recall that, indeed, the identification of matching operator pairs is also possible in unstructured nets. Additionally, it might be argued that models need to be much larger before highlighting starts to have a significant effect on experts' performance.“				p. 347	CA	IIb	no

Nr.	**study**	**context**	**overall method of the study**	**treatment / independent variable (IV)**	**conceptualization of understanding / dependent variable (DV)**			
S54	REIJERS ET AL. (2011b)	investigation of the factors influencing process model understandability	survey + replication (n=73 + n=8, p. 454f.)	1. personal factors (experience, education etc.) 2. model factors (size, structural properties etc.)	correctly answering questions about the model content (SCORE value)			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F11	“Model factors have been hypothesized to have notable effects on their understanding, see [17], [21] for the related discussions. ***In short, the higher a process model’s sequentiality, separability, or structuredness the easier it is to understand such a model; lower values have the opposite effect.*** Similarly, understandability of a process model will also increase by a lower number of nodes, arcs, tasks, and connectors – regardless of its kind – on the one hand, or lower values for its diameter, connectivity, density, token splits, average connector degree, maximum connector degree, mismatch, depth, control flow complexity, connector heterogeneity, and crossconnectivity on the other. Higher values of these model factors will have the opposite effect. This set of expectations can be summarized as hypothesis **H2**: The more complex the model is, the less it will be understood.”				p. 454	CA	IIb	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)			
S59	WEBER ET AL. (2011)	proposition and investigation of a catalogue of process model "smells" for identifying refactoring opportunities	design-oriented, exemplary application in two case studies (healthcare and automotive)	refactoring techniques	n. e.			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F2	"PMS2: Contrieved Complexity [.] It is often possible to express a piece of control-flow logic within a process model in different ways. However, one alternative may be more difficult to comprehend for humans than another, despite their equivalence with respect to the (partial) execution traces they produce. Using the more complex alternative may negatively affect model understanding, and thus make maintenance of the model more difficult. […] ***Various studies have investigated the impact of structural model properties on model understandability. For example, [9] is centered around an adaptation of the cyclomatic number (one of the most widely used SE metrics) for business processes.*** Other research has analyzed process model understandability as aspect of maintainability, and has identified several correlations [8,1]. Further metrics take their motivation from cognitive research [91] or are based on concepts of modularity [93,88]. Most notably, an extensive set of metrics has been validated as factor influencing both error probability [48] and understandability [42]. ***The various validations show that factors like structuredness of a process model (i.e., the proper nesting of its gateways) and its density (i.e., the number of connections between its model elements) are influential. Both aspects can be manipulated by restructuring a process model; e.g., [91] presents three different, but trace-equivalent process models displaying different degrees of connectivity between model elements.*** Similarly, [75] proposes a metric for structural appropriateness, which can be used to determine how different models compare in their ability to capture a process in a compact and meaningful way. Relevant Refactoring. RF3 (Substitute Process Fragment)."				p. 472f.	CA	IIb	yes

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)			
S67	DUMAS ET AL. (2012)	exploration of the trade-off between size and structuredness of process model	survey and comparative analysis of process models with complexity metrics and laboratory experiment (n=110, p. 37ff.)	structuredness of process models, measured with complexity metrics	1. correctly answering questions about the model content 2. perceived complexity			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F2	"Sometimes, hundreds or thousands of process models are created and maintained in order to document large information systems. ***Given that such model collections are consulted, validated and updated by a wide range of stakeholders with various levels of expertise, ensuring the understandability of process models is a key concern in such settings. In this respect, a central guideline for business process modeling is to use structured building blocks as much as possible [19].***"				p. 31	CA	IIb	yes
F6	"***Another study confirms the significance of structuredness, albeit that different definitions are used [13]. These and other experiments are summarized in the seven process modeling guidelines [19]. Specifically, one of these guidelines is to model processes as structured as possible, which ranked as the guideline with the highest relative potential for improving process model understandability.***"				p. 35	CA	IIb	yes

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)			
S70	MENDLING ET AL. (2012)	investigation size and complexity as influence on error probability and understanding	design-oriented, introduction of new error detection method, case study (survey of 429 process models, p. 1193) and refinement of modeling guidelines	model characteristics such as size and complexity expressed by adequate measures and according thresholds	n. e.			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F3	"***Several factors have been found to be relevant factors for process model understanding*** and error probability. They include model purpose, problem domain, modeling notation, and layout (Ware et al., 2002; Hahn and Kim, 1999; Agarwal et al., 1999; Recker and Dreiling, 2007; Reijers and Mendling, 2011). ***In this paper, we focus on those factors that refer to the structure of a process model.*** […] Cardoso reports upon the results of an experiment to correlate process measures with the perceived complexity of process models (Cardoso, 2006). A team of researchers which includes Canfora, Rolón, and García correlate understandability and maintainability with size, complexity, and coupling of a process model (Canfora et al., 2005; Rolón Aguilar et al., 2007). Further measures are defined based on cognitive considerations (Vanderfeesten et al., 2008) and concepts of modularity (Vanhatalo et al., 2007; van der Aalst and Lassen, 2008). A set of measures is validated; these measures are seen as predictors of error probability in Mendling et al. (2008). ***Other works demonstrate that size is an important model factor along with additional measures like structuredness (Mendling, 2008)***."				p. 1190	CA	IIb	no
F5	"General guidelines of process modeling such as SEQUAL (Krogstie et al., 2006) or the Guidelines of Modeling (Becker et al., 2000) have been available for some time. ***Recent work*** in this area has aimed to define guidelines in a more quantitative and operational way, as well as to ***base them on empirical evidence***. The seven process modeling guidelines are a result of these efforts. ***These guidelines formulate the following modeling directives (Mendling et al., 2010)***: **G1** Use as few elements in the model as possible. **G2** Minimize the routing paths per element.				p. 1194f.	CA	IIb	yes

	G3 Use one start and one end event. ***G4 Model as structured as possible.*** **G5** Avoid OR routing elements. **G6** Use verb-object activity labels. **G7** Decompose a model with more than 50 elements.”				

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S73	OTTENSOOSER ET AL. (2012)	experimental comparison of understandability of graphical and textual process descriptions	laboratory experiment (n=196, p. 600)	textual vs. graphical business process descriptions (written use cases vs. BPMN), order of presentation	1. recall 2. accuracy of answering questions about the model content

code	statements / findings	ref.	meth.	LoE	TR
F14	“***In other works it has been shown that whether the information in the model is well organized in terms of labeling (Mendling et al., 2010), secondary notation (Reijers et al., 2011), iconic symbol design (Siau and Tian, 2009; Moody, 2009), or structuredness (Laue and Mendling, 2010) has an important influence on understanding.***”	p. 604	CA	IIb	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S87	MENDLING (2013)	overview on how empirical research informs structural and textual quality assurance of business process models	literature review (“essential contributions”, p. 100) and conceptual analysis	structural and textual characteristics of business process models (p. 101)	1. correctly answering questions on model content 2. recall of model elements 3. problem-solving based on the model (p. 104f.)

code	statements / findings	ref.	meth.	LoE	TR
F5	“[…], several guidelines of the 7PMG could be refined in [42]. Table 1 provides an overview of the results showing, among others, that process models with more than 30 nodes should be decomposed.” […] ***G4.a Structuredness Model as structured as possible.***	p. 104	CA	IIb	yes

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S91	WEITLANER ET AL. (2013)	investigation of factors supporting intuitive understandability of process models	field experiment (n=43, p. 56) and survey (n=77, p. 63)	1. personal characteristics (previous knowledge, education etc.) 2. model characteristics (language: EPC, BPMN, UML)	correctly answering comprehension questions (order, repetition, concurrency)

code	statements / findings	ref.	meth.	LoE	TR
F6	“***In that regard, it was found so far that e.g. structured models are less error prone than unstructured ones [21], learning a specific modeling language is sufficient in order to be able to understand also other ones equally well [14], systematic BPM labeling practices could improve the models' comprehensibility [17], and the size or rather complexity of a model impacts its understandability as well [6].*** The second mentioned finding, however, seems to be contradictory to the discoveries of Mendling et al. [6] that the amount of theoretical modeling knowledge plays indeed a role in this particular context.“	p. 55	CA	IIb	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S100	MORENO-MONTES DE OCA ET AL. (2014)	assessment of the acceptance of process modeling guidelines through a survey	literature review to collect modeling guidelines, survey (n=40, p. 78)	guidelines regarding: 1. size 2. modularity and structuredness 3. complexity 4. layout and label style	n. e.

code	statements / findings	ref.	meth.	LoE	TR
F2	“Modularity and Structuredness: Modularity is achieved by using subprocesses [22]. This entails reducing the size of the model at the top level in the model hierarchy ***to improve understandability of the model***. There are various guidelines in the literature that guide the modeler in the number of items from which the modularity should be included in the business process models and criteria for subprocess discovery [23]. Since model size is a prerequisite to introduce modularization, guideline S1 is also related to modularity. ***The structuredness property on the other hand, has been discussed as a guideline to avoid errors, first in research on programming, and later also in business process modeling [24]. A business process model is structured if every split gateway matches a respective join gateway of the same type [8]. In this group we collected six guidelines.*** ***- M1: Model as structured as possible: every split gateway should match a respective join gateway of the same type.*** - M2: Avoid deeply nesting structured blocks. - M3: Avoid decompositions into subprocesses with less than 5-7 activities. - M4: Good candidates for subprocesses are fragments of a model that are components with a single input and a single output control flow arc. - M5: Good candidates for subprocesses are those fragments of a model of which the nodes are more strongly connected by arcs to each other than the nodes outside this collection. - M6: Avoid inclusion of many small process models.”	p. 77	CA	IIb	yes

Level of Evidence III:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S13	MENDLING ET AL. (2007)	investigation of the influence of personal and model characteristics on process model understandability	field experiment (n=73, p. 52) and expert interview (n=12, p. 60)	1. personal characteristics of model readers 2. model characteristics	1. perceived ease of understanding 2. correctly answering questions about the model (reg. order, concurrency, exclusiveness, repetition) 3. relative perceived understandability (ranking of models)

code	statements / findings	ref.	meth.	LoE	TR
F12	"Finally, ***experts*** indicated a decreasing relevance of (a) model-related factors, (b) person-related factors, and (c) domain knowledge for the understanding of process models. ***The model-related factors that were mentioned most as positively influencing model understandability: unambiguity (7 times), simplicity (4 times), structuredness (4 times) and modularity (4 times).*** From the less mentioned factors, the supposed positive effects of textual support is interesting to mention, i.e. well-chosen textual descriptions of model elements (3 times) and textual context information on the model in general (3 times). Part of the factors mentioned seem to overlap with the factors considered in this study (e.g. simplicity and structuredness), while others are food for further research (e.g. modularity and textual support)."	p. 60	EI	III	no

Level of Evidence IV:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S70	MENDLING ET AL. (2012)	investigation size and complexity as influence on error probability and understanding	design-oriented, introduction of new error detection method, case study (survey of 429 process models, p. 1193) and refinement of modeling guidelines	model characteristics such as size and complexity expressed by adequate measures and according thresholds	n. e.

code	statements / findings	ref.	meth.	LoE	TR
F9	"***Guideline G4 emphasizes the importance of structured modeling. This guideline is confirmed by the threshold of 0.79. Beyond this value, we observed an error probability of almost 10%. While structuredness has a recall of only 30%, it has by far the best precision of roughly 25% for the insurance sample. The overall accuracy of prediction is greater than 90%. The central importance of this measure is therefore confirmed by our study. In order to avoid problems with structuredness, it seems desirable to use well-formed design patterns (van der Aalst et al., 2003; Wohed et al., 2006). This observation is further emphasized by the connector mismatch measure. It has the second largest AUC value of about 87% and shows a good balance of precision and recall in the validation sample.***"	p. 1195	SU	IV	yes
F12	***G4.a Structuredness Model as structured as possible.***	p. 1195	SU	IV	yes

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S75	SÁNCHEZ-GONZÁLEZ ET AL. (2012)	definition and investigation of thresholds for gateway complexity measures	two field experiments (n=28 + n=23, p. 1163ff.)	different structural measures: 1. CFC (Control-Flow Complexity) 2. GM (Gateway Mismatch) 3. GH (Gateway Heterogeneity) 4. AGD (Average Gateway Degree) 5. MGD (Max. Gateway Degree) 6. TNG (Total Number of Gateways)	1. time needed to answer questions (understand. time) 2. number of correct answers related to understandability 3. ratio between Nr. of correct answers and time (efficiency) 4. perceived complexity of understandability exercise

code	statements / findings	ref.	meth.	LoE	TR
F4	"First of all, it is ***important to define the most suitable number of decision nodes***. Following the thresholds for the TNG measure, the gateway complexity is high when the model has more than 18 decision nodes, and very high with more than 22. For this reason, we establish the number of nodes as being between 18 and 22. But it is not only the number of decision nodes that increases the complexity of the model; it is also the diversity of their types (XOR, OR and AND). Following the CFC measure, OR-split nodes create more mental states, a total of $2^n - 1$, which means that the focus of reducing gateway complexity should be in this type of decision nodes, while AND nodes imply a lower increase of complexity for models. Since heterogeneity of decision nodes is an important point in the evaluation of complexity, the thresholds for the GH measure indicate to us that more than 10 XOR decision nodes, 7 AND nodes or 4 OR nodes endanger the quality of the model. Input/output sequence flows from decision nodes are another key aspect in gateway complexity. Specifically, more than 7 input/output sequence flows increase the complexity of the model and more than 9 is not acceptable, due to the fact that the modeler would take into account a very ''difficult'' number of mental states. ***Finally, an important aspect in a good design is about the number of output/input in split/join nodes. A good design has the same output sequence flows for splits and input sequence flows for joins. To be precise, if that difference is higher than 15, the complexity could increase too much – higher than 20 is not appropriate.*** All of this information can be summarized in the following set of rules for business process modeling: - Include no more than 18–22 decision nodes. - Minimize the number of OR split nodes.	p. 1169	LE	IV	yes

	- Include no more than 10 XOR, 7 AND and 4 OR decision nodes. - Each decision node should have fewer than 7–9 input/output sequence flows. - ***A difference higher than 15–20 in the number of input/output sequence flows between split/join nodes is not acceptable.***"				

3.5.2 Aggregated Statement "*AS_M_14*"

Level of Evidence I:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S11	GRUHN ET AL. (2006)	investigation of software complexity metrics and their adoption in business process modeling	conceptual analysis and discussion	1. factors influencing the control flow complexity of process models 2. complexity metrics	n. e.

code	statements / findings	ref.	meth.	LoE	TR
F4	"***In general, using not well-structured models […] can be regarded as bad modeling style which makes understanding of the model more complicate.***"	p. 7	CA	I	no

Level of Evidence IIa:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S9	CARDOSO ET AL. (2006)	description of the scientific discourse on process model complexity	literature survey of complexity metrics and adaption to process models, report on an experiment (n=19), detailed method description is missing	process model complexity measure	n. e.

code	statements / findings	ref.	meth.	LoE	TR
F2	"On the other hand, we also have to consider that some languages allow the construction of processes that are not well-structured. As we have already mentioned, examples of such languages include EPC and Workflow nets. ***In these modeling languages, splits do not have to match a corresponding join. These processes are generally more difficult to understand and result often in design errors.***"	p. 119	CA	IIa	no

Level of Evidence IIb:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S11	GRUHN ET AL. (2006)	investigation of software complexity metrics and their adoption in business process modeling	conceptual analysis and discussion	1. factors influencing the control flow complexity of process models 2. complexity metrics	n. e.

code	statements / findings	ref.	meth.	LoE	TR
F1	"The easiest complexity measurement for software is the "lines of code" (LOC) count which represents the program size. While for assembler programs a line of code is the same as an instruction statement, for programs written in a modern programming language, the LOC count usually refers to the number of executable statements (ignoring comments, line breaks etc.) [9]. For BPMs, the number of activities in the model can be regarded as an equivalent to the number of executable statements in a piece of software. For this reason, the "number of activities" is a simple, easy to understand measure for the size of a BPM. ***However, the "number of activities" metric does not take into account the structure of the model: A BPM with 50 activities may be written using a well-structured control flow which is easy to understand or in an unstructured way which makes understanding very hard.***"	p. 3	CA	IIb	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S38	Mendling et al. (2010)	investigation of existing research on the relationship of model structure on the one hand and error probability and understanding on the other hand	literature review and synthesis of research results into modeling guidelines, survey of experts concerning a ranking of the guidelines concerning their importance	model characteristics concerning structure and label style	n. e., "degree to which a process model can be easily understood" (p. 130)

code	statements / findings	ref.	meth.	LoE	TR
F9	"***G4: Model as structured as possible***. A process model is structured if every split connector matches a respective join connector of the same type. Structured models can be seen as formulas with balanced brackets, i.e., every opening bracket has a corresponding closing bracket of the same type. ***Unstructured models are not only more likely to include errors [44], people also tend to understand them less easily [31].***"	p. 130	CA	IIb	yes

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S54	Reijers et al. (2011b)	investigation of the factors influencing process model understandability	survey + replication (n=73 + n=8, p. 454f.)	1. personal factors (experience, education etc.) 2. model factors (size, structural properties etc.)	correctly answering questions about the model content (SCORE value)

code	statements / findings	ref.	meth.	LoE	TR
F11	"Model factors have been hypothesized to have notable effects on their understanding, see [17], [21] for the related discussions. ***In short, the higher a process model's sequentiality, separability, or structuredness the easier it is to understand such a model; lower values have the opposite effect.*** Similarly, understandability of a process model will also increase by a lower number of nodes, arcs, tasks, and connectors – regardless of its kind – on the one hand, or lower values for its diameter, connectivity, density, token splits, average connector degree, maximum connector degree, mismatch, depth, control flow complexity, connector heterogeneity, and crossconnectivity on the other. Higher values of these model factors will have the opposite effect. This set of expectations can be summarized as hypothesis **H2**: The more complex the model is, the less it will be understood."	p. 454	CA	IIb	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S70	Mendling et al. (2012)	investigation size and complexity as influence on error probability and understanding	design-oriented, introduction of new error detection method, case study (survey of 429 process models, p. 1193) and refinement of modeling guidelines	model characteristics such as size and complexity expressed by adequate measures and according thresholds	n. e.

code	statements / findings	ref.	meth.	LoE	TR
F1	"The example of this process model also shows that a combination of different connectors can easily result in errors. The model cannot always terminate properly. Whenever the OR-split activates both branches, the AND-join can synchronize them and forward control towards a good completion. In any other case, the execution gets stuck at the AND-join, because control from one of the two incoming branches, which would bring the model to completion, is missing. Such an error is called a deadlock. It has been found that many process models in practice include such errors, and that often about 20% of the models have deadlocks or other behavioral problems (Mendling, 2009). ***Clearly, such deadlocks point to bad design. If a business process model is used for communication purposes and requirement analysis, a deadlock might lead to confusion in the stakeholders consulting this model.***"	p. 1189	CA	IIb	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S87	Mendling (2013)	overview on how empirical research informs structural and textual quality assurance of business process models	literature review ("essential contributions", p. 100) and conceptual analysis	structural and textual characteristics of business process models (p. 101)	1. correctly answering questions on model content 2. recall of model elements 3. problem-solving based on the model (p. 104f.)

code	statements / findings	ref.	meth.	LoE	TR
F5	"[…], several guidelines of the 7PMG could be refined in [42]. Table 1 provides an overview of the results showing, among others, that process models with more than 30 nodes should be decomposed." […]***G4.a Structuredness Model as structured as possible.***	p. 104	CA	IIb	yes

Level of Evidence III:

Level of Evidence IV:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S40	SÁNCHEZ-GONZÁLEZ ET AL. (2010a)	introduction and investigation of structural metrics for process models (BPMN)	design-oriented, evaluation of introduced complexity measures by means of six experiments (n_1=22, n_2=40, n_3=9, n_4=29, n_5=15, n_6=12, p. 82)	model characteristics influencing structural complexity (13 structural complexity measures, p. 81)	1. time needed to solve the understandability tasks (time) 2. number of correct answers (accuracy) 3. ratio between nr. of correct answers and time (efficiency)

code	statements / findings	ref.	meth.	LoE	TR
F2	"***Understanding time is strongly correlated with most of the probability error measures*** (number of nodes, diameter, density, average gateway degree, depth, ***gateway mismatch***, and gateway heterogeneity in all three experiments). There is no significant correlation with the connectivity coefficient, and the separability ratio was only correlated in the first experiment.	p. 83	LE	IV	no
F4	"***With regard to efficiency, we obtained evidence of the correlation of all the measures with the exception of separability.***"	p. 84	LE	IV	no
F5	"***The correlation analysis results indicate that there is a significant relationship between structural metrics and the time and efficiency of understandability.*** The results for correct answers are not as conclusive, since there is only a correlation of 3 of the 11 analyzed measures. In conclusion, measures with a significant correlation value (nºnodes, diameter, density, average gateway degree, maximum gateway degree, depth, ***gateway mismatch*** and gateway heterogeneity) can be traced back to particular BPMN elements, such as number of nodes (task, decision nodes, events, subprocesses, and data objects), decision nodes and sequence flow. We have therefore found evidence to reject the null hypothesis H0,1. The alternative hypothesis suggests that these BPMN elements affect the level of understandability of conceptual models in the following way: If there are more nodes, it is more difficult to understand models. If the path from a start node to the end is longer, it is more difficult to understand models. If there are more nodes connected to decision nodes, it is more difficult to understand models. If there is higher gateway heterogeneity, it is more difficult to understand models."	p. 84	LE	IV	no
F6	"We consider these p0 values to constitute different levels of understandability and modifiability, which is described as follows: Level 1: there is a 10% of probability of considering the model efficient Level 2: there is a 30% of probability of considering the model efficient Level 3: there is a 50% of probability of considering the model efficient Level 4: there is a 70% of probability of considering the model efficient The values described in Table 6 […] could be interpreted as follows: if number of nodes of a model is between 30 and 32, ***gateway mismatch is between 0 an 2, depth is 1, connectivity coefficient is 0,4 and sequentially is between 0,7 and 0,84 the probability of considering the model efficient in understandability tasks is about 70%, which means model has an acceptable level of quality***. It is interesting to note that many of the threshold values are rather close to each other. This is a good indication that the thresholds can be considered to be rather stable. […] The information contained in Table 6 can be interpreted as the following: ***if number of nodes is less or equal to 31, gateway mismatch is 1 or depth is 1, the model is considered as "very efficient" in understandability tasks***, while if gateway is 1, density 0 or sequentiality is 0,86, the model is considered as "very efficient" in modifiability tasks. In the same way, if a model has more than 65 nodes, gateway mismatch is more than 29 or CFCxor is more than 30, the model is considered as very inefficient in understandability tasks and if gateway mismatch is about 46 or density is 0,6, the models is considered as very inefficient in modifiability tasks."	p. 90f.	LE	IV	no
F7	***GatewayMismatch:*** ***29 (1: very inefficient understandability);*** ***16 (2: rather inefficient understandability);*** ***6 (3: rather efficient understandability);*** ***1 (4: very efficient understandability)***	p. 91	LE	IV	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S41	SÁNCHEZ-GONZÁLEZ ET AL. (2010b)	investigation and validation of structural metrics for business process models	analysis of experimental data from six experiments (p. 460)	model characteristics influencing structural complexity (13 structural complexity measures, p. 459f.)	1. time needed to solve the understandability tasks (time) 2. number of correct answers (accuracy) 3. ratio between nr. of correct answers and time (efficiency)

code	statements / findings	ref.	meth.	LoE	TR
F1	"***Understanding time is strongly correlated with number of nodes, diameter, density, average gateway degree, depth, gateway mismatch***, and gateway heterogeneity in all three experiments. There is no significant correlation with the connectivity coefficient, and the separability ratio was only correlated in the first experiment."	p. 460	LE	IV	no
F3	"***With regard to efficiency, we obtained evidence of the correlation of all the measures with the exception of separability.***"	p. 460	LE	IV	no
F5	"The statistical analyses suggest rejecting the null hypotheses, since the structural metrics apparently seem to be closely connected with understandability and modifiability. ***For understandability these include*** Number of Nodes, ***Gateway Mismatch***, Depth, Coefficient of Connectivity and Sequentiality."	p. 462	LE	IV	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S70	MENDLING ET AL. (2012)	investigation size and complexity as influence on error probability and understanding	design-oriented, introduction of new error detection method, case study (survey of 429 process models, p. 1193) and refinement of modeling guidelines	model characteristics such as size and complexity expressed by adequate measures and according thresholds	n. e.

code	statements / findings	ref.	meth.	LoE	TR
F9	"***Guideline G4 emphasizes the importance of structured modeling. This guideline is confirmed by the threshold of 0.79. Beyond this value, we observed an error probability of almost 10%. While structuredness has a recall of only 30%, it has by far the best precision of roughly 25% for the insurance sample. The overall accuracy of prediction is greater than 90%. The central importance of this measure is therefore confirmed by our study. In order to avoid problems with structuredness, it seems desirable to use well-formed design patterns (van der Aalst et al., 2003; Wohed et al., 2006). This observation is further emphasized by the connector mismatch measure. It has the second largest AUC value of about 87% and shows a good balance of precision and recall in the validation sample.***"	p. 1195	SU	IV	yes
F12	***G4.a Structuredness Model as structured as possible.***	p. 1195	SU	IV	yes

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)
S75	SÁNCHEZ-GONZÁLEZ ET AL. (2012)	definition and investigation of thresholds for gateway complexity measures	two field experiments (n=28 + n=23, p. 1163ff.)	different structural measures: 1. CFC (Control-Flow Complexity) 2. GM (Gateway Mismatch) 3. GH (Gateway Heterogeneity) 4. AGD (Average Gateway Degree) 5. MGD (Max. Gateway Degree) 6. TNG (Total Number of Gateways)	1. time needed to answer questions (understand. time) 2. number of correct answers related to understandability 3. ratio between Nr. of correct answers and time (efficiency) 4. perceived complexity of understandability exercise

code	statements / findings	ref.	meth.	LoE	TR
F3	"All the correlation results were significant and Spearman rho's values are the following: ***Understandability efficiency*** and measures CFC ["Control Flow Complexity"], ***GM ["Gateway Mismatch"]***, GH ["Gateway Heterogeneity"], AGD ["Average Gateway Degree"], MGD ["Maximum Gateway Degree"] and TNG ["Total Number of Gateways"] have correlation values of (-0.460, -0.452, -0.358, -0.423, -0.447 and -0.458). […] ***Results show that there is an inverse relationship between measures and understandability*** [.] efficiency, which means that the higher the measure values are, the lower the efficiency is."	p. 1165	LE	IV	no
F4	"First of all, it is important to define the most suitable number of decision nodes. Following the thresholds for the TNG measure, the gateway complexity is high when the model has more than 18 decision nodes, and very high with more than 22. For this reason, we establish the number of nodes as being between 18 and 22. But it is not only the number of decision nodes that increases the complexity of the model; it is also the diversity of their types (XOR, OR and AND). […] ***Finally, an important aspect in a good design is about the number of output/input in split/join nodes. A good design has the same output sequence flows for splits and input sequence flows for joins. To be precise, if that difference is higher than 15, the complexity could increase too much – higher than 20 is not appropriate.*** All of this information can be summarized in the following set of rules for business process modeling: - Include no more than 18–22 decision nodes. - Minimize the number of OR split nodes. - Include no more than 10 XOR, 7 AND and 4 OR decision nodes. - Each decision node should have fewer than 7–9 input/output sequence flows. - ***A difference higher than 15–20 in the number of input/output sequence flows between split/join nodes is not acceptable.***"	p. 1169	LE	IV	yes

3.5.3 Aggregated Statement "*AS_M_60*"

Level of Evidence I:

Level of Evidence IIa:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)			
S1	KIEPUSZEWSKI ET AL. (2000)	improvement of workflow models by means of structured modeling	conceptual and argumentative analysis	structure of workflow definitions	not explicated in detail (n. e.)			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F2	"***An alternative technique to transform arbitrary models into a structured form requires node duplication.*** As has been proved earlier, it cannot be used for every model, but even when it can be used, it is not without associated problems. Consider once again the model in figure 3. If activity D in the left model is followed by a large workflow specification, the transformation presented in the right model would need to duplicate the whole workflow specification following activity D. ***The resulting workflow will be almost twice as big as the original and will therefore be more difficult to comprehend.***"				p. 443	CA	IIa	no

Level of Evidence IIb:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)			
S47	LA ROSA ET AL. (2011)	introduction and investigation of patterns for visual process models in order to decrease model complexity based on the "abstract syntax"	design-oriented, review of literature, prototypes and languages, conceptual analysis, introduction of patterns, tool and language survey (n=11) and usability evaluation survey (n=9, p. 625)	complexity reduction mechanisms (introduced modeling patterns) concerning abstract syntax	the usability evaluation: 1. perceived usefulness 2. perceived ease of use			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F5	"**Pattern 2 (Duplication)**: Description: ***Duplication (aka Cloning) introduces controlled redundancy in a process model by repeating model elements.*** Two model elements are duplicated if they point to the same conceptual definition. ***Purpose: To improve understandability and maintenance through a simpler process model structure. Often required to block-structure an unstructured process model.*** […] Metrics: Despite increasing model size, this pattern typically also increases structuredness. ***Rationale: Less cluttered and more structured process models are easier to comprehend [80], [81]*** and less error-prone [76], [69]. Realization: Process modeling languages generally provide the possibility of creating duplicate model elements. […] In the literature, duplication is used to block-structure process models. For instance, the block-structuring approach in [90] uses unfolding techniques from Petri net theory to construct an occurrence net [37]. In an occurrence net, each XOR-join is unfolded by repeating the subsequent net. The result is a structured, but often much bigger model."				p. 616f.	CA	IIb	yes

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)			
S84	GLOWALLA ET AL. (2013)	investigation and survey of approaches for process-driven data quality management (integration of data quality approaches and process modeling)	structured literature review (p. 435ff.)	complexity metrics (model-inherent factors)	"the degree of which information contained in a process model can be easily understood by the reader (Reijers and Mendling 2011, p. 3). A process model is understood if the reader is able to explain the model (Figl and Laue 2011, p. 453)" (p. 435).			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F5	"***Duplication and Compacting. […] Compacting bears the risk of increasing the model structure's complexity due to the need to reroute arcs within the model to remaining representative elements. Besides potential impacts on connectors and according metrics (e.g., separability, structuredness), the layout of the model tends to become more complex (e.g., due to crossing arcs). Consequently, the changes in structure and layout will have a negative impact on the sequence's understandability as an essential characteristic of process models.*** We use the term understandability instead of complexity since the changes in the layout go beyond the impact on the considered metrics. At the same time, applying the compacting pattern, the model size should be reduced (La Rosa et al. *2011b*). […] Since the number of nodes and arcs might increase or decrease, the derived metrics may increase or decrease as well (e.g., repository size, diameter, connectivity, density). Additionally, due to structural model changes, further metrics may increase or decrease (e.g., separability). The impact on the metrics due to the application of this pair of patterns shows two important issues. First, although duplication is applied to improve model structure, related metrics might be impaired and therefore need to be controlled to mitigate undesired effects. Second, the impact of duplication and compacting on complexity is not generally predictable."				p. 441f.	CA	IIb	no

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)			
S87	Mendling (2013)	overview on how empirical research informs structural and textual quality assurance of business process models	literature review ("essential contributions", p. 100) and conceptual analysis	structural and textual characteristics of business process models (p. 101)	1. correctly answering questions on model content 2. recall of model elements 3. problem-solving based on the model (p. 104f.)			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F7	"Insight into factors of process model comprehension provides a solid basis for optimizing its structure. […] The research reported in [50] presents a approach based on the identification of ordering relations which leads to a maximally structured model under fully concurrent bisimulation. Here, two cases have to be distinguished. ***There are process models for which making them structured comes at the price of increasing its size.*** […] This increase stems from the duplication of activities in unstructured paths. There are also cases where a process model can be structured without having to duplicate activities. In practice, making a model structured without duplication appears to be rather rare. ***An investigation with more than 500 models from practice has shown that structuring leads to an increase in size of about 50% on average [53]. It is also important to note that duplication might be more harmful than a usual increase in size.*** The user experiment reported in [53] points to a potential confusion by model readers who are asked about behavioural constraints that involve activities that are shown multiple times in the model. The problem of duplicating activities is a key challenge in this area. It is an open research question how the beneficial effects of structuring can be best balanced with the harmful introduction of duplicate activities."				p. 106	CA	IIb	no

Level of Evidence III / Level of Evidence IV:

3.5.4 Aggregated Statement "*AS_M_91*"

Level of Evidence I / Level of Evidence IIa:

Level of Evidence IIb:

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)			
S45	Genon et al. (2011)	investigation of the cognitive effectiveness of BPMN 2.0 from the perspective of the Physics of Notations framework	theoretical assessment based on the Physics of Notations framework, and in-depth discussion	language constructs of the BPMN 2.0	cognitive effectiveness – the speed, ease and accuracy with which a representation can be processed by the human mind (p. 378)			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F11	"***One of the major flaws of visual notations is their diagrammatic complexity,*** which is mainly due to their poor scaling capability [41]. This complexity is measured by the number of elements displayed on a diagram. The degree of complexity management varies according to the ability of a notation to represent information without overloading the human mind. ***The two main solutions to decrease diagrammatic complexity are modularisation and hierarchic structuring.***"				p. 388	CA	IIb	yes

Nr.	study	context	overall method of the study	treatment / independent variable (IV)	conceptualization of understanding / dependent variable (DV)			
S47	La Rosa et al. (2011)	introduction and investigation of patterns for visual process models in order to decrease model complexity based on the "abstract syntax"	design-oriented, review of literature, prototypes and languages, conceptual analysis, introduction of patterns, tool and language survey (n=11) and usability evaluation survey (n=9, p. 625)	complexity reduction mechanisms (introduced modeling patterns) concerning abstract syntax	the usability evaluation: 1. perceived usefulness 2. perceived ease of use			
code	**statements / findings**				**ref.**	**meth.**	**LoE**	**TR**
F4	"**Pattern 1 (Block-Structuring)**: Description: This pattern refers to methods to structure a process model in blocks. ***In a block-structured process model, each split element has a corresponding join element of the same type, and split-join pairs are properly nested [74].*** Purpose: ***To improve understandability and maintenance through a simpler process model structure.*** […] Metrics: Increases structuredness of a process model. Rationale: ***Structured models are easier to understand [80], [81] and less error-prone [76], [69] than unstructured models.***				p. 616	CA	IIb	yes

	Realization: The problem of structuring process models has been extensively analyzed in the literature both from an empirical and from a theoretical point of view. Lau and Mendling [69] report the results of a study showing that structured models are less error-prone than unstructured equivalent models. ***Mendling et al. [81] propose seven guidelines to model easily-understandable process models. One of these guidelines is to model processes as structured as possible, which was ranked by a pool of practitioners as the guideline with the highest relative potential for improving process model understandability.*** Kiepuszewski et al. [56] provide a first attempt to classifying unstructured process models that can be transformed to structured equivalents, and show that structured models are less expressive than unstructured ones, thus unstructured model fragments cannot always be replaced with structured fragments that are behavior-equivalent. […] Finally, Weber et al. [117] propose a set of refactoring mechanisms for process models wherein they devise (but do not operationalize) a mechanism to replace a process fragment with a trace equivalent fragment having simpler structure.“				
F5	"**Pattern 2 (Duplication)**: Description: ***Duplication (aka Cloning) introduces controlled redundancy in a process model by repeating model elements.*** Two model elements are duplicated if they point to the same conceptual definition. Purpose: ***To improve understandability and maintenance through a simpler process model structure. Often required to block-structure an unstructured process model.*** […] Metrics: Despite increasing model size, this pattern typically also increases structuredness. ***Rationale: Less cluttered and more structured process models are easier to comprehend [80], [81]*** and less error-prone [76], [69]. Realization: Process modeling languages generally provide the possibility of creating duplicate model elements. […] In the literature, duplication is used to block-structure process models. For instance, the block-structuring approach in [90] uses unfolding techniques from Petri net theory to construct an occurrence net [37]. In an occurrence net, each XOR-join is unfolded by repeating the subsequent net. The result is a structured, but often much bigger model."	p. 616f.	CA	IIb	yes

Level of Evidence III / Level of Evidence IV:

---

4 Discussion

The above given overview demonstrates comprehensive support for the relationship between the structuredness of business process models and their understandability. In the above sample of supporting sources there are indeed contributions focusing on different research goals but, nevertheless, providing interesting statements on the relationship of structuredness and understandability, even if they were not in the research focus. The bottom-up approach for the development of SF can – although it is a quite laborious method – significantly contribute to a comprehensive and transparent overview of existing knowledge concerning certain topics of interest.

Using this approach, detailed information supported on different levels of evidence can be presented. However, there should not be particular or fixed thresholds for the evaluation of "final statements" as the development and usage of SF is a continuous and never-ending research process. On the basis of given evidence information, we can certainly draw well-founded conclusions but should always be aware of the preliminary character of every research results especially in the context of our relatively young research discipline. However, the exemplary application of the SF approach illustrates the considerable potential of Stylized Facts for theory development in ISR and BI as one of the major goals of our community's research work.[18]

[18] Cf. BICHLER ET AL. (2016).

The following aspects and questions seem to be important when working with the presented approach and in the discussion of the approach's value for ISR and BI:

1. How detailed should the underlying information be documented to have a transparent and at the same time easy to overlook derivation process?
2. The discourse on the presented material is a vital aspect of the approach and its value for ISR and BI. How can the discourse be supported in a comfortable way and how can SF on any topic be documented and further developed?
3. Against the background of BI being a mostly design-oriented research discipline, which contribution can the developed SF make for the design of innovative artefacts?

Considering the results presented in this report, it can be stated that SF can make a significant contribution to design-oriented research by providing vital information and well-founded guidelines concerning the design of business process models.

5 Conclusion

In this report, we gave an overview of an on-going dissertation project which uses the concept of Stylized Facts in the context of business process model understandability. We presented the research procedure for developing SF, an overview of topical clusters for business process model understandability research and a comprehensive application example. The total amount of identified statements is currently transformed into SF. Then propositions of potential theoretical models will be developed describing the different classified domains. It is planned to complete this work soon and the results are expected to significantly contribute to the on-going research stream on process model understandability as well as the discussion on useful research methods and approaches for theory development in ISR and BI.

References

Bichler, M., Frank, U., Avison, D., Malaurent, J., Fettke, P., Hovorka, D., Krämer, J., Schnurr, D., Mueller, B., Suhl, L., and Thalheim, B. (2016): Theories in Business and Information Systems Engineering; Business & Information Systems Engineering, Vol. 58. 2016, I. 4; pp. 291-319.

Cardoso, J., Mendling, J., Neumann, G., and Reijers, H.A. (2006): A Discourse on Complexity of Process Models; in: *J. Eder and S. Dustdar* (Eds.): Business Process Management Workshops, LNCS 4103; Springer, Berlin; 2006, pp. 117-128.

Dumas, M., La Rosa, M., Mendling, J., Mäesalu, R., Reijers, H.A., and Semenenko, N. (2012): Understanding Business Process Models: The Costs and Benefits of Structuredness; in: *J. Ralyté, X. Franch, S. Brinkkemper and S. Wrycza* (Eds.): Advanced Information Systems Engineering (CAiSE 2012), LNCS 7328; Springer, Berlin; 2012, pp. 31-46.

Fettke, P., Houy, C., and Loos, P. (2010): On the Relevance of Design Knowledge for Design-Oriented Business and Information Systems Engineering; Business and Information Systems Engineering, Vol. 2. 2010, I. 6; pp. 347-358.

Genon, N., Heymans, P., and Amyot, D. (2011): Analysing the Cognitive Effectiveness of the BPMN 2.0 Visual Notation; in: *B. Malloy, S. Staab and M. van den Brand* (Eds.): Software Language Engineering, LNCS 6563; Springer, Berlin; 2011, pp. 377-396.

Glowalla, P., and Sunyaev, A. (2013): Process-Driven Data Qualtiy Management - Through Integration of Data Quality into Existing Process Models; Business & Information Systems Engineering, Vol. 5. 2013, I. 6; pp. 433-448.

Grover, V., and Lyytinen, K. (2015): New State of Play in Information Systems Research: The Push to the Edges; MIS Quarterly, Vol. 39. 2015, I. 2; pp. 271-296.

Gruhn, V., and Laue, R. (2006): Complexity Metrics for Business Process Models;" in: *9th International Conference on Business Information Systems (BIS 2006), W. Abramowicz and H.C. Mayr* (Eds.), Klagenfurt, Austria, 2006.

Heine, B.-O., Meyer, M., and Strangfeld, O. (2005): Stylised facts and the contribution of simulation to the economic analysis of budgeting; Journal of Artificial Societies and Social Simulation, Vol. 8. 2005, I. 4; pp. 34.

Heine, B.-O., Meyer, M., and Strangfeld, O. (2007): Das Konzept der stilisierten Fakten zur Messung und Bewertung wissenschaftlichen Fortschritts; Die Betriebswirtschaft (DBW), Vol. 67. 2007, I. 5; pp. 583-601.

Helfat, C.E. (2007): Stylized facts, empirical research and theory development in management; Strategic Organization, Vol. 5. 2007, I. 2; pp. 185-192.

Houy, C., Fettke, P., and Loos, P. (2009): Stilisierte Fakten der Ereignisgesteuerten Prozesskette - Anwendung einer Methode zur Theoriebildung in der Wirtschaftsinformatik; in: *M. Nüttgens, F.J. Rump, J. Mendling and N. Gehrke* (Eds.): EPK 2009. 8. Workshop der GI und Treffen ihres Arbeitkreises

"Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten (WI-EPK), Gesellschaft für Informatik, Bonn, Berlin, 2009, pp. 22-41.

Houy, C., Fettke, P., and Loos, P. (2011): Stilisierte Fakten in der gestaltungsorientierten Wirtschaftsinformatik - Allgemeine Potentiale und erste Erfahrungen;" in: *Proceedings of the 10. International Conference on Wirtschaftsinformatik, A. Bernstein and G. Schwabe* (Eds.), Zürich, Switzerland, 2011, pp. 1157-1166.

Houy, C., Fettke, P., and Loos, P. (2013): The concept of stylized facts as an instrument for cumulative IS research;" in: *Pre-ECIS-Workshop "Building up or Piling Up? The Literature Review in Information Systems Research", M. Tate, E. Furtmueller, J. Evermann and W. Bandara* (Eds.), Utrecht University, Utrecht, Netherlands, 2013.

Houy, C., Fettke, P., and Loos, P. (2014): On the Theoretical Foundations of Reserach into the Understandability of Business Process Models;" in: *Proceedings of the 22nd European Conference on Information Systems (ECIS 2014)*, Tel Aviv, Israel, 2014.

Houy, C., Fettke, P., and Loos, P. (2015): Stylized Facts as an Instrument for Literature Review and Cumulative Information Systems Research; Communications of the AIS (CAIS), Vol. 37. 2015, I. 1; pp. 225-256.

Kaldor, N. (1961): Capital accumulation and economic growth;" in: *The Theory of Capital, Proceedings of a Conference Held by the International Economic Association, F.A. Lutz and D.C. Hague* (Eds.), Macmillan, London, 1961.

Kiepuszewski, B., ter Hofstede, A.H.M., and Bussler, C.J. (2000): On Structured Workflow Modelling; in: *B. Wangler and L. Bergman* (Eds.): Advanced Information Systems Engineering (CAiSE 2000), LNCS 1789 Springer, Berlin; 2000, pp. 431-445.

La Rosa, M., Wohed, P., Mendling, J., Ter Hofstede, A.H.M., Reijers, H.A., and van der Aalst, W.M.P. (2011): Managing process model complexity via abstract syntax modifications; IEEE Transactions on Industrial Informatics, Vol. 7. 2011, I. 4; pp. 614-629.

Lassen, K.B., and van der Aalst, W.M.P. (2009): Complexity metrics for Workflow nets; Information and Software Technology, Vol. 51. 2009, I. 3; pp. 610-626.

Loos, P., Fettke, P., Weißenberger, B.E., Zelewski, S., Heinzl, A., Frank, U., and Iivari, J. (2011): What in fact is the role of stylized facts in fundamental research of business and information systems engineering?; Business and Information Systems Engineering, Vol. 3. 2011, I. 2; pp. 107-117.

Mendling, J. (2013): Managing Structural and Textual Quality of Business Process Models; in: *P. Cudre-Mauroux, P. Ceravolo and D. Gašević* (Eds.): Data-Driven Process Discovery and Analysis (SIMPDA 2012), LNBIP 162; Springer, Berlin; 2013, pp. 100-111.

Mendling, J., Reijers, H.A., and Cardoso, J. (2007): What makes process models understandable?; in: *G. Alonso, P. Dadam and M. Rosemann* (Eds.): Business Process Management, LNCS 4714; Springer, Berlin; 2007, pp. 48-63.

Mendling, J., Reijers, H.A., and van der Aalst, W.M.P. (2010): Seven Process Modeling Guidelines (7PMG); Information and Software Technology, Vol. 52. 2010, I. 2; pp. 127-136.

Mendling, J., Sánchez-González, L., García, F., and La Rosa, M. (2012): Thresholds for error probability measures of business process models; Journal of Systems and Software, Vol. 85. 2012, I. 5; pp. 1188-1197.

Mendling, J., and Strembeck, M. (2008): Influence factors of understanding business process models; in: *W. Abramowicz and D. Fensel* (Eds.): Business Information Systems, LNBIP 7; Springer, Berlin; 2008, pp. 142-153.

Moreno-Montes de Oca, I., Snoeck, M., and Casas-Cardoso, G. (2014): A Look into Business Process Modeling Guidelines through the Lens of the Technology Acceptance Model; in: *U. Frank, P. Loucopoulos, Ó. Pastor and I. Petrounias* (Eds.): The Practice of Enterprise Modeling. LNBIP 197; Springer, Berlin; 2014, pp. 73-86.

Ottensooser, A., Fekete, A., Reijers, H.A., Mendling, J., and Menictas, C. (2012): Making sense of business process descriptions: An experimental comparison of graphical and textual notations; Journal of Systems and Software, Vol. 85. 2012, I. 3; pp. 596-606

Reijers, H.A., Freytag, T., Mendling, J., and Eckleder, A. (2011a): Syntax highlighting in business process models; Decision Support Systems, Vol. 51. 2011a, I. 3; pp. 339-349.

Reijers, H.A., and Mendling, J. (2011b): A Study Into the Factors That Influence the Understandability of Business Process Models; IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, Vol. 41. 2011b, I. 3; pp. 449-462.

Reiter, M., Fettke, P., and Loos, P. (2013): A Contribution to Theory Building for the Successful Implementation of ERP and BPR – an Application of the Method of Stylized Facts; in: *R.H. Sprague* (Eds.): Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS-46), Maui, USA, 2013, pp. 4045-4054.

Sánchez-González, L., García, F., Mendling, J., and Ruiz, F. (2010a): Quality Assessment of Business Process Models Based on Thresholds; in: *R. Meersman, T. Dillon and P. Herrero* (Eds.): On the Move to Meaningful Internet Systems: OTM 2010, LNCS 6426; Springer, Berlin; 2010a, pp. 78-95.

Sánchez-González, L., García, F., Mendling, J., Ruiz, F., and Piattini, M. (2010b): Prediction of Business Process Model Quality Based on Structural Metrics; in: *J. Parsons, M. Saeki, P. Shoval, C. Woo and Y. Wand* (Eds.): Conceptual Modeling, ER 2010, LNCS 6412; Springer, Berlin; 2010b, pp. 458-463.

Sánchez-González, L., García, F., Ruiz, F., and Mendling, J. (2012): Quality indicators for business process models from a gateway complexity perspective; Information and Software Technology, Vol. 54. 2012, I. 11; pp. 1159-1174.

Schalles, C., Creagh, J., and Rebstock, M. (2011): Usability of Modelling Languages for Model Interpretation: An Empirical Research Report; in: *A. Bernstein and G. Schwabe* (Eds.): Proceedings of the 10. International Conference on Wirtschaftsinformatik, Zurich, Switzerland, 2011, pp. 787-796.

Strangfeld, O. (2012): Die Vorteilhaftigkeit relativer Anreizschemata unter Berücksichtigung der Kollusionsgefahr. Eine agentenbasierte Simulation; Gabler, Wiesbaden; 2012.

Vanderfeesten, I., Reijers, H.A., Mendling, J., Van Der Aalst, W.M.P., and Cardoso, J. (2008): On a quest for good process models: The cross-connectivity metric; in: *Z. Bellahsène and M. Léonard* (Eds.): Advanced Information Systems Engineering, LNCS 5074; Springer, Berlin; 2008, pp. 480-494.

Weber, B., Reichert, M., Mendling, J., and Reijers, H.A. (2011): Refactoring Large Process Model Repositories; Computers in Industry, Vol. 62. 2011, I. 5; pp. 467-486.

Weißenberger, B.E., and Löhr, B.W. (2007): Planung und Unternehmenserfolg: Stylized Facts aus der empirischen Controllingforschung im deutschsprachigen Raum von 1990 - 2007; Zeitschrift für Planung & Unternehmenssteuerung, Vol. 18. 2007, I. 4; pp. 335-363.

Weitlaner, D., Guettinger, A., and Kohlbacher, M. (2013): Intuitive Comprehensibility of Process Models; in: *H. Fischer and J. Schneeberger* (Eds.): S-BPM ONE - Running Processes, CCIS 360; Springer, Berlin; 2013, pp. 52-71.

The Publications of the Institute for Information Systems (IWi) at the German Research Center for Artificial Intelligence (DFKI) appear at irregular intervals:

Issue 202: Marius Schönberger, Thomas Kleinert, Tobias Dumont: Praxisleitfaden - Geschäfte richtig dokumentieren, November 2015

Issue 201: Patrick Lübbecke: Praxisleitfaden – Green IT für kleine und mittlere Unternehmen, November 2015

Issue 200: Constantin Houy, Tim Niesen, Philip Hake, Peter Fettke, Peter Loos: Automatisierte Identifikation und Analyse von Argumentationsstrukturen im Digital-Humanities-Projekt ARGUMENTUM – Ergänzende Materialien und Ergebnisse, Oktober 2015

Issue 199: Tom Thaler, Sharam Dadashnia, Andreas Sonntag, Peter Fettke, Peter Loos: The IWi Process Model Corpus, Oktober 2015

Issue 198: Constantin Houy, Tim Niesen, Johannes Frank, Peter Fettke, Peter Loos: Zur Verwendung von Theorien in der Wirtschaftsinformatik – Eine quantitative Literaturanalyse, Dezember 2014. Heft 198 wurde auch in englischer Sprache herausgegeben:

Constantin Houy, Tim Niesen, Johannes Frank, Peter Fettke, Peter Loos: On the Usage of Theories in the Field of *Wirtschaftsinformatik* – A Quantitative Literature Analysis, December 2014

Issue 197: Peter Fettke, Constantin Houy, Philipp Leupoldt, Peter Loos: Discourse-Orientation in Conceptual Model Quality Research - Foundations, Procedure Model and Applications, January 2014

Issue 196: Constantin Houy, Peter Fettke, Peter Loos: Understanding understandability of conceptual models. What are we actually talking about? − Supplement, Juni 2013

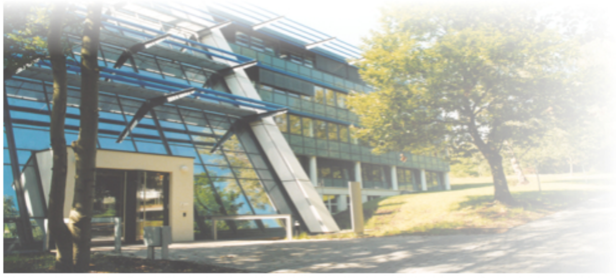
Issue 195: Constantin Houy, Markus Reiter, Peter Fettke, Peter Loos: Prozessorientierter Web-2.0-basierter integrierter Telekommunikationsservice (PROWIT) - Anforderungserhebung, Konzepte, Implementierung und Evaluation, Oktober 2012

Issue 194: Isabelle, Aubertin, Constantin Houy, Peter Fettke, Peter Loos: Stand der Lehrbuchliteratur zum Geschäftsprozessmanagement - Eine quantitative Analyse, Mai 2012

Issue 193: Silke Balzert, Thomas Kleinert, Peter Fettke, Peter Loos: Vorgehensmodelle im Geschäftsprozessmanagement - Operationalisierbarkeit von Methoden zur Prozesserhebung, November 2011

Issue 192: Constantin Houy, Peter Fettke, Peter Loos: Einsatzpotentiale von Enterprise-2.0-Anwendungen - Darstellung des State-of-the-Art auf Basis eines Literaturreviews, November 2010

Earlier Issues can be downloaded at: http://www.uni-saarland.de/lehrstuhl/loos/publikationen/iwi-hefte.html



Unter der wissenschaftlichen Leitung von Professor Dr. Peter Loos sind am Institut für Wirtschaftsinformatik (IWi) im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) mehr als 60 Mitarbeiter im Bereich der anwendungsnahen Forschung beschäftigt. Seit das Institut vor 30 Jahren durch Prof. Dr. Dr. h.c. mult. August-Wilhelm Scheer gegründet wurde, wird hier in Forschung und Lehre das Informations- und Prozessmanagement in Industrie, Dienstleistung und Verwaltung vorangetrieben. Ein besonderer Anspruch liegt dabei auf dem Technologietransfer von der Wissenschaft in die Praxis.

Die interdisziplinäre Struktur der Mitarbeiter und Forschungsprojekte fördert zusätzlich den Austausch von Spezialwissen aus unterschiedlichen Fachbereichen. Die Zusammenarbeit mit kleinen und mittelständischen Unternehmen (KMU) hat einen bedeutenden Einfluss auf die angewandte Forschungsarbeit - wie auch Projekte im Bildungs- und Wissensmanagement eine wichtige Rolle spielen. So werden in virtuellen Lernwelten traditionelle Lehrformen revolutioniert. Das Institut für Wirtschaftsinformatik berücksichtigt den steigenden Anteil an Dienstleistungen in der Wirtschaft durch die Unterstützung servicespezifischer Geschäftsprozesse mit innovativen Informationstechnologien und fortschrittlichen Organisationskonzepten. Zentrale Themen sind Service Engineering, Referenzmodelle für die öffentliche Verwaltung sowie die Vernetzung von Industrie, Dienstleistung und Verwaltung.

Am Standort im DFKI auf dem Campus der Universität des Saarlandes werden neben den Lehrtätigkeiten im Fach Wirtschaftsinformatik die Erforschung zukünftiger Bildungsformen durch neue Technologien wie Internet und Virtual Reality vorangetrieben. Hier führt das Institut Kooperationsprojekte mit nationalen und internationalen Partnern durch: Lernen und Lehren werden neu gestaltet; Medienkompetenz und lebenslanges Lernen werden Realität. Zudem beschäftigen sich die Mitarbeiterinnen und Mitarbeiter mit dem Einsatz moderner Informationstechniken in der Industrie. In Kooperation mit industrieorientierten Lehrstühlen der technischen Fakultäten saarländischer Hochschulen werden Forschungsprojekte durchgeführt. Hauptaufgabengebiete sind die Modellierung und Simulation industrieller Geschäftsprozesse, Workflow- und Groupware-Systeme sowie Konzepte für die virtuelle Fabrik.

Stuhlsatzenhausweg 3
D-66123 Saarbrücken
Tel.: +49 (0) 681 / 85775 - 3106
Fax: +49 (0) 681 / 85775 - 3696
iwi@iwi.uni-sb.de
www.iwi.uni-sb.de
www.dfki.de