Comprehension of Procedural Visual Business Process Models – A Literature Review

Kathrin Figl

Bus Inf Syst Eng 59(1) (2017)

Appendix (available online via http://link.springer.com)

Appendix A

Table 1 Statistical methods used in empirical studies

Type of Data	Type of Question	Number of Samples	Parametric (P)/ Non-Parametric Test (NP)	Statistical Test	Count	Percentages
	Description			Descriptive statistics only, no inferential statistics	11	8%
Discrete, categorical				Chi-Square test	1	1%
Continuous	Relationships			Regression analysis	22	15%
			P	Correlation analysis (Pearson's r)	11	8%
			NP	Correlation analysis (Spearman's rank correlation)	14	10%
	Differences	2 samples	P	Unpaired/paired Student's t-test	5	3%
			NP	Wilcoxon rank-sum test, Mann-Whitney U-test	10	7%
		> 2 samples	P	ANOVA	49	34%
			NP	Kruskal-Wallis test/Sheirer-Ray-Hare test (for more than one independent variables)	8	6%
			P	MANCOVA (Multivariate)	8	6%
				SEM (Structural equation modeling)	4	3%
			Information missing		4	
				Total	141	100%

Table 2 Categorization of sample sizes for all empirical studies

Quartile	n	Label
1	0-42	Small
2	43-70	Medium
3	71-113	Large
4	114-273	Very large

Table 3 Categorization of empirical studies' level of evidence of effects

Level of Evidence	Statistical Results
-	• Effect was not investigated or not reported. (E.g., in a variety of cases, subjective comprehension was measured as a control variable, but not all effects of independent variables on subjective comprehension were reported.)
No evidence (of effect)	• Effect was reported to be statistically not significant.
Conflicting	• Effect was reported to be both negative and positive in a study. (E.g., a factor led to lower comprehension in one model but higher comprehension in another.)
Weak	• Effect was reported based on descriptive statistics only, with no inferential statistics. • Effect was reported as a trend ($0.05 \leq p \leq 0.01$). • Effect was not calculated directly for measurement of comprehension accuracy but only for derived measures (e.g., efficiency, score time ratio).
Moderate	• Effect was significant in some but not all cases. For example: o "reached significance on a 0.05 level in at least two of five sub-samples" o significance for model 1 but not for model 2 o significance for experts but not for novices • Statistical significant test results were reported, but no numbers.
Strong	• Significance was reported in all cases (e.g., models). Numbers were reported.

Appendix B: Empirical Articles on Comprehension – Article-Based Overview

Table 4 Studies' characteristics and variables measured

Source Reference	# of Participants	Type of Participants	# of Models	Notation of Models	Measurement of Objective Comprehension	Measure-ment of Time	Measurement of Subjective Comprehension and User Preferences	Independent Variables [Category: Medium/Notation/Secondary Notation/Label/Model/Task/User; Research Design: Between/Within/Mixed]
(Aguilar et al. 2008)	110 (large)	students (5 samples, mostly master's and PhD students in information systems or software technology)	10	BPMN	six comprehension tasks – efficiency (ratio between the number of right answers and time) used as dependent variable	time taken	-	sixty defined measures – forty-six base measures that count significant elements of the model and fourteen derived measures (proportions between elements); nonsignificant measures are not mentioned in the results sections—only those that yielded significance in at least two of five sub-samples are mentioned as relevant influence factors: • number of end message events [Model, Within] • number of events of the model [Model, Within] • number of exclusive decision data-based [Model, Within] • number of intermediate events of the model [Model, Within] • number of intermediate message events [Model, Within] • number of sequence flows from events [Model, Within] • number of looping sequence flows [Model, Within]
(Bera 2012)	51 (medium)	students (undergraduate MIS, novice modelers)	2	BPMN	• nine comprehension tasks (yes/no) • three open-ended problem-solving tasks (transfer tasks) - coded by two coders	time taken	-	• with/without swim lanes [Secondary Notation, Between] • domain knowledge [User, Within] • modeling knowledge [User, Within]
(Döhring et al. 2014)	14 (small)	users familiar with process modeling: BPM researchers and practitioners	2	C-YAWL and vBPMN	comprehension tasks (yes/no) on order, concurrency, exclusiveness (8 for a simple travel-booking model, 12 for a complex municipality-name registration model); results analyzed in combination with modeling tasks	time taken	items on self-confidence and ease of answering (on a 5-point scale) for each task	• C-YAWL and vBPMN (process variant management notations with differing configuration and adaptation mechanisms; C-YAWL without, v-BPMN with modularization support, in an execution tool; also included modeling tasks) [Notation, Within] • simple and complex model [Model, Within] • professional level: senior: post-docs and industry employees, student-level: students up to PhD [User, Between]
(Dumas et al. 2012)	55 (medium)	students attending business process modeling courses	8	BPMN	six comprehension tasks per model	-	rating of model complexity on a 5-point scale	• structuredness (4 groups: structuredness (structured and unstructured) × cyclicity (cyclic, acyclic)) [Model, Mixed] • theoretical knowledge on process modeling [User, Between]
(Figl and Laue 2011)	199 (very large)	students (business)	4	BPMN-like	eight comprehension tasks per model (on concurrency, exclusiveness, order, repetition)	time taken (per model)	subjective cognitive load (7-point scale per comprehension task)	• item wording: concurrency, exclusiveness, order, repetition [Task→Model, Within] • element interactivity (PST-distance, cut vertex) per task [Task→Model, Within]
(Figl and Laue 2015)	156 (very large)	students (business)	4	BPMN	eight comprehension tasks per model (on concurrency, exclusiveness, order, repetition)	time taken (per model)	subjective cognitive load (7-point scale per comprehension task)	• modeling knowledge (low, high) [User, Between] • control-flow pattern (sequence, AND, XOR, loop, compound) [Task→Model, Within] • element interactivity: process structure tree distance [Task→Model, Within] • element interactivity: cut-vertex [Task→Model, Within] • validity of conclusion (valid, wrong) [Task, Within]
(Figl and Recker 2016)	120 (very large)	students (business, undergraduate)	1	BPMN	- [empirical article on subjective comprehension and user preferences]	-	preference for process model representation for comprehension tasks [scores from 0 to 100 in sliding scales]	• process model representation (text, structured text, diagram (BPMN)) [Notation, Within] • use of icons (with/without icons) [Notation, Within] • participants' cognitive style (verbal, spatial visual, object visual) [User, Between] • knowledge on conceptual modeling [User, Between]

Source Reference	# of Participants	Type of Participants	# of Models	Notation of Models	Measurement of Objective Comprehension	Measure-ment of Time	Measurement of Subjective Comprehension and User Preferences	Independent Variables [Category: Medium/Notation/Secondary Notation/Label/Model/Task/User; Research Design: Between/Within/Mixed]
(Figl et al. 2013a)	136 (very large)	students (business)	3	BPMN, UML AD, EPC, YAWL	comprehension tasks (yes/no) on order, concurrency, repetition, exclusiveness; model text comparison task	time taken (per model)	subjective cognitive load (7-point scale per comprehension task)	• modeling competencies: training on modeling basics, work experience, process modeling knowledge test [User, Between] • perceptual discriminability deficiencies [Notation, Between] • semiotic clarity deficiencies [Notation, Between]
(Figl et al. 2013b)	154 (very large)	students (information systems and business)	4	BPMN-like with differing routing symbols (BPMN, UML AD, and EPC)	eight comprehension tasks per model (on concurrency, exclusiveness, order, repetition)	time taken (per model)	subjective cognitive load (7-point scale per comprehension task), scale on perceived control flow comprehension (self-developed, 4 items on a 5-point Likert scale)	• symbol quality (aesthetics) [Notation, Between] • symbol quality (perceptual pop-out) [Notation, Between] • symbol quality (semantic transparency) [Notation, Between] • symbol quality (visual discriminability) [Notation, Between] • process modeling knowledge [User, Between]
(Figl and Strembeck 2015)	44 (medium)	students	2	BPMN	sixteen (2*8) true/false comprehension tasks	time taken	perceived ease of use	• flow direction (left-to-right, right-to-left, top-to-bottom, bottom-to-top) [Secondary Notation, Between] • abstract and concrete labels [Label, Between] • process modeling knowledge [User, Between]
(Heggset et al. 2015)	18 (small)	students and employees	2	BPMN	fifteen and ten questions	-	-	• models before and after revising syntactic quality according to a guideline [Model, Mixed]
(Hipp et al. 2014)	22 (small)	students, academics, industry participants	1	new visualizations	- [empirical article on subjective comprehension and user preferences]	-	• perceived comprehensibility • perceived comprehensibility of sequence flow • perceived clarity and overview 5-point scale	• new visualizations (bubble visualization concept, BPMN3D, network visualization concepts, thin line concept) [Notation, Within]
(Jeyaraj and Sauter 2014)	142 in study 1, 131 in study 2 (very large)	students (business)	1	similar to UML AD	identification of activities and internal/external actors	time taken	-	• swim lanes versus no swim lanes [Secondary Notation, Between]
(Johannsen et al. 2014a)	53 (medium)	students attending a process modeling course (bachelor level)	1 with 3 levels and 8-10 sub-processes	eEPC	open questions, fill-in-the-blank test	-	perceived ease of understanding with four items on a 7-point scale	• domain knowledge [User, Between] • decomposition: full compliance versus moderate violation (minimality, determinism, losslessness violated) versus strong violation (minimality, determinism, losslessness, minimum coupling, strong cohesion violated) [Secondary Notation, Between] • personal factors [User, Between]
(Jošt et al. 2016)	103 (large)	students (undergraduate, no prior training in modeling)	4	BPMN, UML AD, EPC	questions on the semantics of the process; no other details found	time taken	-	• BPMN, UML AD, EPC [Notation, Mixed] • complexity [Model, Mixed]
(Kock et al. 2008)	210 (very large)	students	1 (self-modeled)	flow diagram	- [empirical article on subjective comprehension and user preferences]	-	perceived ease of understanding (2 Likert-type scale items, 7-point scale)	• activity flow representation versus communication flow representation [Notation, Within]
(Kock et al. 2009)	78 (large)	employees of 18 organizations (participants in business process redesign groups)	1 (self-modeled)	communication flow diagrams versus functional flowcharts with swim lanes	- [empirical article on subjective comprehension and user preferences]	-	perceived ease of understanding (3 items, 7-point scale)	• modeling notation (high versus low communication flow orientation) [Notation, Within] • ease of generating the models [Task, Within]
(Koschmider et al. 2015b)	49 (medium)	process modeling beginners (graduates) and experts	2	no particular process modeling notation	- [empirical article on subjective comprehension and user preferences]	-	perceived ease of understanding on a 5-point scale	• naming conventions for labels (unrevised versus revised activity labels from a linguistic perspective) [Label, Mixed]
(Kummer et al. 2016)	127 (very large)	students (postgraduate, business)	2	BPMN	eight comprehension tasks per model	time taken	six-item Likert scale adopted from the ease of understanding measure	• culture (Germanic – Germany and Austria, Confucian – China) [User, Between] • theoretical knowledge on process modeling [User, Between]

Source Reference	# of Participants	Type of Participants	# of Models	Notation of Models	Measurement of Objective Comprehension	Measure-ment of Time	Measurement of Subjective Comprehension and User Preferences	Independent Variables [Category: Medium/Notation/Secondary Notation/Label/Model/Task/User; Research Design: Between/Within/Mixed]
							(Burton-Jones and Meso 2008)	• familiarity with BPMN [User, Between] • color (black/white versus bright colors with high contrast based on Asian preferences) [Secondary Notation, Between]
(Laue and Gadatsch 2011)	22 (small)	students (business)	1	EPC-like, without intermediary events	sixteen comprehension tasks on concurrency, nineteen on exclusiveness, thirty on order, six on forward dependency/response, and six on backward dependency/precedence	-	-	• question type (concurrency, exclusiveness and order) [Task→Model, Within] • question wording (style 1 and style 2 questions) [Task, Between]
(Melcher and Seese 2008)	18 (small)	students attending a workflow management course	1	BPMN-like (routing symbols labelled with XOR and AND)	comprehension tasks on order, repetition, concurrency, and exclusiveness	-	-	• comprehension tasks (group A: questions about order and repetition; group B: questions about concurrency and exclusiveness) [Task→Model, Between]
(Melcher et al. 2010)	178 (very large)	students attending workflow management courses	1	BPMN-like (routing symbols labelled with XOR and AND)	276 comprehension tasks on order, repetition, concurrency, exclusiveness distributed over nine groups	-	-	• control flow element (order, repetition, concurrency, exclusiveness) [Task→Model, Between]
(Mendling and Strembeck 2008)	42 (small)	German mailing lists EMISA and WI, as well as students attending courses on process modeling at the Vienna University of Economics and Business Administration	6	EPC-like, without events	six yes/no comprehension tasks per model	time taken	-	• abstract activity labels (e.g., A, B, C) versus illustrative, textual labels (e.g., “check credit limit”) [Label, Between] • diameter (length of the longest path from a start node to an end node in the process model) [Model, Within] • heterogeneity of gateways [Model, Within] • number of nodes [Model, Within] • separability [Model, Within] • soundness [Model, Within] • string length of all textual activity labels [Label, Within] • structuredness of the process graph [Model, Within] • token split [Model, Within] • theoretical knowledge on process modeling [User, Between] • intensity of work with process models (how often they work with process models) [User, Between] • duration of involvement with business process modeling [User, Between]
(Mendling et al. 2010c)	29 (small)	students (after attending a course on process modeling)	1	EPC	- [empirical article on subjective comprehension and user preferences]	-	perceived usefulness for each label on a 7-point scale	• labels of various styles (verb-object style, action-noun style, other styles) [Label, Within] • perceived ambiguity of labels [Label, Within]
(Mendling et al. 2012b)	113 (very large)	academics and practitioners with interest in modeling (EMISA and WI mailing list), graduate students (modeling courses, system analysis and design course)	6	EPC-like, without events (routing symbols labelled with XOR and AND)	six comprehension tasks per model	time taken (per question)	-	• label semantics (abstract versus concrete labels) [Label, Between] • modeling expertise [User, Between] • modeling intensity [User, Between] • formal theoretical process knowledge [User, Between] • paper versus computer [Medium, Between]
(Natschläger 2011)	22 (small)	post-graduate computer scientists	2*4	BPMN, deontic BPMN	thirty-four (2*17) comprehension tasks on exclusiveness and order of tasks	-	-	• modeling notation (BPMN, deontic BPMN) [Notation, Between]
(Ottensooser et al. 2012)	196 (very large)	• users trained in process models (students who had received explicit training in business process modeling and flow-chart notations	6	BPMN notation versus written use cases	six multiple-choice comprehension tasks answered three times (increase in domain understanding measured); for example “What happens if stakeholders change the project scope?”	-	-	• BPMN notation (graphical) vs. written use cases (textual) [Notation, Within] • comfort with flow charts [User, Between] • frequency of use of flow charts [User, Between]

Source Reference	**# of Participants**	**Type of Participants**	**# of Models**	**Notation of Models**	**Measurement of Objective Comprehension**	**Measure-ment of Time**	**Measurement of Subjective Comprehension and User Preferences**	**Independent Variables [Category: Medium/Notation/Secondary Notation/Label/Model/Task/User; Research Design: Between/Within/Mixed]**
		as business analyst proxies: o 129 (post-graduate industrial engineering) o 26 (post-graduate BPM and enterprise systems) • users not trained in process models (without training in flow-charts, 41 students as business-user proxies from varying courses)						
(Petrusel et al. 2016)	75 (large)	experienced modelers from industry and academia	16	BPMN	sixteen comprehension tasks (yes, no), eye-tracking	time taken	confidence rating (5-point scale)	• color highlighting of gateways for task-relevant region [Secondary Notation, Within] • task-specific layout [Secondary Notation, Within]
(Pichler et al. 2012)	27 (small)	students (Berlin, Innsbruck)	4	BPMN versus ConDec	thirty-two (4*2*4) tasks: • four sequential tasks (how input conditions lead to a certain outcome; e.g., “Activity X must be directly preceded by activity Y”) • four circumstantial tasks (e.g., what (combination of) circumstances will cause/lead to/follow from a particular outcome? “If activity X or Y has been executed, is it possible to terminate a process instance by executing at least one additional activity?”)	time taken	-	• sequential and circumstantial tasks [Task, Within] • modeling paradigms: declarative versus imperative models [Notation, Between]
(Recker and Dreiling 2007)	69 (medium)	students (postgraduate, information systems)	2	EPC versus BPMN	multiple-choice comprehension tasks (yes, no, undecided), problem-solving task, cloze test filling in missing words in a textual process description	time taken	perceived ease of use scale (Moore, Benbasat 1991)	• business domain knowledge [User, Between] • modeling competencies: level of EPC competency [User, Between] • modeling notation (EPC versus BPMN; participants were familiar with EPC but not with BPMN) [Notation, Between]
(Recker and Dreiling 2011)	68 (medium)	students (postgraduate, information systems)	2	EPC versus BPMN	• multiple-choice comprehension tasks (yes, no, unknown) • inferential problem-solving task (surface understanding; retention ability) • cloze test filling in missing words in a textual process description (deep understanding; transferability test) [without process models]	time taken	perceived ease of use scale (Moore, Benbasat 1991)	• modeling notation (EPC versus BPMN; participants were familiar with EPC but not with BPMN) [Notation, Between] • English as a second language (ESL) [User, Between] • modeling experience with EPC (number of models created or read) [User, Between] • working experience with BPM [User, Between]

Source Reference	**# of Participants**	**Type of Participants**	**# of Models**	**Notation of Models**	**Measurement of Objective Comprehension**	**Measurement of Time**	**Measurement of Subjective Comprehension and User Preferences**	**Independent Variables [Category: Medium/Notation/Secondary Notation/Label/Model/Task/User; Research Design: Between/Within/Mixed]**
(Recker et al. 2005)	16 (small)	students (postgraduate, information technology)	2	C-EPCs versus EPC	- [empirical article on subjective comprehension and user preferences]	-	perceived usefulness and perceived ease for various tasks	• configurable EPCs (C-EPCs) versus EPC [Notation, Within]
(Recker 2013)	98 (large)	students (information systems)	3	BPMN with/without gateway constructs	four comprehension tasks per model measure participants' understanding of four fundamental aspects of the control flow logic of the models presented, viz., reachability, deadlocks, liveness, and option to complete	-	perceived ease of understanding with four items (Maes and Poels 2007)	• number of arcs and nodes [Model, Within] • implicit/explicit representation of parallel split and simple merge scenarios (with/without BPMN gateway constructs, perceptual discriminability effect through use of gateways) [Notation, Between] • knowledge of control flow logic [User, Between] • self-reported knowledge of BPMN grammar [User, Between]
(Recker et al. 2014)	92 (large)	three groups: domain experts (35 staff members from a government agency), method experts (22 postgraduate students enrolled in a business process management course), control group (35 mixed, without high domain or method knowledge score)	2	unknown	• multiple-choice tasks (domain surface understanding) • comprehension test about the domain modeled • ability to understand the models (modularity, concurrency, exclusiveness, and repetition of the control flow logic)	-	-	• selection ability [User, Between] • abstraction ability [User, Between] • users' surface learning motive and strategy [User, Between] • users' deep learning motive and strategy [User, Between] • prior domain knowledge [User, Between] • method knowledge [User, Between] • self-efficacy [User, Between] • sensing versus intuitive learning style [User, Between] • paper versus computer [Medium, Between]
(Reijers and Mendling 2011)	73 (large)	students from three universities, trained in Petrinets, EPCs, or both	12 (6 versions of the questionnaire, 847 complete model evaluations)	EPC-like	seven multiple-choice comprehension tasks related to execution order, exclusiveness, concurrency and repeatability issues (yes/no/I don't know) and one open question to describe problems with this process (0-2 points)	-	perceived ease of understanding the model (one question)	• approximately twenty-four model-based metrics; only those correlations are reported that "displayed the direction of the influence that we hypothesized upon" [Model, Within]: o average gateway degree o gateway heterogeneity o cross-connectivity o density o gateway mismatch o number of OR joins • educational background (undergraduate versus graduate students with higher level of knowledge of workflow concepts) [User, Between] • modeling experience [User, Between] • theoretical knowledge of process modeling [User, Between]
(Reijers et al. 2011a)	70 (103 data sets, 33 excluded) (medium)	62 experienced modelers (from industry and academia) and 41 students (from a business process management course)	1	workflow net	closed comprehension tasks on concurrency, exclusiveness, order, repetition; adapted from (Mendling et al. 2007); response options: yes, no, I don't know	time taken	-	• syntax highlighting (evaluated overall and separately for novices and experts) [Secondary Notation, Between]
(Reijers et al. 2011b)	28 (small)	experienced consultants	2	workflow net modeled with Protos tool	twelve comprehension tasks per model	time taken	-	• modularization (model with subprocesses, flattened model without subprocesses) [Secondary Notation, Mixed]
(Rodrigues et al. 2015)	73 (large)	students and practitioners from various IT companies (experienced and inexperienced subjects)	1	BPMN, textual process description	ten multiple-choice questions	-	-	• BPMN, textual process description modeling notation [Notation, Between]
(Sánchez-González et al. 2010)	22+40+9=71 (large)	students (in three computer engineering/information systems courses, analyzed separately, 9-40 students each)	15	BPMN	unclear	time taken	-	• coefficient of connectivity [Model, Within] • depth [Model, Within] • diameter [Model, Within] • gateway heterogeneity [Model, Within] • gateway mismatch [Model, Within] • number of nodes [Model, Within]

Source Reference	# of Participants	Type of Participants	# of Models	Notation of Models	Measurement of Objective Comprehension	Measurement of Time	Measurement of Subjective Comprehension and User Preferences	Independent Variables [Category: Medium/Notation/Secondary Notation/Label/Model/Task/User; Research Design: Between/Within/Mixed]
								• sequentiality [Model, Within]
(Sánchez-González et al. 2012)	28 + 23 (medium)	students (three courses, pre- and postgraduates)	15	BPMN	"a set of comprehension tasks"	-	personal opinion on difficulty of completing comprehension tasks (1=very easy, 5=very difficult)	no factorial design, but ten models cover a wide range of these measures; six independent variables [Model, Within]: • average gateway degree: average number of both incoming and outgoing arcs of the gateway nodes • control-flow complexity (CFC): complexity of split gateways • gateway heterogeneity • gateway mismatch: sum of gateway pairs that do not match each other • maximum gateway degree: maximum number of incoming and outgoing arcs of a decision node • number of gateways
(Sandkuhl and Wiebring 2015)	113 (large)	employees of a company who may work with process models	1	flow diagrams, eEPC, UML AD, BPMN	seventeen true/false questions (identifying and counting elements in the process)	-	five Likert-scale items for perception of notation (e.g., illustration of roles and relationship to process tasks)	• flow diagrams, eEPC, UML AD, BPMN [Notation, Between]
(Sarshar and Loos 2005)	50 (medium)	students with business and economy backgrounds	1	EPC versus Petrinet	multiple-choice tasks	-	perceived ease of use of model and of control flow	• questions for AND, XOR, OR, and multilevel AND/XOR situations [Task→Model, Within] • EPC versus Petrinet notation [Notation, Between]
(Soffer et al. 2015)	54 (medium)	students (in information systems attending a course on enterprise resource planning (ERP) systems and business process design)	5	EPC-like, with routing symbols left blank and a textual process description	assigning rules: • identifying the specific case from the catalog for each gateway • providing a logical expression that specifies the behavior of the process; five true/false questions related to the possible process behavior	-	-	• provision of a catalog of routing possibilities or a catalog of workflow patterns [Task, Between]
(Stitzlein et al. 2013)	16 (small)	professionals (healthcare or engineering)	2	BPMN, health process notation (HPN)	twenty-eight (2*14) items (multiple-choice and open questions with short answers)	time taken	perceived ease of use, perceived usefulness	• BPMN, health process notation (HPN) [Notation, Mixed] • professional background - healthcare or engineering [User, Between]
(Turetken et al. 2016)	60 (medium)	company representatives (majority with engineering university degree, 26 domain experts)	2	BPMN	nine comprehension tasks per model: • three local tasks: can be answered based on single sub-process • three global tasks: can be answered based on modularized model • three global-local tasks: require information from modularized model and a sub-process	time taken	• perceived ease of use (4 items) • perceived usefulness use (4 items)	• modularity [Secondary Notation, Mixed – balanced block design] • paper versus computer [Medium, Mixed – balanced block design] • experience and knowledge in process modeling and BPMN [User, Between] • familiarity with process and domain [User, Between]
(Weitlaner et al. 2013)	77 (large)	employees from companies in an Austrian group	4	UML, BPMN, EPC, SBD (storyboard design)	six comprehension tasks: • single-choice dichotomous (true/false) questions • multiple-choice (4 possible answers)	-	-	• previous knowledge [User, Between] • level of education [User, Between] • control flow element (order, repetition, concurrency) [Task→Model, Within] • modeling notation (UML AD, BPMN, EPC, SBD (storyboard design)) [Notation, Between]
(Wiebring and Sandkuhl 2015)	113 (large)	employees of a medium-sized organization from the utility industry	1	BPMN	sixteen comprehension tasks (e.g., identifying and counting elements of the process, such as the roles	-	perception of process models (e.g., connection between tasks and roles)	• modeling notation (flow Diagram, BPMN, UML AD, eEPC) [Notation, Between]

Source Reference	**# of Participants**	**Type of Participants**	**# of Models**	**Notation of Models**	**Measurement of Objective Comprehension**	**Measure-ment of Time**	**Measurement of Subjective Comprehension and User Preferences**	**Independent Variables [Category: Medium/Notation/Secondary Notation/Label/Model/Task/User; Research Design: Between/Within/Mixed]**
					involved; yes/ no-questions on, e.g., mutual exclusion, interpreting the process flow correctly)			

Appendix C: Empirical Articles on Comprehension – Variable-Based Overview

Table 5 Main results of empirical articles

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
Presentation Medium									
(Mendling et al. 2012b)	presentation medium	paper versus computer	computer: online questionnaire	no significant effect	ANOVA	not significant → no evidence	-	113 (very large)	6
(Recker et al. 2014)	presentation medium	paper versus computer	computer: online questionnaire	no significant effect	ANOVA	not significant → no evidence	-	92 (large)	2
(Turetken et al. 2016)	presentation medium	paper versus computer	computer: interactive web-based visualizations	no significant differences between comprehension accuracy and efficiency (accuracy/time spent) for computer and paper versions	ANOVA	not significant → no evidence	• perceived usefulness: F=9.54, p=0.002 • perceived ease of understanding: F=4.32, p=0.038 → strong	60 (medium)	2
Notation									
(Ottensooser et al. 2012)	notation: representation paradigm	BPMN notation (graphical) versus written use cases (textual)	manipulated factor constituting experimental groups	• Reading a BPM model increased comprehension accuracy for users trained in process modeling but not significantly for users not trained in process modeling (H2). • A BPM model increased comprehension accuracy more than a written use case did (H4). • Reading a process model in written use case notation, followed by a BPM model, increases comprehension accuracy more than a written use case alone (H5).	Wilcoxon rank-sum test	• H2: users trained in process modeling (p<0.001), users not trained in process modeling (p=0.15) • H4: users trained in process modeling (p=0.05), users not trained in process modeling (p=0.77) → moderate	-	196 (very large)	6
(Rodrigues et al. 2015)	notation: representation paradigm	BPMN versus textual process-description modeling notation	manipulated factor constituting experimental groups	• experienced group: BPM increases comprehension accuracy more than text does. • inexperienced group: no significant difference	Mann-Whitney test	• experienced group: BPM leads to higher comprehension accuracy than text (79.2% versus 71.6%, p=0.049) • inexperienced group: not significant → moderate	-	73 (large)	1
(Figl and Recker 2016)	notation: representation paradigm	text, structured text, diagram (BPMN)	manipulated factor	Users preferred BPMN over structured text and text (least preferred option).	descriptive statistics	[empirical article on subjective comprehension and user preferences]	user preferences [scores from 0 to 100; scores > 50 indicate preference for the first option] • diagrammatic representations over text (M=80.05, SD=2.71) • diagrammatic representations over	120 (very large)	1

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Mod els
						Objective	Subjective		
							structured text (M=74.42, SD=3.07) - structured text over text (M=70.24, SD=3.42) → weak		
(Pichler et al. 2012)	notation: representation paradigm	declarative versus imperative models	manipulated factor constituting experimental groups	Higher comprehension accuracy was achieved and less time was taken for imperative models than for declarative models.	Sheirer-Ray-Hare test	- comprehension accuracy: p=0.001 - [time taken: p=0.002] → strong	-	27 (small)	4
(Figl and Recker 2016)	notation: representation paradigm	use of icons (with/without icons)	manipulated factor	weak preference for icons over no icons	descriptive statistics	[empirical article on subjective comprehension and user preferences]	user preferences [scores from 0 to 100; scores > 50 indicate preference for the first option] - icons over no icons (M=62.99, SD=3.50) → weak	120 (very large)	1
(Hipp et al. 2014)	notation: representation paradigm	new visualizations (bubble visualization concept, BPMN3D, network visualization concepts, thin line concept)	manipulated factor: - bubble visualization concept: activities as bubbles - BPMN3D: BPMN, but data objects are outsourced in a third dimension - network visualization concepts: one activity of interest is dynamically emphasized; others are grayed out - thin line concept: sequence flow of process activities is separated from data objects	BPMN3D was evaluated best for all three dimensions, followed by Bubble, thin line, and finally network visualization.	unclear	[empirical article on subjective comprehension and user preferences]	- perceived comprehensi-bility (p=0.04) - perceived comprehensi-bility of sequence flow (p=0.012) - perceived clarity and overview (p=0.011) → strong (BPMN3D rated best)	22 (small)	1
(Döhring et al. 2014)	notation: primary notation	C-YAWL and vBPMN	manipulated factor constituting experimental groups; process-variant management notations with differing configuration and adaptation mechanisms; C-YAWL without modularization support, v-BPMN with modularization support in an execution	- There was a significant effect on task success (combination of comprehension and modeling tasks), time taken—although there was also an interaction effect with modeling versus the understanding task type, as modeling in the tool is especially time-consuming in C-YAWL—and easiness of tasks. - There was no significant effect on task confidence. - vBPMN easier to comprehend and participants took less time than they did with C-YAWL.	Mann-Whitney U-test, Wilcoxon rank-sum test, t-tests (paired), ANOVA	- task success: Wilcoxon-test: p=0.048 (after Bonferroni) - [processing time: ANOVA: $F_{1,12}=19.94$, p<0.001] → strong	easiness of tasks: Wilcoxon-test: p=0.025 (after Bonferroni) → strong	14 (small)	2 (1 per tool)

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Mod els
						Objective	Subjective		
			tool; also included modeling tasks						
(Sarshar and Loos 2005)	notation: primary notation	EPC versus Petrinet	manipulated factor constituting experimental groups	• There was higher comprehension accuracy for EPC than for Petrinets. (XOR situations were understood better in EPC, and AND was understood equally well.) • Perceived ease-of-use of the EPC notation was higher (tendency). • There was no significant difference concerning perceived ease-of-use of control flow.	information missing	no numbers → weak	• perceived ease of use: statistical trend • perceived ease of use of control flow: not significant → weak	50 (medium)	1
(Recker and Dreiling 2007)	notation: primary notation	modeling notation (EPC versus BPMN; participants were familiar with EPC but not with BPMN)	informationally equivalent experimental groups; higher number of semantically different language constructs used in BPMN than in EPC	There was no significant effect; although BPMN was unfamiliar to participants, it received higher (though not significantly higher) comprehension accuracy.	ANCOVA	not significant → no evidence	not significant → no evidence (numbers available for only one of two models)	69 (medium)	2
(Recker and Dreiling 2011)	notation: primary notation	modeling notation (EPC versus BPMN; participants were familiar with EPC but not with BPMN)	informationally equivalent experimental groups; higher number of semantically different language constructs used in BPMN than in EPC	• BPMN had a positive effect on model-based transferability test scores. (Previous knowledge of a modeling notation assists users in understanding the business context as depicted in this model; however, it does not assist in developing deep transfer abilities.) There was no significant effect on: • comprehension accuracy (model comprehension, inferential transfer ability test scores, retention test scores); • task completion times (model comprehension, transfer ability test scores, retention ability test scores); • ease of understanding.	MANCOVA	• no effect on comprehension accuracy • effect on model-based transfer ability test scores: o goods receipt task: $F_{1,11}=2.76$, $p=0.01$ o claims handling task: $F_{1,11}=3.69$, $p<0.001$ → no evidence	not significant → no evidence	68 (medium)	2
(Recker et al. 2005)	notation: primary notation	C-EPC versus EPC	manipulated factor; configurable EPC (C-EPC) versus EPC	• C-EPC was found to be more useful than EPC. • Some visualizations should be improved (e.g., definitions for configurable gateways).	descriptive statistics only	[empirical article on subjective comprehension and user preferences]	no numbers (only a subset of numbers for results were in the paper) → weak	16 (small)	2
(Weitlaner et al. 2013)	notation: primary notation	UML AD, BPMN, EPC, SBD (storyboard design)	• manipulated factor constituting experimental groups • informationally equivalent	Accuracy was highest for SBD and lowest for EPC.	descriptive statistics only	no numbers → weak	-	77 (large)	4
(Sandkuhl and Wiebring 2015)	notation: primary notation	flow diagrams, eEPC, UML AD, BPMN	manipulated factor constituting experimental groups	• Comprehension accuracy: highest accuracy for UML AD (M=11.93), followed by eEPC (M=11.50), flow diagram (M=11.32), and finally BPMN (M=11.29). • Perception of notation: highest performance for eEPC (M=2.70), UML AD (M=2.29), followed by	descriptive statistics only	descriptive statistics only → weak	“perception”: descriptive statistics only → weak	113 (large)	1

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Models
						Objective	Subjective		
				BPMN (M=2.19), and finally flow diagram (M=2.14).					
(Natschläger 2011)	notation: primary notation	BPMN, deontic BPMN	deontic BPMN: highlights mandatory and optional activities with color and uses textual decisions and deontic constructs in the corresponding activity symbols (e.g., “O() for obligatory activities, P() for permissive activities, (A|B) for preconditions). For instance, optionality is not expressed through gateways, but through textual decisions.	Deontic BPMN could reduce comprehension errors (higher comprehension accuracy); however, preconditions were difficult to understand in deontic BPMN.	descriptive statistics only	204 versus 176 comprehension errors in BPMN compared to deontic BPMN (-13.7%) → weak	-	22 (small)	2*4 models
(Jošt et al. 2016)	notation: primary notation	BPMN, UML AD, EPC	manipulated factor constituting experimental groups	• Model 1: Comprehension accuracy was higher for UML AD than for EPC or BPMN. • Model 2: Time taken was lower for EPC than for BPMN. • Model 3: Comprehension accuracy was higher for UML AD than for EPC. • Model 4: no significant differences	Mann Whitney post hoc tests with Bonferroni correction	• Model 1: comprehension accuracy was higher for UML AD than for EPC (std. test statistics=3.052, p=0.007) and BPMN (std. test statistics=2.708, p=0.02) • [Model 2: time taken was lower for EPC than for BPMN (std. test statistics=-2.602, p=0.028)] • Model 3: comprehension accuracy was higher for UML AD than for EPC (std. test statistics=3.511, p=0.001) → moderate	-	103 (large)	4
(Stitzlein et al. 2013)	notation: primary notation	BPMN, health process notation (HPN)	manipulated factor constituting experimental groups	• There was no significant overall effect on comprehension accuracy but there was a trendwise effect on comprehension efficiency (accuracy/time taken). • BPMN models increased comprehension accuracy for simple tasks; HPN models increased it for complex tasks. • There were statistically significant differences for neither perceived ease of use nor perceived usefulness.	ANOVA	comprehension accuracy: • overall effect: not significant • simple tasks: $F_{1,12}$=14.11, p=0.003 • complex tasks: $F_{1,12}$=8.90, p=0.01 [comprehension efficiency: $F_{1,12}$=10.73, p=0.007] → conflicting	not significant → no evidence	16 (small)	2
(Kock et al. 2008)	notation: primary notation	activity flow representation versus communication flow representation	activity flow representation versus communication flow representation	Communication-flow orientation was not perceived as significantly more difficult to understand than activity-flow representation. (No information was offered on whether it was easier.)	unclear	[empirical article on subjective comprehension and user preferences]	not significant → no evidence	210 (very large)	1

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
(Kock et al. 2009)	notation: primary notation	high versus low communication-flow orientation	• high communication-flow orientation: communication-flow diagrams • low communication-flow orientation: functional flowcharts with swim lanes	Comprehension accuracy was higher for models with a greater communication-flow orientation.	SEM	[empirical article on subjective comprehension and user preferences]	$\beta=0.269$, $p<0.01$ → strong	78 (large)	1
(Figl et al. 2013a)	notation: primary notation/ notational characteristics	perceptual discriminability deficiencies	notation with perceptual discriminability deficiencies (YAWL), and two notations without global deficiencies (UML AD, BPMN) as a reference value	• Perceptual discriminability deficiencies significantly influenced comprehension accuracy of the large model and tended to influence comprehension accuracy of the small model, but not the accuracy in the text-model comparison task. • There was a significant effect for time taken with the large model and the text-model comparison (average increase of 1 minute). • There was a significant effect of perceptual discriminability deficiencies on cognitive load for the tasks with both the small model and the large model.	ANCOVA	comprehension accuracy: • small model ($F=4.53$, $p=0.06$) • large model ($F=6.69$, $p=0.01$) • text-model comparison: not significant [time taken: • large model ($F=7.02$, $p=0.01$) • text-model comparison ($F=5.17$, $p=0.03$)] → moderate	cognitive load: • small model ($F=3.97$, $p=0.05$) • large model ($F=5.26$, $p=0.02$) • text-model comparison: not significant → strong	136 (very large)	3
(Figl et al. 2013a)	notation: primary notation/ notational characteristics	semiotic clarity deficiencies	notation with semiotic clarity deficiencies (EPC) and 2 notations without global deficiencies (UML AD, BPMN) as reference values	• Semiotic clarity deficiencies had a statistically significant influence on comprehension accuracy of the large model but not on comprehension accuracy of the small model or the text-model comparison task. • Semiotic clarity deficiencies were found to have a trendwise influence on time taken in only one of three tasks (text-model comparison). • Symbol sets with semiotic clarity deficiencies imposed a higher cognitive load (trendwise).	ANCOVA	comprehension accuracy: • small model: not significant • large model ($F=7.03$, $p=0.01$) • text-model comparison: not significant [time taken for text-model comparison ($F=3.47$, $p= 0.07$)] → moderate	cognitive load: • small model: $F=2.89$, $p=0.09$ • large model: $F=5.01$, $p=0.07$ • text-model comparison: $F=2.89$, $p=0.09$ → weak	136 (very large)	3
(Recker 2013)	notation: notational characteristics	implicit/explicit representation of parallel split and simple merge scenarios	manipulated factor constituting experimental groups; • explicit: with BPMN gateway constructs; perceptual discriminability effect through use of gateways • implicit: without BPMN gateway constructs	• The use (versus non-use) of gateway constructs had a consistently significant effect on comprehension accuracy: in all cases, comprehension accuracy increased when gateway constructs were used in the model. • The effects on time taken and perceived ease of understanding were not significant.	ANCOVA	$F=3.78$, $p=0.005$, $\eta^2=0.03$ → strong	not significant → no evidence	98 (large)	2*3 models
(Figl et al. 2013b)	notation: notational characteristics	symbol aesthetics	self-developed user evaluation of symbols questionnaire on four dimensions: semantic transparency, perceptual discriminability, pop-out, and aesthetics with 3-4	The aesthetic design of symbols lowered the perceived cognitive load but did not significantly affect comprehension accuracy.	regression analysis	not significant → no evidence	perceived cognitive load: aesthetic design of symbols ($\beta=0.09$, $p\leq0.05$) → strong	154 (very large)	4

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
			items on each dimension (and a 5-point Likert scale)						
(Figl et al. 2013b)	notation: notational characteristics	perceptual pop-out of symbols	self-developed user evaluation of symbols questionnaire on four dimensions: semantic transparency, perceptual discriminability, pop-out, and aesthetics with 3-4 items on each dimension (and a 5-point Likert scale)	Pop-out improved comprehension accuracy and perceived control flow comprehension and lowered the perceived cognitive load.	regression analysis	comprehension accuracy: $\beta=0.09$, $p\leq0.05$ → strong	- perceived control flow comprehension: $\beta=0.13$, $p\leq0.001$ - perceived cognitive load: $\beta=0.13$, $p\leq0.001$ → strong	154 (very large)	4
(Figl et al. 2013b)	notation: notational characteristics	semantic transparency of symbols	self-developed user evaluation of symbols questionnaire on four dimensions: semantic transparency, perceptual discriminability, pop-out, and aesthetics with 3-4 items on each dimension (and a 5-point Likert scale)	Semantic transparency lowered the perceived cognitive load but did not significantly affect comprehension accuracy.	regression analysis	not significant → no evidence	perceived cognitive load: semantic transparency ($\beta=0.11$, $p\leq0.001$) → strong	154 (very large)	4
(Figl et al. 2013b)	notation: notational characteristics	perceptual discriminability of symbols	self-developed user evaluation of symbols questionnaire on four dimensions: semantic transparency, perceptual discriminability, pop-out, and aesthetics with 3-4 items on each dimension (and a 5-point Likert scale)	Perceptual discriminability improved comprehension accuracy and perceived control flow comprehension and lowered the perceived cognitive load.	regression analysis	comprehension accuracy: $\beta=0.09$, $p\leq0.05$ → strong	- perceived control flow comprehension: $\beta=0.10$, $p\leq0.05$ - perceived cognitive load: $\beta=0.14$, $p\leq0.00$ → strong	154 (very large)	4
(Kock et al. 2009)	notation: notational characteristics	ease of generating the models	- experimental task: redesigning a process in groups, with two modeling approaches - ease of generating the models (4 items on a 7-point scale (e.g. "It is easy to conceptualize a process using this approach"))	A model that is perceived as easier to generate was also perceived as easier to understand.	SEM	[empirical article on subjective comprehension and user preferences]	$\beta=0.503$, $p<0.01$ → strong	78 (large)	1
Secondary Notation									
(Johannsen et al. 2014a)	secondary notation: decomposition	decomposition: full compliance versus moderate violation versus strong violation	three groups: 0/3/5 conditions violated; - moderate violation: minimality, determinism, losslessness violated - strong violation: minimality,	Full compliance and moderate violation increased comprehension accuracy and the models were perceived to be easier to understand than in the strong-violation group; there was no significant effect (unclear result) for problem-solving.	regression analysis (with strong-violation group as base level)	comprehension accuracy: - no-violation group: $t=3.68$, $p=0.001$ - moderate-violation group: $t=3.45$, $p=0.001$ → strong	perceived ease of understanding: - no-violation group: $t=3.698$, $p=0.001$ - moderate-violation group: $t=9.994$, $p<0.001$ → strong	51-53 (medium)	1

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Models
						Objective	Subjective		
			determinism, losslessness, minimum coupling, strong cohesion violated						
(Reijers et al. 2011b)	secondary notation: decomposition	modularization	manipulated factor constituting experimental groups; model with sub-processes versus flattened model without sub-processes	There was a significant difference for one of the models: Modularity was positively connected with comprehension accuracy. (This model had a higher fan in-out metric and a higher number of subprocesses, so the difference between flattened and modular version was larger.)	t-test (based on a median-split variable)	one of the models: p=0.001 → moderate	-	28 (small)	2
(Turetken et al. 2016)	secondary notation: decomposition	modularity presentation	manipulated factor constituting experimental groups; fully flattened model versus flattened view with visual “groups” for expanded sub-processes versus models with sub-processes (sub-processes collapsed and shown in separate models)	There was a significant effect of modularity on • comprehension accuracy—both flattened models increase comprehension accuracy more than the use of collapsed sub-processes, the largest difference for scores of local questions, which involve information about sub-processes— • perceived usefulness—the difference between fully flattened models versus sub-processes collapsed was significant— • perceived ease of understanding—the difference between fully flattened models versus both other versions was significant.	Kruskal-Wallis test	comprehension accuracy: H=8.49, p=0.001 [comprehension efficiency: not significant] → strong	• perceived usefulness: F=13.12, p=0.001 • perceived ease of understanding: F=13.59, p=0.001 → strong	60 (medium)	2

(Bera 2012)	secondary notation: Gestalt theory	with swim lanes versus without swim lanes	with swim lanes versus without swim lanes (lanes were cut, but actors were left on the side of the models)	• Swim lanes reduced the time taken for answering comprehension tasks but did not significantly affect comprehension accuracy.
				• Performance in problem-solving tasks was higher, but there was no significant effect on time taken.
					ANCOVA	comprehension accuracy: not significant
						• [time taken:
						  o Model 1: 11.00, p<0.01
						  o Model 2: 11.29, p<0.01
						• problem-solving tasks:
						  o Model 1: 8.40, p=0.01
						  o Model 2: 10.15, p<0.01]
						→ no evidence
							-	51 (medium)	2

(Jeyaraj and Sauter 2014)	secondary notation: Gestalt theory	with swim lanes versus without swim lanes	• without swim lanes: names of actors shown as first words in the activity labels • study 1: similar physical position of activities, decisions, and arrows • study 2: different physical position of activities, decisions and arrows • identification of internal/external	• higher comprehension of non-swim lane diagrams for selected information • study 1: no significant differences, with the exception of identification of external actors • study 2: no significant differences, with the exception of identification of external actors and external activities and time taken identifying activities (participants took longer for non-swim lane diagrams)	t-test	comprehension accuracy: not significant • [identification of external actors: t=-3.70, p<0.001 (study 1) • identification of external actors: t=-2.34, p<0.05 (study 2) • identification of external activities: t=-3.48, p<0.001 (study 2) • time taken identifying activities: t=-3.47, p<0.001 (study 2)]	-	142 in study 1, 131 in study 2 (very large)	1 model on a manufacturing process (with/without swim lane

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Mod els
						Objective	Subjective		
			actors as well as activities			→ moderate			version)
(Kummer et al. 2016)	secondary notation: Gestalt theory	different colors for different symbols	black/white versus bright colors with high contrast based on Asian preferences	- Color had no significant main effect on comprehension accuracy. - Members of Confucian cultures perceived colored models as less difficult than uncolored models. - Members of Confucian cultures perceived colored models as less difficult than members of Germanic cultures did.	MANCOVA	not significant → no evidence	perceived difficulty: - Model 1: interaction effect between culture and color: $F_{1,\ 115}=9.22$, $p < 0.001$ - Model 2: interaction effect between culture and color: $F_{1,\ 115}= 4.90$, $p < 0.05$ → moderate	127 (very large)	2
(Reijers et al. 2011a)	secondary notation: Gestalt theory	syntax highlighting	highlighting of matching operator pairs in different colors (e.g., AND-split/AND-join pair)	- overall: higher comprehension accuracy with highlighting - novices: higher comprehension accuracy; no significant difference in time taken - experts: significant difference in neither comprehension accuracy nor time taken	Mann-Whitney test, two-tailed	effect of highlighting on comprehension accuracy - overall: p=0.049 - novices: p=0.017 - experts: not significant → moderate	-	70 (medium)	1
(Petrusel et al. 2016)	secondary notation: Gestalt theory	color highlighting	model elements in the task-relevant region are colored red	- Colored relevant model elements did not significantly affect comprehension accuracy but lowered time taken. - "Mental effort," measured by fixations and fixation durations, decreased significantly.	Wilcoxon test	comprehension accuracy: not significant [time taken: Z=3.039, p=0.002] → no evidence	-	75 (large)	16
(Petrusel et al. 2016)	secondary notation: layout	task-specific layout	model elements in the task-relevant region are larger-sized than the rest of the gateways and repositioned close to each other	- Task-specific layout of relevant model elements significantly affected neither comprehension accuracy nor time taken. - "Mental effort," measured by fixations and fixation durations, decreased significantly.	Wilcoxon test	comprehension accuracy: not significant → no evidence	-	75 (large)	16
(Figl and Strembeck 2015)	secondary notation: layout	flow direction	manipulated factor constituting experimental groups, left-to-right, right-to-left, top-to-bottom, bottom-to-top	There was no significant main effect of flow direction on comprehension accuracy, perceived ease of use of model, or time taken.	ANCOVA	not significant → no evidence	not significant → no evidence	44 (medium)	2
Label									
(Mendling et al. 2010c)	label: syntactic naming conventions	labels of various styles	- verb-object style - action-noun style - other styles	There was a significant effect of label style on perceived usefulness. (Verb-object label style was rated highest in perceived usefulness, followed by action-noun label style, and finally the rest of the labels.)	ANOVA	[empirical article on subjective comprehension and user preferences]	perceived usefulness: F=18.495, p<0.001 (analysis based on labels: 6 labels×29 participants=174 data points) → strong	29 (small)	1
(Mendling et al. 2010c)	label:	perceived ambiguity of labels	participants were asked to identify the three most ambiguous labels	Usefulness was rated higher for unambiguous labels than for ambiguous labels.	ANOVA	[empirical article on subjective comprehension and user preferences]	usefulness: F=31.553, p<0.001	29 (small)	1

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Models
						Objective	Subjective		
	semantic naming conventions						(analysis based on labels: 6 labels×29 participants=174 data points) → strong		
(Koschmider et al. 2015b)	label: semantic naming conventions	unrevised and revised activity labels from a linguistic perspective	unrevised versus revised activity labels: revised by students of German linguistics (e.g., using the most frequent words, considering labels in the neighborhood, replacement if synonyms are used in the process model)	• For the model that did not use domain-specific vocabulary, users rated linguistically superior labels easier to understand (based on a standard language dictionary). • For the model with domain-specific vocabulary, the unrevised version was rated easier to understand. (It is hypothesized that a domain glossary or ontology would be better suited to revising the labels.)	t-tests	[empirical article on subjective comprehension and user preferences]	p<0.001 for all models and orders, but different directions of effect → conflicting	49 (medium)	2
(Mendling and Strembeck 2008)	label: semantic naming conventions	abstract versus concrete labels	manipulated factor constituting experimental groups; abstract activity labels (e.g., A, B, C) versus illustrative, textual labels (e.g., “check credit limit”)	Both groups had similar comprehension accuracy.	descriptive statistics only	not significant → no evidence	-	42 (small)	6
(Mendling et al. 2012b)	label: semantic naming conventions	abstract versus concrete labels	• abstract: capital letters • concrete: verb-object style	Comprehension accuracy was higher for abstract labels; time taken was lower for abstract labels.	ANOVA	• comprehension accuracy: F=5.05, p=0.03 • [time taken: F=3.90, p=0.05] → strong	-	113 (very large)	6
(Figl and Strembeck 2015)	label: semantic naming conventions	abstract versus concrete labels	• abstract: capital letters • concrete: verb-object style	• There was no significant effect on comprehension accuracy. • Participants took more time to answer questions on the model with concrete labels than the model with abstract labels. • There was no significant effect on perceived ease of use of the model.	ANCOVA	[time taken: $F_{1,29}$=6.39, p=0.02] → no evidence	not significant → no evidence	44 (medium)	2
(Mendling and Strembeck 2008)	label: label design	length of textual labels	string length of all textual activity labels	Longer text labels reduced comprehension accuracy.	correlation analysis (Pearson)	r=0.84, p=0.01 → strong	-	42 (small)	6
Model Characteristics									
(Recker 2013)	model characteristics: size measures	arcs and nodes	based on arcs and nodes: • “low complexity” model (10 arcs, 7 nodes) • “average complexity” model (17 arcs, 13 nodes)	• More complex models reduced comprehension accuracy. • Significant differences were also found when comparing time taken for the pair-average-versus-high-complexity models, but not for the pair-low-versus-average-complexity model.	t-test (paired), ANCOVA	• comprehension accuracy: F=22.76, p=0.00; t-test p<0.01 for both comparisons (low/average, average/high) • [time taken: overall test not significant; t-test p=0.01 for comparison	-	98 (large)	2*3 mod-els

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
			• “high complexity” model (61 arcs, 50 nodes)			(average/high), but not significant for comparison (low/average)] → strong			
(Mendling and Strembeck 2008)	model characteristics: size measures	number of nodes	number of nodes	no correlation (but positive)	correlation analysis (Spearman)	not significant → no evidence	-	42 (small)	6
(Sánchez-González et al. 2010)	model characteristics: size measures	number of nodes	number of activities and routing elements	The higher the nodes are, the lower the comprehension accuracy.	correlation and regression analysis	r=-0.70, p=0.003 → strong		71 (large)	15
(Mendling and Strembeck 2008)	model characteristics: size measures	diameter	length of the longest path in the process model from a start node to an end node	There was no significant correlation with comprehension accuracy (but a negative sign).	correlation analysis (Spearman)	not significant → no evidence	-	42 (small)	6
(Sánchez-González et al. 2010)	model characteristics: size measures	diameter	length of the longest path in the process model from a start node to an end node	Higher diameter led to lower comprehension accuracy.	correlation and regression analysis	r=-0.70, p=0.004 → strong	-	71 (large)	15
(Aguilar et al., 2008)	model characteristics: size measures	number of exclusive data-based decisions	number of exclusive data-based decisions	There was a negative correlation with time taken to answer tasks and comprehension efficiency (score/time ratio).	correlation analysis (Spearman)	reached significance on a 0.05 level in at least two of five sub-samples → weak	-	110 participants in 5 samples (large)	10
(Sánchez-González et al. 2012)	model characteristics: size measures	number of gateways	number of gateways	fairly easy ≤ 9 ≤ easy ≤ 12 ≤ moderately difficult ≤ 16 ≤ difficult ≤ 18 ≤ fairly difficult	ANOVA tests (significant) between models were used to identify “thresholds” between different models.	no numbers → weak	-	28 for obtaining thresholds, 23 for cross-validation (medium)	10
(Reijers and Mendling 2011)	model characteristics: size measures	number of OR joins	number of OR joins	no significant correlation	correlation analysis (Pearson)	not significant → no evidence	-	73 (large)	12
(Aguilar et al. 2008)	model characteristics: size measures	number of end message events	number of end message events	There was a correlation with comprehension efficiency (score/time ratio) (direction unclear).	correlation analysis (Spearman)	reached significance on a 0.05 level in at least two of five sub-samples → weak (direction of effect unclear)	-	110 (large)	10
(Aguilar et al. 2008)	model characteristics: size measures	number of events in the model	number of events in the model	There were negative correlations with time taken to answer tasks and comprehension efficiency (score/time ratio).	correlation analysis (Spearman)	reached significance on a 0.05 level in at least two of five sub-samples → weak	-	110 (large)	10

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
(Aguilar et al. 2008)	model characteristics: size measures	number of intermediate events in the model	number of intermediate events in the model	There was a negative correlation with time taken to answer tasks.	correlation analysis (Spearman)	reached significance on a 0.05 level in at least two of five sub-samples → weak	-	110 (large)	10
(Aguilar et al. 2008)	model characteristics: size measures	number of intermediate message events	number of intermediate message events	There were correlations with time taken to answer tasks and comprehension efficiency (score/time ratio) (direction unclear)	correlation analysis (Spearman)	reached significance on a 0.05 level in at least two of five sub-samples → weak (direction of effect unclear)	-	110 (large)	10
(Aguilar et al. 2008)	model characteristics: size measures	number of sequence flows from events	number of sequence flows from events	There were negative correlations with time taken to answer tasks and comprehension efficiency (score/time ratio).	correlation analysis (Spearman)	reached significance on a 0.05 level in at least two of five sub-samples → weak	-	110 (large)	10
(Dumas et al. 2012)	model characteristics: modularity/ structuredness	structuredness	four groups of models: structuredness (structured and unstructured) × cyclicity (cyclic, acyclic)	• There was no overall effect, but in two models the structured version increased comprehension accuracy, and in two other models it reduced comprehension accuracy. • There were no significant differences in the other four models; the authors hypothesized that structuring may be preferred only when it does not increase the number of gateways.	ANOVA	higher comprehension effectiveness for structured models: • eff. size = 0.18, p = 0.009 • eff. size = 0.17, p = 0.011 higher comprehension effectiveness for unstructured models: • eff. size = 0.12, p = 0.0032 • eff. size = 0.14, p = 0.023 → conflicting	-	55 (medium)	8
(Mendling and Strembeck 2008)	model characteristics: modularity/ structuredness	structuredness of the process graph	"one minus the number of nodes in structured blocks divided by the number of nodes"	no significant effect	correlation analysis (Spearman)	not significant → no evidence	-	42 (small)	6
(Sánchez-González et al. 2010)	model characteristics: modularity/ structuredness	maximum nesting depth	maximum nesting of structured blocks in a process model ["depth"]	higher depth negatively related to comprehension accuracy	correlation and regression analysis	no numbers → weak	-	71 (large)	15
(Figl and Laue 2011)	model characteristics: modularity/ structuredness	process structure tree (PST) distance (measured for each comprehension task)	process structure tree (PST) distance between two elements A and B = the number of arcs between A and B in the PST minus one	The higher the PST distance is, the lower the comprehension accuracy and the higher the perceived difficulty.	ANCOVA	comprehension accuracy: ($F_{1,55}$ = 4.32, p = 0.042) → strong	perceived difficulty ($F_{1,55}$ = 22.04, p < 0.001) → strong	199 (very large)	4
(Figl and Laue 2015)	model characteristics: modularity/ structuredness	PST distance (measured for each comprehension task)	PST distance between two elements A and B = the number of arcs between A and B in the PST minus one	The higher the PST distance is, the lower the comprehension accuracy and the higher the perceived difficulty.	ANCOVA	comprehension accuracy: ($F_{1,53}$=22.08, p<0.001) → strong	perceived difficulty ($F_{1,53}$=17.79, p<0.001) → strong	155 (very large)	4

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
(Mendling and Strembeck 2008)	model characteristics: modularity/ structuredness	separability	relates the number of cut-vertices to the number of nodes	Higher separability is associated with higher comprehension accuracy.	correlation analysis (Spearman)	r=0.89, p=0.019 (n=6 models) → strong	-	42 (small)	6
(Figl and Laue 2011)	model characteristics: modularity/ structuredness	cut-vertex (measured for each comprehension task)	cut vertex: presence of a single arc in the BPM separates the BPM into two disjointed parts	no significant effect	ANCOVA	not significant → no evidence	not significant → no evidence	199 (very large)	4
(Figl and Laue 2015)	model characteristics: modularity/ structuredness	cut-vertex (measured for each comprehension task)	cut vertex: presence of a single arc in the BPM separates the BPM into two disjointed parts	no significant effect	ANCOVA	not significant → no evidence	not significant → no evidence	155 (very large)	4
(Sánchez-González et al. 2010)	model characteristics: modularity/ structuredness	sequentiality	degree to which the model is constructed out of pure sequences of task	the higher the degree of sequentially, the higher the comprehension accuracy	correlation and regression analysis	no numbers → weak	-	71 (large)	15
(Döhring et al. 2014)	model characteristics: gateway interplay/ control structures	simple and complex model	• complex: name registration for a child • simple: travel booking • (no further details on the kind of "complexity" are given in the paper)	• There was no significant effect on task success—comprehension accuracy and modeling tasks are analyzed in combination—or easiness of tasks or task confidence. • However, there was an effect on time taken: time taken was lower for lower-complexity models than it was for higher-complexity models.	Mann-Whitney U-test, Wilcoxon rank-sum test, t-tests (paired), ANOVA	• [processing time: ANOVA: $F_{1,12}$=29.97, p<0.001] • task success—comprehension accuracy and modeling tasks are analyzed in combination—not significant → no evidence	easiness of tasks, task confidence: not significant → no evidence	14 (small)	2 (1 per complexity)
(Reijers and Mendling 2011)	model characteristics: gateway interplay/ control structures	gateway mismatch	sum of gateway pairs that do not match each other	no significant correlation	correlation analysis (Pearson)	not significant → no evidence	-	73 (large)	12
(Sánchez-González et al. 2010)	model characteristics: gateway interplay/ control structures	gateway mismatch	sum of gateway pairs that do not match each other	The higher the gateway mismatch is, the lower comprehension accuracy.	correlation and regression analysis	no numbers → weak	-	71 (large)	15
(Sánchez-González et al. 2012)	model characteristics: gateway	gateway mismatch	sum of gateway pairs that do not match each other	easy ≤ 6 ≤ moderately difficult ≤ 15 ≤ difficult ≤ 20 ≤ fairly difficult	ANOVA tests (significant) between models	no numbers → weak	-	28 for obtaining thresholds,	10

<table>
<tr><th rowspan="2">Source Reference</th><th rowspan="2">Category</th><th rowspan="2">Independent Variable</th><th rowspan="2">Measurement of Independent Variable</th><th rowspan="2">Results</th><th rowspan="2">Analysis Method</th><th colspan="2">Statistical Significance and Level of Evidence of Effect on Comprehension</th><th rowspan="2"># of Partici-pants</th><th rowspan="2"># of Mod els</th></tr>
<tr><th>Objective</th><th>Subjective</th></tr>
<tr><td></td><td>interplay/ control structures</td><td></td><td></td><td></td><td>were used to identify “thresholds” between different models.</td><td></td><td></td><td>23 for cross-validation (medium)</td><td></td></tr>
<tr><td>(Sánchez-González et al. 2012)</td><td>model characteristics: gateway interplay/ control structures</td><td>control-flow complexity (CFC)</td><td>complexity of split gateways</td><td>fairly easy ≤ 13 ≤ easy ≤ 22 ≤ moderately difficult ≤ 37 ≤ difficult ≤ 51 ≤ fairly difficult</td><td>ANOVA tests (significant) between models were used to identify “thresholds” between different models.</td><td>no numbers
→ weak</td><td>-</td><td>28 for obtaining thresh-olds, 23 for cross-validation (medium)</td><td>10</td></tr>
<tr><td>(Sánchez-González et al. 2010)</td><td>model characteristics: gateway interplay/ control structures</td><td>gateway heterogeneity</td><td>frequency with which different types of gateways are used in a model</td><td>Higher gateway heterogeneity lowered comprehension accuracy.</td><td>correlation and regression analysis</td><td>r=-0.62, p=0.014
→ strong</td><td>-</td><td>71 (large)</td><td>15</td></tr>
<tr><td>(Sánchez-González et al. 2012)</td><td>model characteristics: gateway interplay/ control structures</td><td>gateway heterogeneity</td><td>frequency with which different types of gateways are used in a model</td><td>fairly easy ≤ 0.62 ≤ easy ≤ 0.79 ≤ moderately difficult ≤ 0.92 ≤ difficult ≤ 0.94 ≤ fairly difficult</td><td>ANOVA tests (significant) between models were used to identify “thresholds” between different models.</td><td>no numbers
→ weak</td><td>-</td><td>28 for obtaining thresh-olds, 23 for cross-validation (medium)</td><td>10</td></tr>
<tr><td>(Mendling and Strembeck 2008)</td><td>model characteristics: gateway interplay/ control structures</td><td>gateway heterogeneity</td><td>frequency with which different types of gateways are used in a model</td><td>no significant correlation (but negative)</td><td>correlation analysis (Spearman)</td><td>not significant
→ no evidence</td><td>-</td><td>42 (small)</td><td>6</td></tr>
<tr><td>(Reijers and Mendling 2011)</td><td>model characteristics: gateway interplay/ control structures</td><td>gateway heterogeneity</td><td>frequency with which different types of gateways are used in a model</td><td>no significant correlation</td><td>correlation analysis (Pearson)</td><td>not significant
→ no evidence</td><td>-</td><td>73 (large)</td><td>12</td></tr>
<tr><td>(Aguilar et al. 2008)</td><td>model characteristics: gateway interplay/ control structures</td><td>number of looping sequence flows</td><td>number of looping sequence flows</td><td>correlation with comprehension efficiency (score/time ratio); direction unclear</td><td>correlation analysis (Spearman)</td><td>reached significance on a 0.05 level in at least two of five sub-samples
→ weak (direction of effect unclear)</td><td>-</td><td>110 partici-pants in 5 samples (large)</td><td>10</td></tr>
<tr><td>(Mendling and</td><td>model characteristics: gateway</td><td>concurrency</td><td>sums up all concurrent threads that can be activated</td><td>correlation not significant (but negative)</td><td>correlation analysis (Spearman)</td><td>not significant
→ no evidence</td><td>-</td><td>42 (small)</td><td>6</td></tr>
</table>

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Models
						Objective	Subjective		
Strembeck 2008)	interplay/ control structures		by AND-splits and OR-splits ("token split")						
Melcher and Seese 2008)	model characteristics: gateway interplay/ control structures	order, repetition, concurrency, exclusiveness (wording of comprehension task)	• group A: questions about order and repetition • group B: questions about concurrency and exclusiveness	Order (μ=0.48, SD=0.24) had lower comprehension accuracy than repetition (μ=0.93, SD=0.10), concurrency (μ=0.88, SD=0.23), or exclusiveness (μ=0.90, SD=0.20).	Wilcoxon rank-sum test for independent values and for paired values	$p<0.05$ → strong for order versus repetition/concurrency/exclusiveness → no evidence for other comparisons	-	18 (small)	1
(Melcher et al. 2010)	model characteristics: gateway interplay/ control structures	order, repetition, concurrency, exclusiveness (wording of comprehension task)	between-subject design, each group got thirty-three (different) comprehension tasks on one aspect only (276 different questions and some dummy questions)	Order (estimated μ=0.58, SD=0.16) had lowest comprehension accuracy, repetition (estimated μ=0.95, SD=0.15) and exclusiveness (estimated μ=0.93, SD=0.11) were easiest, and concurrency (estimated μ=0.86, SD=0.14) was in the middle. (Order was the only normal distributed variable.)	Wilcoxon-rank-sum test for independent variables ("virtual subjects procedure"; subjects with similar personal partial process comprehension accuracy are combined to a new virtual subject)	significant difference for all pair comparisons on the α=0.05 level, with the exception of repetition and exclusiveness → moderate	-	178 partici-pants in 9 samples (very large)	1
(Figl and Laue 2011)	model characteristics: gateway interplay/ control structures	order, repetition, concurrency, exclusiveness (wording of comprehension task)	two item wordings per type of comprehension question (concurrency, exclusiveness, order, repetition) on pairs of activities, which are either close (1 activity between them) or distant (> 1 activity between them)	• Order was the easiest relationship (80% correct answers) with the lowest subjective difficulty (3.08), followed by concurrency (83%, 3.20). • Exclusiveness was the most difficult relationship concerning comprehension accuracy (70%, 3.19), and repetition was rated as the most difficult by participants (71%, 3.58).	ANCOVA	comprehension accuracy: only trend ($F_{3,55}$ = 2.65, p = 0.058) → weak	perceived difficulty ($F_{3,55}$ = 4.20, p = 0.010) → strong	199 (very large)	4
(Laue and Gadatsch 2011)	model characteristics: gateway interplay/ control structures	order, concurrency, exclusiveness (wording of comprehension task)	construction of comprehension tasks (wordings of order 1 and order 2 differ)	Comprehension accuracy (percentage of correct answers) was given, but there was no information on statistical difference between types of question.	descriptive statistics only	• order (style 1: 90%, style 2: 39%) • exclusiveness (style 1: 79%, style 2: 80%) • concurrency (style 1: 72%, style 2: 59%) • forward dependency/response: 75%, • backward dependency/ precedence: 85% →weak	-	22 (small)	1
(Weitlaner et al. 2013)	model characteristics: gateway interplay/ control structures	order, repetition, concurrency (wording of comprehension task)	construction of comprehension tasks (24 points per category, 72 points total)	Comprehension accuracy was higher for order and repetition than it was for concurrency.	descriptive statistics only	no numbers → weak	-	77 (large)	4

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
(Sarshar and Loos 2005)	model characteristics: gateway interplay/ control structures	questions for AND, XOR, OR, and multilevel AND/XOR situations (wording of comprehension task)	construction of comprehension tasks (details missing)	• EPC multi-level AND/XOR-gateways had similar comprehension accuracy as AND-gateways and XOR-gateways. • OR-gateways had lower comprehension accuracy than AND-gateways and XOR-gateways.	information missing	no numbers →weak	no numbers →weak	50 (medium)	1
(Figl and Laue 2015)	model characteristics: gateway interplay/ control structures	control-flow pattern (order/sequence, concurrency/AND, exclusiveness/XOR, repetition/loop, compound) (measured for each comprehension task)	based on a consensus-building rating approach, the authors determined which control-flow patterns had to be considered to perform each deductive-reasoning task	• There was a significant impact of different control-flow patterns on comprehension accuracy. (Tasks were most difficult if they demanded the reader understand repetition, followed by compound control-flow patterns (a combination of at least two patterns other than order), concurrency, and exclusiveness. Tasks for which only the control-flow pattern order had to be understood had the highest comprehension accuracy. • Trend-wise effect on subjective difficulty: Compound patterns were more difficult than order, concurrency, or exclusiveness. Repetition was more difficult than order.	ANCOVA	comprehension accuracy ($F_{1,53}=3.19$, p=0.02) → strong	subjective comprehension difficulty ($F_{1,53}=1.99$, p=0.11) → weak	155 (very large)	4
(Reijers and Mendling 2011)	model characteristics: connection	average gateway degree	the average of the number of both incoming and outgoing arcs of the gateway nodes (Reijers and Mendling 2011)	The effect on comprehension accuracy was negative.	correlation analysis (Pearson)	r=-0.67, p=0.02 → strong	-	73 (large)	12
(Sánchez-González et al. 2012)	model characteristics: connection	average gateway degree	the average of the number of both incoming and outgoing arcs of the gateway nodes (Reijers and Mendling 2011)	fairly easy ≤ 3.67 ≤ easy ≤ 3.83 ≤ moderately difficult ≤ 4.06 ≤ difficult ≤ 4.18 ≤ fairly difficult	ANOVA tests (significant) between models were used to identify “thresholds” between different models.	no numbers → weak	-	28 for obtaining thresholds, 23 for cross-validation (medium)	10
(Sánchez-González et al. 2012)	model characteristics: connection	maximum gateway degree	maximum number of incoming and outgoing arcs of a decision node	fairly easy ≤ 4 ≤ easy ≤ 5 ≤ moderately difficult ≤ 7 ≤ difficult ≤ 9 ≤ fairly difficult	ANOVA tests between models were used to identify “thresholds” between different models.	significant, but no numbers → weak	-	28 for obtaining thresholds, 23 for cross-validation (medium)	10
(Reijers and Mendling 2011)	model characteristics: connection	extent to which all the nodes in a model are connected to each other	the extent to which all the nodes in a model are connected to each other (Reijers and Mendling 2011) [“cross-connectivity”]	no significant correlation	correlation analysis (Pearson)	not significant → no evidence	-	73 (large)	12

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Models
						Objective	Subjective		
(Sánchez-González et al. 2010)	model characteristics: connection	ratio of arcs to nodes	ratio of the total number of arcs in a process model to its total number of nodes [“coefficient of connectivity“]	A higher coefficient of connectivity is negatively related to model comprehension.	correlation and regression analysis	no numbers → weak (direction of effect unclear)	-	71 (large)	15
(Reijers and Mendling 2011)	model characteristics: connection	ratio of arcs to theoretically maximum number of arcs	“ratio of the total number of arcs in a process model to the theoretically maximum number of arcs (i.e., when all nodes are directly connected)” (Reijers and Mendling 2011, p. 5) [“density“]	Higher density is negatively related to model comprehension.	correlation analysis (Pearson)	r=-0.62, p=0.03 → strong	-	73 (large)	12
(Mendling and Strembeck 2008)	model characteristics: syntax rules	soundness	soundness criteria based on (Mendling and van der Aalst 2006); for instance, soundness can be violated by the incorrect insertion of OR-joins	Correlation is not significant (but positive).	correlation analysis (Spearman)	not significant → no evidence	-	42 (small)	6
(Heggset et al. 2015)	model characteristics: syntax rules	models before and after revising syntactic quality according to a guideline	syntactic quality according to a guideline (e.g., naming conventions, allowed symbols)	• no significant difference in the first model which included only a few improvements • significant difference in the second model: higher comprehension accuracy for improved model version	descriptive statistics only	• students: 69% (old) versus 90% (improved model) • employees: 55% (old) versus 85% (improved model) → weak	-	18 (small)	1
User									
(Recker and Dreiling 2007)	user: domain knowledge	domain knowledge	perceived knowledge of domain (no further details found)	There was no significant effect on comprehension accuracy, perceived ease of use scale, or time taken.	ANCOVA	not significant → no evidence	not significant → no evidence	69 (medium)	2
(Bera 2012)	user: domain knowledge	domain knowledge	two items (experience in the last 2 years, extent of knowledge), 7-point scale	no significant effect	ANCOVA	not significant → no evidence	-	51 (medium)	2
(Recker et al. 2014)	user: domain knowledge	domain knowledge	adapted from Burton-Jones and Meso (2008), median-split groups (high-low)	no significant effect	ANCOVA/ regression analysis (hierarchical)	not significant → no evidence	-	92 (large)	2
(Turetken et al. 2016)	user: domain knowledge	domain knowledge	self-rated familiarity with process and domain (26 of 60 participants were domain experts who had taken part in the execution of the process)	no significant effect	unknown	not significant → no evidence	not significant → no evidence	60 (medium)	2
(Stitzlein et al. 2013)	user: domain knowledge	professionals of different domains	healthcare or engineering professionals	• There was no overall significant effect on comprehension accuracy or on complex items. • However, for simple items, healthcare workers understood both models (BPMN, health process	ANOVA	• simple tasks: interaction effect of notation and professional background $F_{1,12}$=6.37, p=0.03	-	16 (small)	2

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
				notation (HPN)) equally well, while engineers understood BPMN better than HPN.		- comprehension accuracy: overall effect not significant → no evidence			
(Ottensooser et al. 2012)	user: experience and familiarity with modeling	frequency of use of flow charts	5-point scale	Participants who work with flowcharts more often perform better.	regression analysis (OLS, t-values above 1.96 are significant for a two-tailed test)	- participants at the higher end of the scale (coefficient: 0.27, standard error=0.16, t=2.31) - participants at the lower end of the scale (coefficient: -0.09, standard error=0.14, t=-0.64) → moderate	-	196 (very large)	6
(Mendling et al. 2012b)	user: experience and familiarity with modeling	modeling expertise	how long they have been involved with the process, measured on four levels: less than one month, less than a year, less than three years, and longer than three years	- Inverse u-shaped curve: Participants with medium experience (less than three years) seem to have better comprehension accuracy than participants with low experience (less than a month) and those with high experience (longer than 3 years). - There was no significant effect on time taken.	Kruskal-Wallis test	- comprehension accuracy: X^2=24.48, p=0.00 → moderate	-	113 (very large)	6
(Figl et al. 2013a)	user: experience and familiarity with modeling	training on modeling basics	training on modeling basics at a university or school (Yes/No)	There was a significant influence of modeling basics training on comprehension accuracy and subjective cognitive load.	ANCOVA	- small model: F= 3.98, p= 0.05 - large model: F=4.34, p= 0.04 → strong	subjective cognitive load: - small model (F=4.24, p= 0.04) - large model: not significant → moderate	136 (very large)	3
(Reijers and Mendling 2011)	user: experience and familiarity with modeling	educational background	undergraduate versus graduate students with greater knowledge (and more hours spent on modeling) on workflow concepts (e.g., model soundness)	Eindhoven students had higher comprehension accuracy than students of Vienna and Madeira did.	Kruskal-Wallis test	no numbers → moderate	-	73 (large)	12
(Recker and Dreiling 2011)	user: experience and familiarity with modeling	modeling experience with EPC (number of models created or read)	median-split groups (high-low) based on the number of process models created or read	Modeling experience has a positive effect on - inferential transfer ability test scores (deep understanding; but not on model-based answers) and - time taken (retention ability test, partial effect on claims handling task). There was no significant effect on: - comprehension scores (retention test scores) or - time taken (transfer ability test scores).	MANCOVA	comprehension scores (retention test scores): not significant [inferential transfer ability test scores: - goods receipt task: $F_{2,11}$=3.23, p=0.05 - claims handling task: $F_{2,11}$=3.25, p=0.05] retention ability task completion times: - claims handling task: $F_{2,11}$=3.16, p=0.05] → no evidence	-	69 (medium)	2

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
(Mendling and Strembeck 2008)	user: experience and familiarity with modeling	duration of involvement with business process modeling	four levels: less than one month, less than a year, less than three years, and longer than three years	no significant correlation (but positive)	correlation analysis (Pearson)	not significant → no evidence	-	42 (small)	6
(Mendling and Strembeck 2008)	user: experience and familiarity with modeling	intensity of work with process models (how often they work with process models)	four levels: daily, monthly, less frequent than monthly, never	no significant correlation (but positive)	correlation analysis (Pearson)	not significant → no evidence	-	42 (small)	6
(Mendling et al. 2012b)	user: experience and familiarity with modeling	modeling intensity	how often they work with process models: four options—daily, monthly, less frequent than monthly, never	• Inverse u-shaped curve: Participants with medium modeling intensity (monthly) seem to achieve better comprehension accuracy than do participants with low modeling intensity (never, less than monthly) and those with high modeling intensity (daily). • Users with higher modeling intensity took less time.	Kruskal-Wallis test	• comprehension accuracy: not significant • [time taken: X^2=9.09, p=0.03] → no evidence	-	113 (very large)	6
(Reijers and Mendling 2011)	user: experience and familiarity with modeling	modeling experience	self-assessment of process modeling experience on a 4-point ordinal scale	no significant effect	Kruskal-Wallis test	not significant → no evidence	-	73 (large)	12
(Turetken et al. 2016)	user: experience and familiarity with modeling	experience and knowledge in process modeling and BPMN	self-rated experience and knowledge in process modeling and BPMN	no significant effect	unknown	not significant → no evidence	not significant → no evidence	60 (medium)	2
(Recker and Dreiling 2011)	user: experience and familiarity with modeling	working experience with business process modeling	yes/no	Work experience has a positive effect on: • inferential transfer ability test scores • time taken (for retention ability test) There was no significant effect on: • comprehension scores (model-based retention test scores) • time taken (for transfer ability test)	MANCOVA	comprehension scores (model-based retention test scores): not significant [transfer ability test scores: • goods receipt task: $F_{2,11}$=3.75, p=0.03, retention ability task completion times: • goods receipt task: $F_{2,11}$=6.80, p=0.00 • claims handling task: $F_{2,11}$=8.47, p=0.00] → no evidence	-	69 (medium)	2
(Recker 2013)	user: experience and familiarity	self-reported familiarity of the BPMN grammar	validated three-item grammar familiarity scale from (Recker 2010)	no significant effect	t-test (based on median-split variable)	not significant → no evidence	-	98 (large)	2*3 models

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Mod els
						Objective	Subjective		
	with modeling								
(Kummer et al. 2016)	user: experience and familiarity with modeling	self-reported familiarity of the BPMN grammar	validated three-item grammar familiarity scale from (Recker 2010)	There was no significant effect on either comprehension accuracy or perceived difficulty.	MANCOVA	no numbers → no evidence	no numbers → no evidence	127 (very large)	2
(Ottensooser et al. 2012)	user: experience and familiarity with modeling	comfort with flowcharts	5-point scale for each notation	no significant effect	regression analysis (OLS, t-values above 1.96 are significant for a two-tailed test)	not significant → no evidence	-	196 (very large)	6
(Recker and Dreiling 2007)	user: experience and familiarity with modeling	level of EPC competency	familiarity, confidence and competence with EPC	no significant effect	ANCOVA	not significant → no evidence	-	69 (medium)	2
(Johannsen et al. 2014a)	user: experience and familiarity with modeling	personal factors	questions on modeling expertise, theoretical knowledge concerning modeling with process models, domain knowledge (questions were aggregated because many of the questions had no explanatory power on their own)	no significant effect	regression analysis	not significant → no evidence	-	51-53 (medium)	1
(Weitlaner et al. 2013)	user: experience and familiarity with modeling	self-assessed previous knowledge on BPM	participants' theoretical and practical elementary knowledge on BPM on a subjective basis, measured on a 5-point Likert scale	no significant correlation	correlation analysis (Pearson)	not significant → no evidence	-	77 (large)	4
(Reijers and Mendling 2011)	user: experience and familiarity with modeling	self-assessed theoretical knowledge on process modeling	self-assessment on a 5-point ordinal scale	no significant effect	Kruskal-Wallis test	not significant → no evidence	-	73 (large)	12
(Figl et al. 2013a)	user: modeling knowledge	process modeling test knowledge	process modeling test scores (based on Mendling et al. 2012b)	For selected tasks prior knowledge about process modeling had significant influence on subjective cognitive load but not on comprehension accuracy.	ANCOVA	not significant → no evidence	subjective cognitive load: • large model: F=3.65, p= 0.06 • small model: not significant	136 (very large)	3

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
							• text-model comparison: not significant → moderate		
(Recker et al. 2014)	user: modeling knowledge	process modeling test knowledge	adapted from Mendling et al. (2012b), median-split groups (high-low)	no significant effect	ANCOVA/ regression analysis (hierarchical)	not significant → no evidence	-	92 (large)	2
(Figl et al. 2013b)	user: modeling knowledge	process modeling test knowledge	process modeling test scores (based on Mendling et al. 2012b)	Process modeling knowledge improves comprehension accuracy, subjective cognitive load, and subjective difficulty of control-flow comprehension.	regression analysis	$\beta=0.14$, $p \leq 0.001$ → strong	• perceived cognitive load ($\beta=0.14$, $p \leq 0.001$) • perceived control flow comprehension ($\beta=0.14$, $p \leq 0.001$) → strong	154 (very large)	4
(Kummer et al. 2016)	user: modeling knowledge	process modeling test knowledge	process modeling test scores (based on Mendling et al. 2012b)	• There was a significant effect on model comprehension in the first model and a trendwise effect in the second model. • There was no significant effect on perceived difficulty.	MANCOVA	• Model 1: $F_{1,115}=12.11$, $p<0.001$ • Model 2: $F_{1,115}=3.17$, $p<0.1$ → moderate	not significant → no evidence	127 (very large)	2
(Figl and Laue 2015)	user: modeling knowledge	process modeling test knowledge (low, high)	process modeling test scores (based on Mendling et al. 2012b), median-split groups (low, high)	Participants with lower modeling knowledge had lower comprehension accuracy and rated tasks as more difficult than did participants with higher process-modeling knowledge.	ANCOVA	$F_{1,53}=8.05$, $p=0.006$ → strong	subjective difficulty ($F_{1,53}=22.58$, $p<0.001$) → strong	156 (very large)	4
(Recker 2013)	user: modeling knowledge	process modeling test knowledge	ten true/false questions on control flow logic (reachability, deadlocks, liveness, and option to complete), (based on Mendling et al. 2012b) ("knowledge of control flow logic")	Control flow knowledge was a significant covariate for explaining comprehension accuracy but not for explaining time taken.	ANCOVA (repeated measures)	$F=9.12$, $p=0.00$, $\eta^2=0.09$ → strong	-	98 (large)	2*3 models
(Figl and Strembeck 2015)	user: modeling knowledge	process modeling test knowledge	process modeling test scores (based on Mendling et al. 2012b)	There was a positive effect on comprehension accuracy, but no significant effect on either perceived ease of use or time taken	ANCOVA	$F_{1,36}=27.64$, $p<0.001$ → strong	not significant → no evidence	44 (medium)	2
(Mendling and Strembeck 2008)	user: modeling knowledge	process modeling test knowledge	six yes/no questions on questions on, for example, choices, concurrency, loops, and deadlocks; prior version of process modeling test (based on Mendling et al. 2012b)	There was a positive relationship with comprehension accuracy.	correlation analysis (Pearson)	$r=0.49$, $p=0.01$ → strong	-	42 (small)	6
(Dumas et al., 2012)	user: modeling knowledge	theoretical knowledge on process modeling	no details found in the paper	There was no overall significant effect, but there was a positive effect on comprehension accuracy for four out of eight models.	ANOVA	effect size = 0.355-0.454, p = 0.005-0.045 → moderate	-	55 (medium)	8

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
(Figl and Recker 2016)	user: modeling knowledge	conceptual modeling knowledge	conceptual modeling test (recognizing different notations test)	Knowledge on conceptual modeling heightens the preference for process model representation for comprehension tasks (more specifically, for diagrams over structured text, although there was no statistically significant effect for diagrams over text).	multiple regression analysis	[empirical article on subjective comprehension and user preferences]	diagrams over structured text: $B=4.86$, $\beta=0.21$ ($p \leq 0.05$) diagrams over text: not significant → moderate	120 (very large)	1
(Weitlaner et al. 2013)	user: education	level of education	three levels of education: - apprenticeship - graduates/high-school - graduates/academics	Academics and high-school graduates performed better (higher comprehension accuracy) than apprenticeship graduates did.	ANOVA (one-way), Tukey post-hoc tests	$F_{2,74}=4.47$, $p=0.015$ (across all processes) → strong	-	77 (large)	4
(Döhring et al. 2014)	user: education	professional level	professional level: - senior-level: post-docs and industry employees - student level: students up to PhD	There was no significant effect on comprehension accuracy, perhaps because of a relatively coarse-granular level of measurement, nor on easiness or time taken.	Mann-Whitney U-test, Wilcoxon rank-sum test, t-tests (paired), ANOVA	not significant → no evidence	not significant → no evidence	14 (small)	2
(Recker and Dreiling 2011)	user: user characteristics	English as a second language (ESL)	English as a second language (ESL): native English speakers and European and Asian English speakers	Native speakers performed better on: - model-based transfer ability test scores and - time taken (retention ability test). (Interpretation: textual semantics are difficult to understand for ESL speakers.) There was no significant effect on: - comprehension scores (inferential transfer ability test scores, retention test scores) or - task completion times (transfer ability test scores).	MANCOVA	comprehension scores (inferential transfer ability test scores): not significant [model-based transfer ability test scores: - goods receipt task $F_{2,11}=8.68$, $p<0.001$ - claims handling task $F_{2,11}=10.74$, $p<0.01$ retention ability task completion times: - goods receipt task $F_{2,11}=4.30$, $p=0.02$ - claims handling task $F_{2,11}=8.12$, $p=0.00$] → no evidence	-	69 (21 native speakers versus 47 non-natives) (medium)	2
(Kummer et al. 2016)	user: user characteristics	culture (Germanic – Germany and Austria; Confucian – China)	experiment was conducted in China, Germany, and Austria; country of origin	- Process models were perceived to be more difficult in the Germanic culture than they are in the Confucian culture. - There was no difference in comprehension accuracy.	MANCOVA	not significant → no evidence	- Model 1: $F_{1,115}=49.61$, $p<0.001$ - Model 2: $F_{1,115}=12.44$, $p<0.001$ → strong	127 (very large)	2
(Recker et al. 2014)	user: user characteristics	users' deep learning motive and strategy	revised learning process questionnaire (R-LPQ-2F) from Kember et al. (2004)	no significant effect	regression analysis (hierarchical)	not significant → no evidence	-	92 (large)	2

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Participants	# of Models
						Objective	Subjective		
(Recker et al. 2014)	user: user characteristics	users' surface learning motive and strategy	revised learning process questionnaire (R-LPQ-2F) from Kember et al. (2004)	A surface learning motive (indicating extrinsic instead of intrinsic motivation or low learning intensity) was negatively associated with comprehension accuracy, while surface learning strategy (indicating simple learning for memorization) was positively associated with comprehension accuracy.	regression analysis (hierarchical)	surface motive: • Model 1: $\beta = -0.34$, $p < 0.05$ • Model 2: $\beta = -0.28$, $p < 0.05$ surface strategy: • Model 1: $\beta = 0.29$, $p < 0.05$ • Model 2: $\beta = 0.32$, $p < 0.01$ → strong	-	92 (large)	2
(Recker et al. 2014)	user: user characteristics	sensing versus intuitive learning style	sensing versus intuitive learning scale from Felder and Spurlin (2005); two median-split groups	Users with a sensing learning style achieve higher process model comprehension accuracy.	ANCOVA	no numbers → moderate	-	92 (large)	2
(Recker et al. 2014)	user: user characteristics	selection ability	"choosing a path" test (Ekstrom et al 1976)	There was a positive association with comprehension accuracy.	regression analysis (hierarchical)	• Model 1: $\beta = -0.25$, $p < 0.05$ • Model 2: $\beta = -0.32$, $p < 0.05$ → strong	-	92 (large)	2
(Recker et al. 2014)	user: user characteristics	self-efficacy	adapted from Phillips and Gully (1997); median-split groups (high-low)	no significant effect	ANCOVA/ regression analysis (hierarchical)	not significant → no evidence	-	92 (large)	2
(Recker et al. 2014)	user: user characteristics	abstraction ability	abstract reasoning: "thinking in figures" test (de Wit and Compaan 2005)	There was a negative association with comprehension accuracy.	regression analysis (hierarchical)	• Model 1: $\beta = 0.56$, $p < 0.001$ • Model 2: $\beta = 0.61$, $p < 0.001$ → strong	-	92 (large)	2
(Figl and Recker 2016)	user: user characteristics	Participants' cognitive style (verbal, spatial visual, object visual)	measured with an instrument from (Blazhenkova and Kozhevnikov 2008)	There were cognitive style effects on the preference for process model representation for comprehension tasks: • Spatial style heightens the preference for diagrams over text. • Verbal style lowers the preference for diagrams over structured text and heightens the preference for structured text over text. • Object style has no significant effect.	multiple regression analysis	[empirical article on subjective comprehension and user preferences]	• Spatial style heightens the preference for diagrams over text ($B=7.35$, $\beta=0.27$, $p \leq 0.05$). • Verbal style lowers the preference for diagrams over structured text ($B=-10.64$, $\beta=-0.22$, $p \leq 0.05$). • Verbal style heightens the preference for structured text over text ($B=8.30$, $\beta=0.17$, $p \leq 0.05$). • object style: not significant → moderate (for spatial style and verbal style)	120 (very large)	1

Source Reference	Category	Independent Variable	Measurement of Independent Variable	Results	Analysis Method	Statistical Significance and Level of Evidence of Effect on Comprehension		# of Partici-pants	# of Models
						Objective	Subjective		
							→ no evidence (for object style)		
Task									
(Soffer et al. 2015)	task	provision of a catalog of routing possibilities or a catalog of workflow patterns	manipulated factor constituting experimental groups	A catalog of routing possibilities increased comprehension accuracy for the three cases that were available only in the catalog and made no significant difference in the other two cases that were part of both catalogs, compared to the catalog of workflow pattern.	Mann-Whitney test	p=0.017 → moderate	-	54 (medium)	5
(Pichler et al. 2012)	task	sequential and circumstantial tasks	sequential tasks (how input conditions lead to a certain outcome; e.g., “Activity X must be directly preceded by activity Y”) and four circumstantial tasks (what (combination of) circumstances will cause/lead to/follow from a particular outcome: “If activity X or Y has been executed, is it possible to terminate a process instance by executing at least one additional activity?)	There was better comprehension accuracy and lower time taken for sequential than for circumstantial tasks.	Sheirer-Ray-Hare test	• comprehension accuracy: p=0.02 (for imperative models) • [time taken: p=0.01 (both for imperative and declarative models)] → strong	-	27 (small)	4
(Laue and Gadatsch 2011)	task	question wording (style 1 and style 2 questions)	manipulated factor constituting experimental groups	exact wording of items for concurrency, exclusivity, and order influence comprehension accuracy	Pearson's Chi-Square test	level of significance 0.001 → strong	-	22 (small)	1
(Figl and Laue 2015)	task	validity of conclusion (valid, wrong)	manipulated factor constituting experimental groups	• Valid reasoning tasks were more difficult to answer than invalid/wrong tasks. • Concerning subjective difficulty, descriptive results pointed in the same unanticipated direction, although they were not significant.	ANCOVA	comprehension accuracy $F_{1,53}$=7.53, p=0.008) → strong	not significant → no evidence	155 (very large)	4

Appendix D: “Theoretical” Discussions versus Empirical Studies

Table 6 Comparison of influence factors for process model comprehension in theoretical discussions and empirical studies

Category of Independent Variables		**Theoretical Discussions**	**Empirical Studies (measuring only subjective comprehension [s] or only preferences/perceived usefulness [p] or measuring objective comprehension [o])**
Presentation Medium			
			• paper versus online questionnaire (Mendling et al. 2012b; Recker et al. 2014) [o] • paper versus interactive web-based visualizations (Turetken et al. 2016) [o]
Notation			
representation paradigm	representation and cognitive fit	• text versus model (graphical notation) (Rosa et al. 2011) • cognitive fit (e.g., different dialects, representational media) (Genon et al. 2010)	• BPMN notation (graphical) versus written use cases (textual) (Ottensooser et al. 2012) [o] • BPMN versus textual process description modeling notation (Rodrigues et al. 2015) [o] • process model representation (text, structured text, diagram (BPMN)) (Figl and Recker 2016) [p]
	declarative versus imperative	• declarative process models versus imperative process models (better comprehension of sequential information in imperative models; better comprehension of circumstantial information in declarative models) (Fahland et al. 2009)	• declarative versus imperative models (Pichler et al. 2012) [o]
	icons	• assigning pictorial elements (e.g., icons, images) to modeling elements (La Rosa et al. 2011) • domain semantic-oriented icons added to labels (Mendling and Recker 2008; Mendling et al. 2010a)	• use of icons (with/without icons) (Figl and Recker 2016) [p]
	animation	• animation and visualization techniques for 2D and 3D models (embedding perspectives, animating control flow, animation scenarios based on event logs) (Aysolmaz and Reijers 2016) • use of narration for process animation (Aysolmaz and Reijers 2016)	• new visualizations (bubble visualization, BPMN3D, network visualization concepts, thin line concept) (Hipp et al. 2014) [s]
primary notation		• BPMN, UML AD, BPMN, YAWL, EPCs, Petri nets (Figl et al. 2009) • BPMN (Genon et al. 2010), (Leopold et al. 2016) • routing symbol design (BPMN, UML AD, BPMN, YAWL, EPCs) (Figl et al. 2010)	• C-YAWL and vBPMN (Döhring et al. 2014) [o] • EPC versus Petrinet notation (Sarshar and Loos 2005) [o] • configurable EPCs (C-EPCs) versus EPC (Recker et al. 2005) [p] • EPC versus BPMN (Recker and Dreiling 2007, 2011) [o] • UML, BPMN, EPC, SBD (storyboard design) (Weitlaner et al. 2013) [o] • flow diagrams, eEPC, UML AD, BPMN (Sandkuhl and Wiebring 2015) [o] • BPMN, deontic BPMN (Natschläger 2011)[o] • BPMN, UML AD, EPC (Jošt et al. 2016) [o] • BPMN, health process notation (HPN) (Stitzlein et al. 2013) [o] • high communication-flow orientation versus low communication-flow orientation (Kock et al. 2008; Kock et al. 2009) [s]
notational characteristics	semiotic clarity	• semiotic clarity (Figl et al. 2009; Genon et al. 2010; Figl et al. 2010) • avoidance of implicit splits and joins (without gateway constructs) (Leopold et al. 2016)	• semiotic clarity deficiencies (Figl et al. 2013a) [o]
	perceptual discriminability	• perceptual discriminability (Figl et al. 2009; Genon et al. 2010; Figl et al. 2010)	• perceptual discriminability (Figl et al. 2013a; Figl et al. 2013b) [o] • with/without BPMN gateway constructs, perceptual discriminability effect through use of gateways (Recker, 2013) [o]
	semantic transparency	• semantic transparency (Figl et al. 2009; Genon et al. 2010; Figl et al. 2010) • low semantic transparency of throwing message events (Leopold et al. 2016)	• semantic transparency (Figl et al. 2013b) [o]

Category of Independent Variables		**Theoretical Discussions**	**Empirical Studies (measuring only subjective comprehension [s] or only preferences/perceived usefulness [p] or measuring objective comprehension [o])**
	visual expressiveness	• visual expressiveness (Figl et al. 2009; Genon et al. 2010; Figl et al. 2010)	• perceptual pop-out of symbols (Figl et al. 2013b) [o] • symbol aesthetics (Figl et al. 2013b) [o]
	graphic economy	• restricting the syntax and semantics of a notation (creating a notation subset, e.g., for novices) (Rosa et al. 2011) • graphic economy (Figl et al. 2009; Genon et al. 2010; Figl et al. 2010) • extension of the syntax and semantics of a notation (e.g., to fit a domain-specific audience) (Rosa et al. 2011)	
			• ease of generating the models in a notation (Kock et al. 2009) [s]
Secondary Notation			
decomposition	use of decomposition/modularization	• complexity management (modularization, hierarchic structuring) (Genon et al. 2010) • decomposition in general (Becker et al. 1995; La Rosa et al. 2011) • decomposition/modularity (Storch et al. 2013) • “decompose a model with more than 50 elements” (Mendling et al. 2010b, p. 131) • “decompose a model with more than 31 nodes” (Mendling et al. 2012a, p. 1195) • modularity (e.g., fan-in/fan-out metrics “can indicate poor modularization”) (Azim et al. 2008; Gruhn and Laue 2006b)	• modularization (model with sub-processes, flattened model without sub-processes) (Reijers et al. 2011b; Turetken et al. 2016) [o]
	decomposition heuristics	• decomposition heuristics (breakpoint heuristic: decompose processes at breakpoints (e.g., sub-goals achieved, different themes); data objects heuristic: do not split activities that share objects; role heuristic (decomposition based on who is performing the activity); shared processes heuristic (redundant process fragments are modeled in one sub-process and called upon); repetition heuristic (based on frequency of process); structuredness heuristic (use single entry-single exit fragments to identify candidates for sub-processes)) (Milani et al. 2016) • decomposition heuristics (minimality, determinism, freedom of losslessness, minimum coupling, strong cohesion violated) (Johannsen et al. 2014b) • vertical (hierarchical, hiding in sub-processes), horizontal (partitioning in peer processes), and orthogonal (along crosscutting concerns) modularization (La Rosa et al. 2011) • decomposition/modularity (trade-off between abstraction, which heightens comprehension, and split-attention effect, which lowers comprehension) (Zugal et al. 2012) • constructing a consolidated model from disjointed models or merging similar process models (a family of process models) into a single process model (La Rosa et al. 2011)	• decomposition: full compliance versus moderate violation (minimality, determinism, losslessness violated) versus strong violation (minimality, determinism, losslessness, minimum coupling, strong cohesion violated) (Johannsen et al. 2014a) [o]
Gestalt theory		• change/use of a visual variable (e.g., line thickness, color) to highlight elements with a shared property (e.g. all functions in green) (La Rosa et al. 2011) • dual coding (alignment of symbols and text) (Genon et al. 2010) • textual annotations (in free-form text) (La Rosa et al. 2011) • close placement of related elements (La Rosa et al. 2011) • visual enclosure to highlight elements with a shared property (e.g., those that need revision) (La Rosa et al. 2011)	• syntax highlighting with color (Reijers et al. 2011a) [o] • color for different symbols (Kummer et al. 2016) [o] • task-specific color highlighting (Petrusel et al. 2016) [o] • with/without swim lanes (Bera 2012; Jeyaraj and Sauter 2014) [o]
layout	edges	• edges (length, broken/simple/curved edges, crossing edges, placement of text on edges, ending points, angles) (Bernstein and Soffer 2015b) • angles between edges (Bernstein and Soffer 2015a) • control flow and message flow arcs should be consistent (Leopold et al. 2016) • avoidance of crossing edges (La Rosa et al. 2011) • avoidance of overlapping edges and nodes (Leopold et al. 2016) • line crossings (Schrepfer et al. 2009)	
	direction	• flow direction (e.g., left-to-right, top-to-bottom) (La Rosa et al. 2011)	• flow direction (Figl and Strembeck 2015) [o]

Category of Independent Variables		**Theoretical Discussions**	**Empirical Studies (measuring only subjective comprehension [s] or only preferences/perceived usefulness [p] or measuring objective comprehension [o])**
		• left-to-right orientation is superior to other flow directions (Figl and Strembeck 2014) • model's direction (general direction, placement of ending events, branching off, change in direction) (Bernstein and Soffer 2015b) • modeling direction (consistent left-to-right orientation) (Leopold et al. 2016)	
	shape and size	• model's structure (shape "e.g., horizontally or vertically, rectangular, square" and size) (Bernstein and Soffer 2015b, 2015a) • excessive diagram size (should fit on an A3 page) (Leopold et al. 2016)	
	symmetry	• symmetry in blocks (Bernstein and Soffer 2015b) • symmetry (Schrepfer et al. 2009) • "neat and tidy" layout instead of "chaotic and cluttered" (La Rosa et al. 2011) • placing incoming and outgoing edges on opposite sides (La Rosa et al. 2011)	
	alignment of elements and spacing	• alignment of elements in the model (Bernstein and Soffer 2015b) • alignment of elements to visualize their characteristics (e.g., tasks with higher frequency on the left) (Becker et al. 1995) • inappropriate spacing (distance between connected elements should be at least 50% of the element size) (Leopold et al. 2016) • use of locality (close placement of related elements) (Schrepfer et al. 2009) • elements' orientation, alignment, and spacing (La Rosa et al. 2011)	• task-specific layout (size of elements and alignment) (Petrusel et al. 2016) [o]
	ending points	• number of ending points (Bernstein and Soffer 2015a)	
Label			
label design		• label design (e.g., "lowercase usage of letters, sans-serif, non-bold fonts," optimal average length hypothesized as five to eight letters per word; constant symbol size, symbol form, and font size; left-alignment; "usage of high levels of contrast for font/background colors"; spatially close placement of labels to corresponding symbols; linguistic-based segmentation of words) (Koschmider et al. 2015a) • label quality (label style and length) (Fettke et al. 2012)	• string length of all textual activity labels (Mendling and Strembeck 2008) [o]
naming conventions	syntactic naming conventions	• syntactic naming conventions (e.g., use of a verb-object label style) (Fettke et al. 2012; La Rosa et al. 2011; Mendling 2013; Mendling et al. 2010a) • inconsistent labels (violations of labeling conventions) (Overhage et al. 2012) • naming conventions, grammatical structure (Fettke et al. 2012) • consistent labeling (object-participle-question style for gateways, verb-object style for activities, object-participle style for events) ((Leopold et al. 2016)	• label styles (Mendling et al. 2010c) [p]
	semantic naming conventions	• semantic naming conventions (e.g., using a domain-specific vocabulary) (La Rosa et al. 2011) • non-normalized labels (e.g., synonymous labels in one process model) (Overhage et al. 2012) • naming conventions (consistency of terms, centralized terminology, e.g., based on glossary) (Fettke et al. 2012) • avoidance of synonyms and homonyms (Mendling 2013) • inconsistencies in activity label names and styles (Weber et al. 2011)	• abstract versus concrete labels (Mendling and Strembeck 2008) [o]; (Mendling et al. 2012b) [o]; (Figl and Strembeck 2015) [o] • perceived ambiguity of labels (Mendling et al. 2010c) [p] • revised labels from a linguistic perspective (Koschmider et al. 2015b) [s]
Model Characteristics			

Category of Independent Variables	**Theoretical Discussions**	**Empirical Studies (measuring only subjective comprehension [s] or only preferences/perceived usefulness [p] or measuring objective comprehension [o])**
size measures	• size (Becker et al. 1995; Rosa et al. 2011; Storch et al. 2013) • “use as few elements in the model as possible” (Mendling et al. 2010b, p. 131) • “do not use more than 31 [nodes]” (Mendling et al. 2012a, p. 1195) • number of activities (Azim et al. 2008; Gruhn and Laue 2006b) • omitting elements or collapsing elements into one to decrease size (Rosa et al. 2011) • “use one start and one end event” (Mendling et al. 2010b, p. 131) • “use no more than 2 start and end events” (Mendling et al. 2012a, p. 1195)	• size (based on arcs and nodes) (Recker, 2013) [o] • diameter (Mendling and Strembeck 2008; Sánchez-González et al. 2010) [o] • number of nodes (Sánchez-González et al. 2010) [o] • number of events in the model (Aguilar et al. 2008) [o] • number of end message events (Aguilar et al. 2008) [o] • number of intermediate events in the model (Aguilar et al. 2008) [o] • number of intermediate message events (Aguilar et al. 2008) [o] • number of sequence flows from events (Aguilar et al. 2008) [o] • number of exclusive data-based decisions (Aguilar et al., 2008) [o] • number of gateways (Sánchez-González et al. 2012) [o] • number of OR joins (Reijers and Mendling 2011) [o]
modularity/structuredness	• block structuredness (corresponding split and join elements; sometimes duplication of repeating model elements is necessary to achieve this goal) (La Rosa et al. 2011; Storch et al. 2013) • “model as structured as possible” (Mendling et al. 2010b, p. 131; Mendling et al. 2012a) • follow patterns in which unstructured modeling enhances comprehension instead of reducing it (e.g., interruption of a sequence of activities or structured modeling would lead to deep nesting of the control flow; e.g., after a parallel execution, one task has to be repeated or structured modeling would lead to repetition of elements) (Gruhn and Laue 2007) • model structure (e.g., nesting depth) (Azim et al. 2008) • model structure (e.g., nesting depth, jump outs of a control structure) (Gruhn and Laue 2006b) • nesting depth (Gruhn and Laue 2006a; Storch et al. 2013)	• separability (Mendling and Strembeck 2008) [o] • structuredness (Dumas et al. 2012) [o] • structuredness of the process graph (Mendling and Strembeck 2008) [o] • maximum nesting depth (Sánchez-González et al. 2010) [o] • element interactivity (of the activities included in a comprehension task), measured by the PST distance (Figl and Laue 2015) [o] • sequentiality (Sánchez-González et al. 2010) [o]
refactoring	• pattern to reduce cognitive load (e.g., removing unnecessary OR gateways, e.g., as a loop entry is used for optional execution; moving redundant elements outside a control block) (Gruhn and Laue 2009) • “unnecessarily complex control-flow structures, which could be simplified without changing the models’ behavior” (Weber et al. 2011)	
redundant elements	• removing redundant or superfluous elements (Rosa et al. 2011) • rules for dealing with redundancies of process tasks (Becker et al. 1995) • redundant process fragments (Weber et al. 2011)	
gateway interplay/control structures	• cognitive weights for different control structures (Gruhn and Laue 2006a): 1(lowest): sequence, cancel activity; 2: cancel case, composite task, XOR-split with corresponding XOR-join; 3: “XOR-split (exactly one of ≥ 3 branches is chosen) with corresponding XOR-join”; 4: parallel split and synchronization; 6: multiple instances patterns; 7: multiple choice and synchronizing merge • control flow complexity (metrics based on McCabe’s cyclomatic number, based on Halstead or on McMabe) (Azim et al. 2008) • control flow complexity (control flow complexity (CFC) metric based on McCabe’s cyclomatic number) (Gruhn and Laue 2006b) • “minimize the heterogeneity of gateway types” (Mendling 2013) • “use design patterns to avoid mismatch” (Mendling 2013) • avoid cycles (Storch et al. 2013) • “avoid OR routing elements” (Mendling et al. 2010b, p. 131; Mendling et al. 2012a) • “minimize the level of concurrency” (Mendling 2013) • avoid anti-patterns (e.g., implicit termination without an explicit end state) (Gruhn and Laue 2006b)	• simple and complex model (Döhring et al. 2014) [o] • gateway heterogeneity (Mendling and Strembeck 2008; Reijers and Mendling 2011; Sánchez-González et al. 2010; Sánchez-González et al. 2012) [o] • gateway mismatch (Reijers and Mendling 2011; Sánchez-González et al. 2010; Sánchez-González et al. 2012) [o] • control-flow complexity (Sánchez-González et al. 2012) [o] • number of sequence flows looping (Aguilar et al. 2008) [o] • concurrency (Mendling and Strembeck 2008) [o] • control flow structures (sequence versus loops versus concurrency (AND), XOR) (Figl and Laue 2011, 2015; Laue and Gadatsch 2011; Melcher et al. 2010; Melcher and Seese 2008; Sarshar and Loos 2005; Weitlaner et al. 2013) [o]

Category of Independent Variables	Theoretical Discussions	Empirical Studies (measuring only subjective comprehension [s] or only preferences/perceived usefulness [p] or measuring objective comprehension [o])
connection	• “minimize the routing paths per element” (Mendling et al. 2010b, p. 131) • “no more than 3 inputs or outputs per gateway” (Mendling et al. 2012a, p. 1195)	• average gateway degree (Reijers and Mendling 2011; Sánchez-González et al. 2012) [o] • maximum gateway degree (Sánchez-González et al. 2012) [o] • ratio of arcs to nodes (Sánchez-González et al. 2010) [o] • ratio of arcs to theoretically maximum number of arcs (Reijers and Mendling 2011) [o] • extent to which all the nodes in a model are connected (Reijers and Mendling 2011) [o]
syntax rules	• syntactically correct use of model elements (activities, gateways, events, flow gateways; e.g., attaching message flows to incorrect nodes) and avoidance of structural inconsistencies (e.g., presence of multi-merges and deadlocks, both of which may be caused by implicit splits and joins; inconsistent connections between main process and sub-process (e.g., different roles)) (Leopold et al. 2016)	• models before and after revising syntactic quality according to a guideline (Heggset et al. 2015) [o] • soundness – can, for instance, be violated by the incorrect insertion of OR-joins (Mendling and Strembeck 2008) [o]
Task		
	-	• question wording of tasks (Laue and Gadatsch 2011) [o] • validity of statements (Figl and Laue 2015) [o] • sequential tasks (how input conditions lead to a certain outcome) versus circumstantial tasks (what combination of circumstances will lead to a particular outcome) (Pichler et al. 2012) [o] • provision of a catalog of routing possibilities (Soffer et al. 2015) [o]
User		
tailoring for personal factors	• tailoring of process models for personal factors (Aysolmaz and Reijers 2016)	
domain knowledge	• domain knowledge (Schrepfer et al. 2009)	• domain knowledge (Bera 2012; Turetken et al. 2016; Recker and Dreiling 2007; Recker et al. 2014) [o]
experience and familiarity with modeling	• expert-novice differences (Genon et al. 2010) • modeler expertise (skills, knowledge, experience) (Schrepfer et al. 2009) • training and competence (Schrepfer et al. 2009) • modeling language expertise and familiarity (Schrepfer et al. 2009) • practical experience (Schrepfer et al. 2009) • students versus practitioners (Schrepfer et al. 2009) • time and intensity of being involved with process modeling (Schrepfer et al. 2009) • years of work experience at process consultancy (Schrepfer et al. 2009) • number of years of experience and deliberate practice (Schrepfer et al. 2009) • years of field experience in process consulting (Schrepfer et al. 2009) • estimated number of modeled processes (Schrepfer et al. 2009) • estimated average size of modeled processes (Schrepfer et al. 2009) • exceptional performance (training, experience, and talent) (Schrepfer et al. 2009) • expert performance (amount and complexity of knowledge and task-specific experience) (Schrepfer et al. 2009)	• duration of involvement with business process modeling (Mendling and Strembeck 2008) [o] • self-assessment of process modeling experience (Reijers and Mendling 2011) [o] • intensity of work with process models (Mendling and Strembeck 2008) [o] • frequency of use of “flow charts” (Ottensooser et al. 2012) [o] • work experience with process models (Recker and Dreiling 2011) [o] • modeling intensity and duration of modeling involvement (Mendling et al. 2012b) [o] • modeling familiarity (Ottensooser et al. 2012; Recker and Dreiling 2007) [o] • training in modeling at a university (Figl et al. 2013a; Reijers and Mendling 2011) [o] • self-assessment of previous modeling knowledge (Johannsen et al. 2014a; Turetken et al. 2016; Reijers and Mendling 2011; Weitlaner et al. 2013; Recker 2013) [o]
modeling knowledge	• theoretical knowledge (Schrepfer et al. 2009)	• process-modeling knowledge test (Figl and Laue 2015; Figl et al. 2013a; Figl et al. 2013b; Figl and Strembeck 2015; Kummer et al. 2016; Mendling and Strembeck 2008; Recker 2013; Recker et al. 2014) [o] • conceptual modeling test (Figl and Recker 2016) [p]
education	• highest educational degree (Schrepfer et al. 2009)	• higher education in general (Weitlaner et al. 2013) [o]
user characteristics	• perceptual expertise (Schrepfer et al. 2009) • ability to recognize patterns (Schrepfer et al. 2009)	• participants’ cognitive style (Figl and Recker 2016) [p] • users’ surface learning motive and strategy (Recker et al. 2014) [o]

Category of Independent Variables	**Theoretical Discussions**	**Empirical Studies (measuring only subjective comprehension [s] or only preferences/perceived usefulness [p] or measuring objective comprehension [o])**
	- traits, beliefs, perceived behavioral control, attitude, subjective norm, positive/negative anticipated emotion, self-efficacy, skills (Reijers et al. 2010)	- sensing versus intuitive learning style (Recker et al. 2014) [o] - native-language of labels (Recker and Dreiling 2011) [o]

Table 7 Framework of relevant influence factors based on articles with references (condensed version of Table 6)

	Influence Factors for Process Model Comprehension (Independent Variables)	**Theoretical and Empirical Contributions**
Presentation Medium	Presentation medium (paper versus computer)	**Empirical Studies:** (Mendling et al. 2012b), (Recker et al. 2014), (Turetken et al. 2016)
Notation	Representation paradigm (e.g., text versus model, differing dialects and cognitive fit, declarative versus imperative process models, assigning domain semantic-oriented pictorial elements like icons and images to modeling elements, animation, narration and visualization techniques)	**Theoretical Discussions:** (Rosa et al. 2011), (Genon et al. 2010), Fahland et al. 2009), (Mendling and Recker 2008; Mendling et al. 2010a), (Aysolmaz and Reijers 2016) **Empirical Studies:** (Ottensooser et al. 2012), (Rodrigues et al. 2015), (Figl and Recker 2016), (Pichler et al. 2012), (Hipp et al. 2014), (Figl and Recker 2016)
	Primary notation (e.g., BPMN, UML AD, BPMN, vBPMN, YAWL, C-YAWL, EPCs, configurable EPCs, SBD)	**Theoretical Discussions:** (Figl et al. 2009), (Genon et al. 2010), (Figl et al. 2010) **Empirical Studies:** (Döhring et al. 2014), (Sarshar and Loos 2005), (Recker et al. 2005), (Recker and Dreiling 2007), (Recker and Dreiling 2011), (Weitlaner et al. 2013), (Sandkuhl and Wiebring 2015), (Natschläger 2011), (Jošt et al. 2016), (Stitzlein et al. 2013), (Kock et al. 2009), (Kock et al. 2008)
	Notational characteristics (e.g., semiotic clarity, perceptual discriminability, semantic transparency, visual expressiveness, graphic economy)	**Theoretical Discussions:** (Figl et al. 2009; Genon et al. 2010; Figl et al. 2010), (Genon et al. 2010), (Rosa et al. 2011) **Empirical Studies:** (Figl et al. 2013a), (Recker, 2013)
Secondary Notation	Decomposition (use of decomposition/modularization, decomposition heuristics)	**Theoretical Discussions:** (Genon et al. 2010), (Rosa et al. 2011), (Storch et al. 2013), (Mendling et al. 2010b), (Mendling et al. 2012a), (Azim et al. 2008), (Gruhn and Laue 2006b), (Milani et al. 2016), (Johannsen et al. 2014b), (La Rosa et al. 2011), (Zugal et al. 2012), (Becker et al. 1995) **Empirical Studies:** (Reijers et al. 2011b), (Johannsen et al. 2014a), (Turetken et al. 2016)
	Gestalt theory (dual coding, highlighting, like using colors for control blocks)	**Theoretical Discussions:** (Genon et al. 2010), (La Rosa et al. 2011) **Empirical Studies:** (Bera 2012), (Jeyaraj and Sauter 2014), (Reijers et al. 2011a), (Kummer et al. 2016), (Petrusel et al. 2016)
	Layout (edges like crossing edges, direction, shape and size, symmetry, alignment of elements and spacing, ending points)	**Theoretical Discussions:** (Bernstein and Soffer 2015a), (Bernstein and Soffer 2015b), (Leopold et al. 2016), (La Rosa et al. 2011), (Schrepfer et al. 2009), (Becker et al. 1995) **Empirical Studies:** (Figl and Strembeck 2015), (Petrusel et al. 2016)
Label	Label design	**Theoretical Discussions:** (Koschmider et al. 2015a), (Fettke et al. 2012) **Empirical Studies:** (Mendling and Strembeck 2008)
	Naming conventions (syntactic like using a verb-object label style for activities, semantic like using a domain-specific vocabulary, avoidance of synonyms and homonyms)	**Theoretical Discussions:** (La Rosa et al. 2011), (Mendling et al. 2010b), (Fettke et al. 2012), (Mendling 2013), (Overhage et al. 2012), (Fettke et al. 2012), (Weber et al. 2011) **Empirical Studies:** (Mendling and Strembeck 2008), (Mendling et al. 2012b), (Figl and Strembeck 2015), (Mendling et al. 2010c), (Koschmider et al. 2015b)
Model Characteristics	Size measures (amount of activities, events, gateways, diameter)	**Theoretical Discussions:** (Rosa et al. 2011), (Storch et al. 2013), (Mendling et al. 2010b), (Mendling et al. 2012a), (Azim et al. 2008), (Gruhn and Laue 2006b), (Rosa et al. 2011), (Becker et al. 1995), (Mendling et al. 2010b), (Mendling et al. 2012a) **Empirical Studies:** (Recker, 2013), (Mendling and Strembeck 2008), (Sánchez-González et al. 2010), (Aguilar et al. 2008)
	Modularity and block structuredness (corresponding split and join elements) and related metrics (separability, maximum nesting depth)	**Theoretical Discussions:** (Rosa et al. 2011), (Storch et al. 2013), (Mendling et al. 2010b), (Mendling et al. 2012a) (Gruhn and Laue 2007), (Azim et al. 2008), (Gruhn and Laue 2006b)

		Empirical Studies: (Dumas et al. 2012), (Mendling and Strembeck 2008), (Sánchez-González et al. 2010), (Sánchez-González et al. 2012), (Figl and Laue 2015)
	Refactoring (simplification without changing the process's behavior)	**Theoretical Discussions:** (Gruhn and Laue 2009), (Weber et al. 2011)
	Removing redundant elements	**Theoretical Discussions:** (Rosa et al. 2011), (Weber et al. 2011), (Becker et al. 1995)
	Gateway interplay/control structures (XOR, cycles, OR, AND, concurrency) and related metrics (control flow complexity, sequentiality, cycles, heterogeneity of gateway types)	**Theoretical Discussions:** (Gruhn and Laue 2006a), (Azim et al. 2008), (Gruhn and Laue 2006b), (Mendling et al. 2010b), (Mendling et al. 2012a), (Storch et al. 2013), (Mendling 2013), (Reijers and Mendling 2011), **Empirical Studies:** (Aguilar et al. 2008), (Sánchez-González et al. 2012), (Reijers and Mendling 2011), (Döhring et al. 2014), (Sánchez-González et al. 2010), (Mendling and Strembeck 2008), (Figl and Laue 2011), (Figl and Laue 2015), (Laue and Gadatsch 2011), (Melcher et al. 2010), (Melcher and Seese 2008), (Weitlaner et al. 2013), (Sarshar and Loos 2005)
	Connection	**Theoretical Discussions:** (Mendling et al. 2010b), (Mendling et al. 2012a) **Empirical Studies:** (Reijers and Mendling 2011), (Sánchez-González et al. 2010), (Sánchez-González et al. 2012)
	Syntax rules	**Theoretical Discussions:** (Leopold et al. 2016) **Empirical Studies:** (Heggset et al. 2015), (Mendling and Strembeck 2008)
Task	Task	**Empirical Studies:** (Laue and Gadatsch 2011), (Figl and Laue 2015), (Pichler et al. 2012), (Soffer et al. 2015)
User	Tailoring of process models for personal factors	**Theoretical Discussions:** (Aysolmaz and Reijers 2016)
	Domain knowledge	**Theoretical Discussions:** (Schrepfer et al. 2009) **Empirical Studies:** (Bera 2012), (Recker and Dreiling 2007), (Recker et al. 2014), (Turetken et al. 2016)
	Experience and familiarity with modeling	**Theoretical Discussions:** (Genon et al. 2010), (Schrepfer et al. 2009) **Empirical Studies:** (Mendling and Strembeck 2008), (Ottensooser et al. 2012), (Recker and Dreiling 2011), (Mendling et al. 2012b), (Recker and Dreiling 2007), (Figl et al. 2013a), (Reijers and Mendling 2011), (Johannsen et al. 2014a), (Weitlaner et al. 2013), (Recker 2013), (Turetken et al. 2016)
	Modeling knowledge	**Theoretical Discussions:** (Schrepfer et al. 2009) **Empirical Studies:** (Figl et al. 2013a), (Figl et al. 2013b), (Figl and Laue 2015), (Figl and Strembeck 2015), (Figl and Recker 2016), (Kummer et al. 2016), (Mendling and Strembeck 2008), (Recker 2013), (Recker et al. 2014)
	Education	**Theoretical Discussions:** (Schrepfer et al. 2009) **Empirical Studies:** (Weitlaner et al. 2013)
	User characteristics	**Theoretical Discussions:** (Schrepfer et al. 2009), (Reijers et al. 2010) **Empirical Studies:** (Figl and Recker 2016), (Recker et al. 2014), (Recker and Dreiling 2011)

References

Aguilar ER, García F, Ruiz F, Piattini M, Visaggio CA, Canfora G (2008) Evaluation of BPMN models quality-a family of experiments. In: 3rd International conference on evaluation of novel approaches to software engineering, Funchal, pp 56–63

Aysolmaz B, Reijers HA (2016) Towards an integrated framework for invigorating process models: a research agenda. In: Reichert M, Reijers HA (eds) Business process management workshops 2015, LNBIP 256, Innsbruck, Austria. Springer International Publishing, pp 552–558. doi:10.1007/978-3-319-42887-1_44

Azim A, Ghani A, Koh Tieng WG, Muketha M, Wen WP (2008) Complexity metrics for measuring the understandability and maintainability of business process models using goal-question-metric. Int J Comput Sci Netw Secur 8(5):219–225

Becker J, Rosemann M, Schütte R (1995) Grundsätze ordnungsmäßiger Modellierung. WIRTSCHAFTSINFORMATIK 37(5):435–445

Bera P (2012) Does cognitive overload matter in understanding BPMN models? J Comput Inf Syst 52(4):59–69

Bernstein V, Soffer P (2015a) How does it look? exploring meaningful layout features of process models. In: Advanced Information Systems Engineering Workshops. Springer, Heidelberg, pp 81-86

Bernstein V, Soffer P (2015b) Identifying and quantifying visual layout features of business process models. In: Enterprise, business-process and information systems modeling. Springer, Heidelberg, pp 200–213

Blazhenkova O, Kozhevnikov M (2008) The new object-spatial-verbal cognitive style model: theory and measurement. Appl Cognit Psychol 23(5):638-663. doi:10.1002/acp.1473

Burton-Jones A, Meso P (2008) The effects of decomposition quality and multiple forms of information on novices understanding of a domain from a conceptual model. J Assoc Inf Syst 9(12):784-802

Döhring M, Reijers HA, Smirnov S (2014) Configuration vs. adaptation for business process variant maintenance: an empirical study. Inf Syst 39:108–133. doi:10.1016/j.is.2013.06.002

Dumas M, La Rosa M, Mendling J, Mäesalu R, Reijers H, Semenenko N (2012) Understanding business process models: the costs and benefits of structuredness. In: Ralyté J, Franch X, Brinkkemper S, Wrycza S (eds) Advanced information systems engineering, LNCS, vol 7328. Springer, Heidelberg, pp 31–46. doi:10.1007/978-3-642-31095-9_3

Fahland D, Lübke D, Mendling J, Reijers HA, Weber B, Weidlich M, Zugal S (2009) Declarative versus imperative process modeling languages: the issue of understandability. In: Halpin T, Krogstie J, Nurcan S et al (eds) Enterprise, business-process and information systems modeling. Proceedings of the 10th international workshop BPMDS 2009 and 14th international conference EMMSAD, pp 353–366

Felder RM, Spurlin J (2005) Applications, reliability and validity of the index of learning styles. Int J Eng Educ 21(1):103-112

Fettke P, Vella AL, Loos P (2012) From measuring the quality of labels in process models to a discourse on process model quality: a case study. In: 45th Hawaii international conference on system sciences (HICSS), pp 197–206. doi:10.1109/HICSS.2012.290

Figl K, Laue R (2011) Cognitive complexity in business process modeling. In: Mouratidis H, Rolland C (eds) Advanced information systems engineering, LNCS, vol 6741. Springer, Heidelberg, pp 452–466. doi:10.1007/978-3-642-21640-4_34

Figl K, Laue R (2015) Influence factors for local comprehensibility of process models. Int J Hum Comput Stud 82:96–110. doi:10.1016/j.ijhcs.2015.05.007

Figl K, Mendling J, Strembeck M (2009) Towards a usability assessment of process modeling languages. In: 8th GI-Workshop Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten (EPK), CEUR-WS, vol 554. Berlin, pp 138-156

Figl K, Mendling J, Strembeck M (2013a) The influence of notational deficiencies on process model comprehension. J Assoc Inf Syst 14(6):312–338

Figl K, Mendling J, Strembeck M, Recker J (2010) On the cognitive effectiveness of routing symbols in process modeling languages. In: 13th international conference on business information systems (BIS), Berlin

Figl K, Recker J (2016) Exploring cognitive style and task-specific preferences for process representations. Requir Eng 21(1):63–85. doi:10.1007/s00766-014-0210-2

Figl K, Recker J, Mendling J (2013b) A study on the effects of routing symbol design on process model comprehension. Decis Support Syst 54(2):1104–1118. doi:10.1016/j.dss.2012.10.037

Figl K, Strembeck M (2014) On the importance of flow direction in business process models. In: 9th International conference on software engineering and applications (ICSOFT-EA). SCITEPRESS, Vienna

Figl K, Strembeck M (2015) Findings from an experiment on flow direction of business process models. In: International workshop on enterprise modelling and information systems architectures (EMISA), Innsbruck

Genon N, Heymans P, Amyot D (2010) Analysing the cognitive effectiveness of the BPMN 2.0 visual syntax. In: Software language engineering, LNCS. Springer, Heidelberg, pp 377–396

Gruhn V, Laue R (2006) Adopting the cognitive complexity measure for business process models. In: 5th IEEE International conference on cognitive informatics, pp 236–241. doi:10.1109/COGINF.2006.365702

Gruhn V, Laue R (2006b) Complexity metrics for business process Models. In: 9th International conference on business information. Springer, Heidelberg, pp 1-12

Gruhn V, Laue R (2007) Good and bad excuses for unstructured business process models. In: 12th European conference on pattern languages of programs (EuroPLoP)

Gruhn V, Laue R (2009) Reducing the cognitive complexity of business process models. In: 8th IEEE International conference on cognitive informatics (ICCI'09), pp 339-345

Heggset M, Krogstie J, Wesenberg H (2015) The influence of syntactic quality of enterprise process models on model comprehension. CAiSE forum. CEUR, Stockholm

Hipp M, Strauss A, Michelberger B, Mutschler B, Reichert M (2014) Enabling a user-friendly visualization of business process models. Business process management workshops, LNBIP. Springer, Heidelberg, pp 395–407

Jeyaraj A, Sauter VL (2014) Validation of business process models using swimlane diagrams. J Inf Technol Manag 25(4):27

Johannsen F, Leist S, Braunnagel D (2014) Testing the impact of wand and weber's decomposition model on process model understandability. In: International conference on information systems, Auckland

Johannsen F, Leist S, Tausch R (2014b) Wand and Weber's good decomposition conditions for BPMN. Bus Process Manag J 20(5):693

Jošt G, Huber J, Heričko M, Polančič G (2016) An empirical investigation of intuitive understandability of process diagrams. Comput Stand Interfac 48:90–111. doi:10.1016/j.csi.2016.04.006

Kember D, Biggs J, Leung DY (2004) Examining the multidimensionality of approaches to learning through the development of a revised version of the Learning Process Questionnaire. Br J Educ Psychol 74(2):261-279

Kock N, Danesh A, Komiak P (2008) A discussion and test of a communication flow optimization approach for business process redesign. Knowl Process Manag 15(1):72

Kock N, Verville J, Danesh-pajou A, Deluca D (2009) Communication flow orientation in business process modeling and its effect on redesign success: results from a field study. Decis Support Syst 46(2):562–575. doi:10.1016/j.dss.2008.10.002

Koschmider A, Figl K, Schoknecht A (2015a) A comprehensive overview of visual design of process model element labels. In: Business process management workshops, Innsbruck

Koschmider A, Ullrich M, Heine A, Oberweis A (2015b) Revising the vocabulary of business process element labels. In: Zdravkovic J, Kirikova M, Johannesson P (eds) Advanced information systems engineering, LNCS, vol 9097. Springer, Heidelberg, pp 69–83. doi:10.1007/978-3-319-19069-3_5

Kummer T-F, Recker J, Mendling J (2016) Enhancing understandability of process models through cultural-dependent color adjustments. Decis Support Syst 87:1–12. doi:10.1016/j.dss.2016.04.004

La Rosa M, ter Hofstede AHM, Wohed P, Reijers HA, Mendling J, van der Aalst WMP (2011) Managing process model complexity via concrete syntax modifications. IEEE Trans Ind Inform 7(2):255–265

Laue R, Gadatsch A (2011) Measuring the understandability of business process models—are we asking the right questions? In: Business process management workshops, LNBIP, vol 66. Springer, Heidelberg, pp 37–48

Leopold H, Mendling J, Günther O (2016) Learning from quality issues of BPMN models from industry. IEEE Softw 33(4):26–33. doi:10.1109/MS.2015.81

Maes A, Poels G (2007) Evaluating quality of conceptual modelling scripts based on user perceptions. Data & Knowledge Engineering 63(3):701-724

Melcher J, Seese D (2008) Towards validating prediction systems for process understandability: measuring process understandability. In: 10th International symposium on symbolic and numeric algorithms for scientific computing (SYNASC'08), pp 564–571

Melcher J, Mendling J, Reijers HA, Seese D (2010) On measuring the understandability of process models. In: Business process management workshops, LNBIP, vol 43. Springer, Heidelberg, pp 465–476

Mendling J (2013) Managing structural and textual quality of business process models. Data-driven process discovery and analysis. Springer, Heidelberg, pp 100–111

Mendling J, Recker J (2008) Towards systematic usage of labels and icons in business process models. In: Halpin T, Proper HA, Krogstie J (eds) 12th International workshop on exploring modeling methods in systems analysis and design, pp 1-13

Mendling J, Recker J, Reijers HA (2010a) On the usage of labels and icons in business process modeling. Int J Inf Syst Model Des 1(2):40-58

Mendling J, Reijers H, van der Aalst WMP (2010b) Seven process modeling guidelines (7PMG). Inf Softw Technol 52(2):127–136

Mendling J, Reijers HA, Cardoso J (2007) What makes process models understandable. In: International conference on business process management, LNCS, Springer, Heidelberg, pp 48-63

Mendling J, Reijers HA, Recker J (2010c) Activity labeling in process modeling: empirical insights and recommendations. Inf Syst 35(4):467–482

Mendling J, Sánchez-González L, García F, La Rosa M (2012a) Thresholds for error probability measures of business process models. J Syst Softw 85(5):1188–1197. doi:10.1016/j.jss.2012.01.017

Mendling J, Strembeck M (2008) Influence factors of understanding business process models. In: Abramowicz W, Fensel D (eds) Business information systems (BIS), vol 7. Springer, Heidelberg, pp 142–153

Mendling J, Strembeck M, Recker J (2012b) Factors of process model comprehension—findings from a series of experiments. Decis Support Syst 53(1):195–206

Mendling J, van der Aalst WM (2006) Towards EPC semantics based on state and context. In: 5th GI-Workshop Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten (EPK), pp 25-48

Milani F, Dumas M, Matulevičius R, Ahmed N, Kasela S (2016) Criteria and heuristics for business process model decomposition. Bus Inf Syst Eng 58(1):7–17

Natschläger C (2011) Deontic BPMN. In: Hameurlain A, Liddle S, Schewe K-D, Zhou X (eds) Database and expert systems applications, LNCS, vol 6861. Springer, Heidelberg, pp 264–278. doi:10.1007/978-3-642-23091-2_23

Ottensooser A, Fekete A, Reijers HA, Mendling J, Menictas C (2012) Making sense of business process descriptions: an experimental comparison of graphical and textual notations. J Syst Softw 85(3):596–606

Overhage S, Birkmeier DQ, Schlauderer S (2012) Quality marks, metrics, and measurement procedures for business process models. Bus Inf Syst Eng 54(5):217–235. doi:10.1007/s11576-012-0335-1

Petrusel R, Mendling J, Reijers HA (2016) Task-specific visual cues for improving process model understanding. Inf Softw Technol 79:63–78

Pichler P, Weber B, Zugal S, Pinggera J, Mendling J, Reijers HA (2012) Imperative versus declarative process modeling languages: an empirical investigation. In: Business process management workshops, LNBIP, vol 99, pp 383–394

Recker J (2010) Continued use of process modeling grammars: the impact of individual difference factors. Europ J Inf Syst 19(1):76-92

Recker J (2013) Empirical investigation of the usefulness of gateway constructs in process models. Eur J Inf Syst 22(6):673–689

Recker J, Dreiling A (2007) Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education. In: Toleman M, Cater-Steel A, Roberts D (eds) Australasian conference on information systems. The University of Southern Queensland, pp 356–366

Recker J, Dreiling A (2011) The effects of content presentation format and user characteristics on novice developers’ understanding of process models. Commun Assoc Inf Syst 28 (1), Article 6, pp. 65-84

Recker J, Reijers HA, van de Wouw SG (2014) Process model comprehension: the effects of cognitive abilities, learning style and strategy. Commun Assoc Inf Syst 34(9):199–222

Recker J, Rosemann M, van der Aalst WMP (2005) On the user perception of configurable reference process models—initial insights. In: 16th Australasian Conference on Information Systems, Paper 66

Reijers HA, Freytag T, Mendling J, Eckleder A (2011a) Syntax highlighting in business process models. Decis Support Syst 51:339–349

Reijers HA, Mendling J (2011) A study into the factors that influence the understandability of business process models. IEEE Trans Sys Man Cybern Part A 41(3):449–462

Reijers HA, Mendling J, Dijkman RM (2011b) Human and automatic modularizations of process models to enhance their comprehension. Inf Syst 36(5):881–897

Reijers HA, Recker JC, van de Wouw SG (2010) An integrative framework of the factors affecting process model understanding: a learning perspective. In: 16th Americas conference on information systems: sustainable IT collaboration around the globe. Association for Information Systems, Paper 184

Rodrigues RDA, Barros MDO, Revoredo K, Azevedo LG, Leopold H (2015) An experiment on process model understandability using textual work instructions and BPMN models. In: 29th Brazilian symposium on software engineering (SBES), pp 41–50. doi:10.1109/SBES.2015.12

Rosa ML, Wohed P, Mendling J, ter Hofstede AHM, Reijers HA, van der Aalst WMP (2011) Managing process model complexity via abstract syntax modifications. IEEE Trans Ind Inform 7(4):614–629. doi:10.1109/TII.2011.2166795

Sánchez-González L, García F, Mendling J, Ruiz F (2010) Quality assessment of business process models based on thresholds. In: Meersman R, Dillon T, Herrero P (eds) On the move to meaningful internet systems: OTM 2010, LNCS, vol 6426. Springer, Heidelberg, pp 78–95. doi:10.1007/978-3-642-16934-2_9

Sánchez-González L, García F, Ruiz F, Mendling J (2012) Quality indicators for business process models from a gateway complexity perspective. Inf Softw Technol 54(11):1159–1174. doi:10.1016/j.infsof.2012.05.001

Sandkuhl K, Wiebring J (2015) Experiences from selecting a BPM notation for an enterprise. Business information systems workshops. Springer, Heidelberg, pp 126–138

Sarshar K, Loos P (2005) Comparing the control-flow of epc and petri net from the end-user perspective. In: International conference on business process management. Springer, Heidelberg, pp 434–439

Schrepfer M, Wolf J, Mendling J, Reijers HA (2009) The impact of secondary notation on process model understanding. In: The practice of enterprise modeling, 2nd IFIP WG8.1 working conference (PoEM 2009), Stockholm, Sweden

Soffer P, Wand Y, Kaner M (2015) Conceptualizing routing decisions in business processes: theoretical analysis and empirical testing. J Assoc Inf Syst 16(5):2

Stitzlein C, Sanderson P, Indulska M (2013) Understanding healthcare processes. An evaluation of two process model notations. In: Human factors and ergonomics society annual meeting, vol 1. Sage, New York, pp 240–244

Storch A, Laue R, Gruhn V (2013) Measuring and visualising the quality of models. In: IEEE 1st International workshop on communicating business process and software models quality, understandability, and maintainability (CPSM), pp 1–8. doi:10.1109/CPSM.2013.6703084

Turetken O, Rompen T, Vanderfeesten I, Dikici A, van Moll J (2016) The effect of modularity representation and presentation medium on the understandability of business process models in BPMN. In: International conference on business process management. LNCS, Springer, Heidelberg, pp 289–307

Weber B, Reichert M, Mendling J, Reijers HA (2011) Refactoring large process model repositories. Comput Ind 62(5):467

Weitlaner D, Guettinger A, Kohlbacher M (2013) Intuitive comprehensibility of process models. In: Fischer H, Schneeberger J (eds) S-BPM ONE—running processes, vol 360. Communications in computer and information science. Springer, Heidelberg, pp 52–71. doi:10.1007/978-3-642-36754-0_4

Wiebring J, Sandkuhl K (2015) Selecting the “right” notation for business process modeling: experiences from an industrial case. In: Matulevičius R, Dumas M (eds) Perspectives in business informatics research, LNBIP, vol 229. Springer, Heidelberg, pp 129-144. doi:10.1007/978-3-319-21915-8_9

Zugal S, Pinggera J, Weber B, Mendling J, Reijers HA (2012) Assessing the impact of hierarchy on model understandability—a cognitive perspective. In: Kienzle J (ed) Models in software engineering: workshops and symposia at MODELS 2011, Wellington, New Zealand, Reports and revised selected papers. Springer, Heidelberg, pp 123–133. doi:10.1007/978-3-642-29645-1_14