6-18-2022

Modeling styles in conceptual data modeling: Reflecting observations in a series of multimodal studies

Kristina Rosenthal
University of Hagen, kristina.rosenthal@fernuni-hagen.de

Johannes Wagner
University of Hagen, johannes.wagner@fernuni-hagen.de

Stefan Strecker
University of Hagen, stefan.strecker@fernuni-hagen.de

MODELING STYLES IN CONCEPTUAL DATA MODELING: REFLECTING OBSERVATIONS IN A SERIES OF MULTIMODAL STUDIES

Research Paper

Kristina Rosenthal, University of Hagen, Germany, kristina.rosenthal@fernuni-hagen.de

Johannes Wagner, University of Hagen, Germany, johannes.wagner@fernuni-hagen.de

Stefan Strecker, University of Hagen, Germany, stefan.strecker@fernuni-hagen.de

Abstract

A modeling style characterizes a modeler's sequencing of processing a modeling task in terms of applying the modeling language and its language concepts while constructing a conceptual model. Presently, surprisingly little is known about the different modeling styles modelers exhibit when performing conceptual data modeling. In this research, we combine complementary modes of observation including audio-visual protocols, recorded modeler-tool interactions, and pre-/post-modeling surveys of modelers to identify modeling styles in 24 data modeling processes performed by modelers at different stages of experience in data modeling. Our study identifies and characterizes three distinct modeling styles refining our current knowledge about data modeling processes and informing design science research on style-specific, targeted modeling (software tool) support for data modelers.

Keywords: Modeling style, Conceptual data modeling, Mixed methods research, Multimodal research.

1 Introduction

Conceptual modeling as an activity, e.g., when constructing a data model as Entity-Relationship diagram, involves an intricate array of cognitive processes and performed actions including goal setting, abstracting, conceptualizing, associating and contextualizing, interpreting and sense-making, evaluating and judging, anticipating, envisioning and thinking ahead, drawing and visualizing, and, in group settings, communicating, discussing and agreeing (Rosenthal and Strecker, 2019; Strecker et al., 2021). Performing conceptual modeling is construed as a complex task based on codified and tacit knowledge; a task that requires modelers to master theoretical foundations, modeling languages—such as the Entity-Relationship Model (ERM), Chen (1976)—modeling methods and modeling tools, to apply them to practical problems and, while performing a modeling process, to critically think about modeling objectives, the application domain and its technical terminology (Frank et al., 2014).

The process (the actual "act") of conceptual modeling has recently received renewed attention in conceptual modeling research (e.g., Bera et al., 2019; Claes et al., 2015a; Hoppenbrouwers et al., 2005; Serral et al., 2016; Wilmont et al., 2010): How conceptual modeling is performed by modelers, how individual modeling processes proceed, how novice modelers' modeling processes differ from more advanced modelers has been subject, in particular, to studies on conceptual data modeling (e.g., Batra and Davis, 1992; Chaiyasut and Shanks, 1994; Srinivasan and Te'eni, 1995; Venable, 1996).

While research on modeling styles has been a recurring research theme in business process modeling research (e.g., Claes et al., 2015b; Pinggera et al., 2015), surprisingly little is known about modeling styles in conceptual data modeling (Topi and Ramesh, 2011). Generally conceptualized as a way of proceeding in a modeling process (Hoppenbrouwers et al., 2005), i.e., activities the modeler performs to construct the conceptual model (Claes et al., 2015a), a modeling style (or modeling strategy) characterizes a modeler's sequencing of modeling activities including adding, changing, (re)labeling, moving, deleting or revising model elements which, in turn, includes frequent revisiting of the modeling task and of the modeling objectives. General modeling activities depend on and are determined by the modeling language the modelers use. Hence, a modeling style characterizes a modeler's sequencing of processing a modeling task in terms of applying the modeling language and its language concepts while constructing a conceptual model. We expect modelers constructing a conceptual data model with the ERM to exhibit differences in the sequencing of processing and applying the concepts entity type, relationship type, attribute, cardinality and generalization/specialization (cf. Claes et al., 2015b; Pinggera et al., 2015). Specifically, differences are expected for modelers at different levels of modeling experience (e.g., Batra and Davis, 1992). By identifying and understanding distinct styles of conceptual data modeling the present research contributes to refine our current knowledge about data modeling processes, informs and guides design science research on style-specific, targeted modeling (software tool) support for data modelers, and contributes to developing the theoretical foundation of conceptual (data) modeling.

In the study at hand, we revisit and reflect upon observations in a series of three multimodal studies on modelers constructing a conceptual data model starting from a natural language description—a common scenario in a learning context in which learners of data modeling are expected to model a universe of discourse as described in a textual description of a modeling tasks by means of an ER diagram, e.g., in higher education courses on data modeling. We identify, classify and compare modeling styles of 24 data modelers at different stages of experience in data modeling ranging from modelers with little prior experience to expert modelers with years of data modeling experience. Following a mixed methods research design, we combine complementary modes of observation including audio-visual protocols, recorded modeler-tool interactions, and pre-/post-modeling surveys of modelers to allow for rich insights into their modeling styles. Different from earlier studies, these multimodal observations allow us to inspect ambiguous and peculiar sequencing of modeling activities when reasoning about the modeling style is unclear. Building upon early studies on modeling difficulties in data modeling (Rosenthal and Strecker, 2019; Rosenthal et al., 2020b), we re-examine recorded data and re-analyze audio-visual protocols to identify, classify and compare distinct data modeling styles. Our primary research objective is to identify and characterize modeling styles as a further step toward understanding and justifying differences in data modeling styles. Given the different stages of experience of modelers in the set of studies, our secondary research objective is to present a preliminary comparison of modeling styles by modeler experience in search of particular indications, e.g., if a specific modeling style is preferred by (more) experienced modelers.

The paper is structured as follows: The next section discusses related work and theoretical foundations (Sect. 2). Section 3 introduces the research design. Findings on identified modeling styles are presented in Sect. 4. Section 5 includes a discussion of the findings and future research directions, followed by a discussion of limitations in Sect. 6 and a conclusion in Sect. 7.

2 Related Work

Research on modeling styles in data modeling traces back to Batra et al. (1988, 1990) observing differences in how novice and more advanced modelers proceed in modeling processes. Batra and Antony (1994) and Batra and Davis (1992) conclude that more advanced modelers exhibit a modeling style characterized by decomposing the overall problem description into meaningful parts, gathering and organizing all relevant information concerning the different parts, and then, by mapping the information to appropriate knowledge structures, followed by mapping them to a specific representation—whereas the novice modelers' modeling style is described as considerably different in that they follow a linear, beginning-to-end processing of the modeling task. Srinivasan and Te'eni (1995) add to these findings on discriminable data modeling styles that better performing modelers move orderly between different levels of abstraction—a known problem-solving heuristic—and spend short ("burst") periods of time at a specific level of abstraction. They conclude that modeling tool support should assist modelers in moving between levels of abstraction. Following earlier research, Venable (1996) states that novice modelers' objectives in data modeling processes are typically rather narrowly framed: Novice modelers appear to focus on capturing the specific semantics of the problem description considering the parts of the problem description isolated from each other. More advanced modelers, by contrast, pursue several modeling objectives at the same time, and try to get a holistic understanding of the problem at hand first. Then, they consider all the system requirements and third, they consider future requirements as well. Research by Hoppenbrouwers et al. (2006, 2005) on conceptual modeling in general takes a communication-based approach and identifies modeling strategies taken by individuals based on linguistics analyses—viewing conceptual modeling as a dialogue. Apart from this work, very few further contributions discuss modeling styles in data modeling: A recent review of research on human factors in data modeling by Topi and Ramesh (2011) only finds Hoffer (1982)'s work to control for individual modeling style. The work by Shanks (e.g., Shanks, 1997) and by Siau and Tan (e.g., Siau and Tan, 2005), while discussing human cognition, does not discuss modeling styles.

Another stream of related research investigates cognitive mechanisms of business process modeling by comparing modeling approaches of novices and experts (Wilmont et al., 2010). The study suggests that main differences between experts and novices relate to experts having a richer mental model of possible modeling concepts and being able to integrate the different actions of modeling, i.a., abstracting, generalizing and reflecting, in a purposeful way—in contrast to novices experiencing difficulties in integrating the actions and not finding the right level of abstraction. A further study investigates cognitive processes while modeling to understand how these processes influence modeling behavior and modeling skills (Wilmont et al., 2013). Further work focuses on the process of process modeling leading to distinct modeling styles in process modeling (Pinggera et al., 2015) as well as cognitive process modeling techniques applied by modelers (Claes et al., 2015a)—resulting in the so called Structured Process Modeling Theory (SPMT) aimed at explaining how the probability of an occurrence of cognitive overload in process modeling processes can be reduced (Claes et al., 2015a).

Different from earlier work on (data) modeling styles, the present study starts from a refined conceptualization of modeling style as introduced in the previous section which—different from prior work—puts modeling language concepts at the center of the analysis of modeling styles, i.e., the analysis is performed at the level of language concepts of the ERM: We study conceptual data modeling processes at the level of processing and modeling entity types, relationship types, generalization hierarchies, attributes with their data types and cardinalities. As prior insights into such sequences are rare, we pursue an explorative approach, i.e., our research does not build on pre-conceptualizations of the sequencing of processing and modeling language concepts. However, an initial intuition suggests, e.g., the presumption that modelers do not start with modeling relationship types but rather entity types in creating the data model. Furthermore, the present research design stands out from extant studies by seizing verbalizations of modelers as primary source of insight into modelers' deliberations in combination with behavioral observations of the modelers' interactions with a modeling software tool enriched by visual cues from recording over-the-shoulder. It is completed with a pre- and post-modeling survey. As the experience level of our participants encompasses non-experienced, medium-experienced and experiemced modelers we also tie on the approaches of Batra and Davis (1992) and Srinivasan and Te'eni (1995) who highlight expert-novice differences. This variety of modeler characteristics enables us to get a better understanding of the applied modeling styles according to the modelers' level of modeling experience.

3 Research Design

3.1 Methodological considerations

The present work employs a mixed methods research design (Creswell and Plano Clark, 2018; Venkatesh et al., 2016) to investigate modeling styles of 24 data modelers at different levels of expertise in a series of three multimodal observation studies at three different locations (in Spain, Belgium and Germany). The research design is characterized by "mix[ing] or combin[ing] quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study" (Johnson and Onwuegbuzie, 2004, p. 17) to achieve insights into modeling styles beyond results from either type of data separately. Following a mixed methods design pursues the objective of diversity of views (Venkatesh et al., 2016, p. 442), including two points of data integration: one during the observations and one during data analysis (see Figure 1). Based on the assumption that modeling styles cannot be identified based on behavioral data alone, analyzing audio-visual protocols of modeling processes is complemented with analyzing modeler-tool interactions, e.g., on the sequencing of modeling activities, and subjects are surveyed before and after performing the modeling. Further details on the multimodal observation and data generation approach, its validation and application is presented in Rosenthal et al. (2020a). Please note that the data collected in the three studies has been analyzed on modeling difficulties in data modeling in prior work (Rosenthal and Strecker, 2019; Rosenthal et al., 2020b)—we revisit and re-analyze recorded data to identify modeling styles in the present research.

We observe subjects individually constructing a conceptual data model in three studies conducted in 2019 and 2020: Study I in January 2019 with eight non-experienced modelers as subjects (P1–P8), Study II with eight medium-experienced modelers as subjects conducted in February 2020 (P9–P16) and Study III with eight expert modelers in May/June 2019 (P17–P24). The purposeful sampling of subjects spanning modelers at different stages of knowledge and experience in data modeling is aimed at selecting individuals able to provide in-depth information about the phenomenon under investigation (Creswell and Plano Clark, 2018). Participation was voluntary and all subjects participated individually. In Study I and Study III, the subjects were offered no other incentives than the opportunity to participate in the study. In Study II, the subjects received 10 Euros as a monetary incentive for their participation. Starting from a data modeling task presented to modelers in form of a natural language description, subjects are instructed to construct a conceptual data model (an ER diagram) reconstructing the statements of the problem representation using a browser-based modeling tool. The modeling tool provides subjects with a graphical modeling editor for creating data models, i.e., ER diagrams, in a notational variant of the ERM data modeling language (showing notation symbols for entity types, relationship types etc. and an—initially, empty—modeling canvas). The modeling tool is integrated with a modeling research observatory supporting multimodal observations and analysis of the collected data (see Ternes et al., 2020a; Ternes et al., 2020b).

The data modeling tasks employed for the studies (see supplementary material, Rosenthal et al., 2022) are based in the library domain (Study I and Study III) respectively refer to a car rental (Study II) in order to reduce effects of varying prior domain knowledge (cf. Bera et al., 2014; Pretz et al., 2003). We chose the car rental modeling task in Study II—instead of the library domain used in Study I and III—because of a prior exposure of participating subjects to a library modeling exercise. The three modeling tasks performed by the subjects are deliberately designed to balance demand on subjects, time to perform the task and modeling complexity—depending on the respective stages of knowledge and experience in data modeling.

3.2 Multimodal observations

As first point of data integration, the multimodal observation setup combines three modes of observation to achieve complementary perspectives on an individual modeling process:

1. **Recording audio-visual protocols** during conceptual modeling by a subject, i.e., while working on a modeling task, comprising a think-aloud (verbal) protocol and a video recorded from an "over-the-shoulder" perspective. Subjects are asked to think out loud about their reasoning on the modeling task, about their modeling decisions, and, more generally, about their deliberations towards working on the modeling task and on arriving at their objective, i.e., creating a purposeful ER diagram. The subjects' verbalizations of their own thinking provides insights into each subject's cognitive processes in their own words (e.g., into the modeler's reasoning regarding the sequencing of her modeling activities).

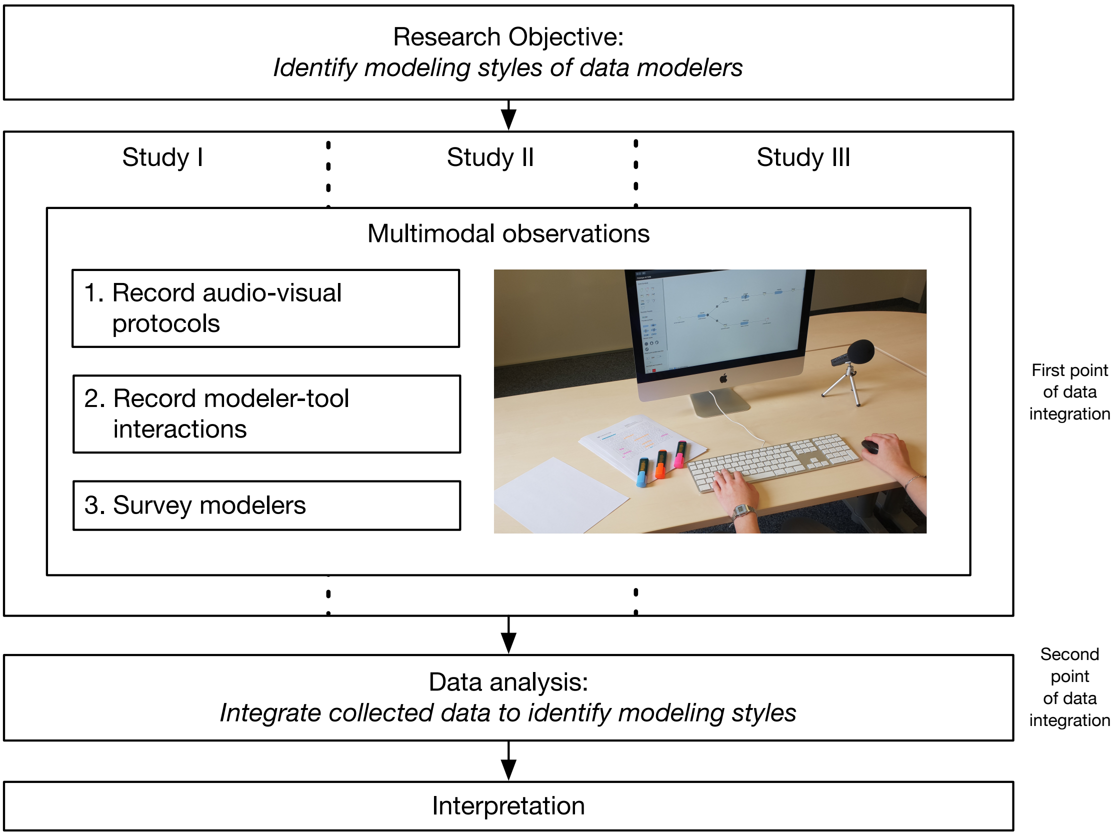

Figure 1. Research design.

The visual "over-the-shoulder" recording provides additional nonverbal cues on each individual's modeling process as conveyed by body language and movement (e.g., reading and marking the modeling task). Verbal protocols, an observation technique common to problem-solving research (e.g., Ericsson and Simon, 1980; van Someren et al., 1994), is thus purposefully combined with visual cues to provide a richer picture of the phenomenon under investigation, e.g., to resolve observations that cannot be analyzed based on verbal protocols alone (e.g.; Zugal et al., 2013).

2. **Recording modeler interactions with the modeling canvas** to observe specific modeling decisions, in particular, decisions with respect to placing a new model element on the modeling canvas, to change an existing model element, to element repositioning, to deletion of model elements, and to renaming a model element. Every such modeler interaction with the graphical editor is tracked and recorded as a time discrete event during the construction of a data model.
3. **Surveying subjects pre- and post-modeling** to collect information on modeler demographics and to obtain self-disclosed information on prior modeling experience and domain knowledge of the modeling task and, post-modeling, a self-assessment regarding the verbalization of own thoughts during think aloud—to identify additional cues regarding the identification of a modeling style.

Each modeling session took place in a quiet environment—to make the subjects "feel at ease" (as suggested, e.g., by van Someren et al., 1994, p. 41)—with one researcher present as an observer who is familiar with the modeling task and tool as well as with the multimodal observations. After being informed about the purpose of the study and completing a consent form, the data collection procedure for all subjects included the following steps:

1. Studying a short description of the semantics of the modeling concepts and their graphical notation to explain the specifics of the chosen variant of the ERM.
2. Watching a short video introduction and tool demonstration of 2–3 minutes. Think aloud instructions by the observer to verbalize all thoughts while modeling "as if alone in the room" and

information that reminders would be given after a predetermined period of silence of 30 seconds—except the subject is reading the modeling task—with the precise wording “Please keep talking”.

4. Performing a warm-up modeling task, i.e., constructing a simple data model, to become familiar with the modeling tool and to practice verbalizing thoughts while modeling.
5. Filling in a pre-modeling survey asking closed-ended and open-ended questions on prior conceptual modeling experience and perceived familiarity with the domain of the modeling task.
6. Performing the main modeling task based on a description in natural language presented on paper. Each subject was instructed to use the modeling tool to construct a conceptual data model based on the description. During modeling, the verbalizations of the participant and a video of the modeler’s behavior were collected using a camcorder. All interactions with the modeling tool were recorded as time-discrete events in the modeling tool.
7. Filling in a post-modeling survey comprising closed-ended and open-ended questions on perceived difficulties with think aloud and domain knowledge, as well as demographic information.

A pre-test of the data collection procedure was conducted with one researcher familiar with conceptual data modeling in December 2018. As results, the description of the semantics of the modeling concepts and the graphical notation of the ER model was revised to reduce complexity and to enhance understandability. In addition, adjustments were made to the questionnaires improving understandability and clarity of the questions. However, data collection in the three studies differed in terms of the sequence of steps performed and the used study material that were both adjusted based on experience gained in the conduct of the previous studies. For example, the three steps of filling in the pre-modeling survey, performing the main modeling task and filling in the post-modeling survey were grouped into a data collection part preceded by the other, introductory steps, e.g., the warm-up modeling task, in Study II and III, which had not been the case in Study I.

3.3 Subject characteristics

Information on open- and closed-ended questions from the pre- and post-modeling surveys are integrated into a description characterizing the sample of subjects with regard to prior conceptual modeling experience and demographic information (discussed as “qualitizing” in literature on mixed methods research, e.g., Venkatesh et al., 2016, pp. 446f).

For the first study in January 2019, we recruited eight bachelor and master students of business informatics or business administration at a German university with little to no experience in conceptual data modeling, subjects we would characterize as non-experienced modelers. Of the eight participants (P1-P8), five subjects were male while three were female with an age ranging between 27 and 52 (with a median of 41,5 years and a mean of 40,5 years). As first language, six subjects stated German, one subject English and the other one German and French. Six subjects were bachelor students of Business Informatics (four subjects) or Business Administration (two subjects), and the remaining two subjects were master students of Business Informatics. Seven of the subjects had completed courses on conceptual modeling during their studies including six subjects who attended an introductory course on “Modeling Business Information Systems” dealing with the notational variant of the ER model used in this study (the remaining subject did not further specify the attended course). In addition, three subjects explained prior experience in conceptual data modeling ranging from two months with constructing three conceptual models and reading six models, over about two and a half years with constructing 15 conceptual models and reading 50 to almost nine years of experience with constructing 20 conceptual models and reading 200. Participant 8 with the outstanding experience constitutes an exceptional case demanding a special consideration in the analysis. However, as practical experience is discussed as only one aspect of being an experienced or expert modeler besides theoretical knowledge and training (e.g., Batra and Davis, 1992, p.87; Venable, 1996, p.50), the outlier is included in further analyses. Besides this exceptional case, the entirety of the characteristics and background of the subjects suits the intention to study subjects which can be characterized as non-experienced modelers.

For the study in February 2020, we recruited eight bachelor and master students who took advanced courses on data modeling or data(base) management at a Belgian university, subjects we expect to have some experience in conceptual data modeling but who we would characterize as medium-experienced modelers. Four of the subjects were female and four male with an age ranging between 20 and 32 (with a median of 22,5 years and a mean of 23,9 years). As first language, three subjects stated Dutch and one subject each stated Arabic, Czech, Spanish, Russian and Yoruba. Seven subjects were master students of Bioinformatics, Biology, Business, Business and Information Systems Engineering (two subjects), Data Science, Information Management and one subject a bachelor student of Business and Information Engineering. All eight subjects had completed one to three courses during their studies that teach conceptual modeling (with a median of 2) with seven subjects who attended an introductory course on “Principles of Database Management” that includes data modeling with the ER model. Further completed courses related to, for example, information systems modeling (i.a., including process modeling with the Business Process Model and Notation, BPMN) or business information systems (i.a., including modeling with the UML). All subjects explained prior experience in conceptual data modeling ranging from four months with constructing five conceptual data models and reading ten models to four years and six month experience with constructing 50 conceptual data models and reading 200. The median of experience in conceptual data modeling is one year and six months with a median of constructing 17,5 conceptual data models and reading 42,5 models. Participant 10 constitutes an exceptional case with the more than four years experience with constructing 50 conceptual data models and reading 200. As in the first study, this outlier is included in further analyses, demanding a special consideration. Altogether, the subjects besides the outlier have some experience in conceptual data modeling with substantial differences between the subjects and, considering the subjects’ background and characteristics, we characterize them as medium-experienced modelers.

In the study conducted in May/June 2019, eight experienced modelers participated (P17-P24), i.e., academics from the context of conceptual modeling, working in an academic context or studying in the final stage of their Master studies at a Spanish university. Four subjects were female and four male with an age between 25 and 45 (with a median of 35 years and a mean of 35.125 years). As first language, five subjects stated Spanish, one Dutch, another one Portuguese and the other one German. Two subjects were postdoctoral researchers holding a PhD, four were doctoral candidates with Master’s degrees and two Master students with Bachelor’s degrees, holding their highest degree in Computer Science (7) or Business and Social Studies (1) and working or studying in subjects within the fields of Computer Science (7) or Information Science (1). All eight subjects reported prior experience in conceptual data modeling ranging from two years with constructing 30 conceptual models and reading 50 models to 20 years with constructing 100 conceptual models and reading 200 models (with a median experience of 10.5 years, a median of constructing 40 conceptual models and reading 75 models). Seven of the subjects had one to several completed courses on conceptual modeling during their studies. Altogether, the characteristics of the subjects suit the intention to study subjects which can be characterized as experienced modelers in an academic context—in alignment with conceptualizations of expert modelers in related studies (e.g., Batra and Davis, 1992, p. 87, Shanks, 1997, p. 65).

With regard to domain knowledge, the answers of participants in all three studies indicate that the participants understood the respective modeling domain well enough to perform the task, i.e., subjects basically understand the service that a library respectively a car rental company offers.

3.4 Data analysis

To identify modeling styles the modelers exhibit when constructing the data model, different types of data are integrated, constituting the second point of data integration including data transformation (Creswell and Plano Clark, 2018, pp. 224–226). In a first step, each audio-visual protocol of a modeling process is systematically inspected for a modeler's sequencing of processing the modeling task and modeling the different concepts of the ERM, i.e., entity types with corresponding attributes with their data types and possible generalization hierarchies, and relationship types with cardinalities. This includes coding the modeling processes for the modeler's interactions with the modeling task, i.e., reading and marking the textual description, to achieve insights into the sequence of processing the modeling task by systematically assigning codes to segments of the audio-visual protocols (e.g., Creswell and Plano Clark, 2018, pp. 213–215). If segments of audio-visual protocols are unclear with respect to identifying the sequences of modeling the different concepts, we supplement the audio-visual analysis with analyzing recorded modeler-tool interactions for closer inspection of the observed sequence of data modeling steps. After analyzing the modeling processes individually, we turn to an aggregated view by comparing the individual modeling processes: Based on observing shared properties in the sequences of processing the modeling task and modeling the different language concepts, the modeling processes are grouped and thus distinguished into groups of modeling processes adopting discernible modeling styles—following the first steps in developing a classification or taxonomic theory (e.g., Gregor, 2006; Kundisch et al., 2021).

4 Modeling Styles

Analyzing the data collected from 24 modeling processes, we observe indicators for different modeling styles with respect to the sequencing of modeling entity types, relationship types, attributes with their data types, cardinalities. Table 1 gives an overview of the participants in the three studies (Study I to III) with their level of modeling experience, the domain of the modeling task and the completion times of the modeling processes. Note that the modeling tasks for the three studies are designed depending on the respective stages of knowledge and experience in data modeling of the participants, resulting in varying times to perform the task between the studies. Moreover, the observed differences among subjects in the length of the modeling process in each study are in line with earlier work on prior modeling knowledge and modeling experience of conceptual data modeling (e.g., Batra and Davis, 1992, p. 94). In the following, we present our observations from analyzing the modeling processes with respect to the modeling style a modeler applies, for each study separately. Subsequently, the modeling styles identified across all three studies are described.

Study I – January 2019 *Non-experienced modelers* *Domain of the modeling task: Library*									
Participants	P1	P2	P3	P4	P5	P6	P7	P8	#
Completion time	23	20	35	18	15	17	35	9	172
Study II – February 2020 *Medium-experienced modelers* *Domain of the modeling task: Car rental*									
Participants	P9	P10	P11	P12	P13	P14	P15	P16	#
Completion time	32	22	32	21	33	32	35	22	229
Study III – May/June 2019 *Expert modelers* *Domain of the modeling task: Library*									
Participants	P17	P18	P19	P20	P21	P22	P23	P24	#
Completion time	41	52	47	49	46	46	38	41	360

Table 1. Overview of participants in the three observation studies with completion times (in minutes) of modeling processes.

In the first study conducted at a German university in 2019 with non-experienced modelers (Study I), we observe participants choosing different sequences of modeling language concepts: Five participants (P3, P4, P5, P6, P8) start by creating entity types, attributes and data types in a first step. Only in a subsequent step, relationship types are created and cardinalities are assigned at the same time. Two further participants (P1, P2) start by creating entity types with associated attributes and data types as well, but continue by creating all relationship types before they determine cardinalities for the modeled relationship types in a separate step. The modeling process of the remaining participant (P7) does not exhibit a comparable, traceable sequencing of modeling activities. Furthermore, we observe that six of the eight non-experienced modelers (P1, P2, P3, P5, P6, P7) read the entire modeling task before starting to model in the modeling tool and use the colored markers provided during the study to mark text segments referring to model elements.

We realized our second study at a Belgian university in 2020 with medium-experienced modelers as participants (Study II). Again, we observe different sequences of modeling language concepts: Two participants start by creating entity types, attributes and data types and only in a second step create relationship types and assign cardinalities at the same time (P11, P15). Three further participants proceed strictly according to the sequence of statements in the modeling task when creating the model elements (P10, P13, P16)—a sequencing we have not observed with the non-experienced modelers in the first study. The remaining three participants (P9, P12, P14) start with modeling entity types with their associated attributes, but then switch between modeling entity types, relationship types and determining cardinalities—a traceable sequencing of modeling steps is not recognizable in the modeling processes of these three participants. With respect to the interaction with the provided modeling task, we observe four subjects that do not use the provided colored markers (P9, P10, P13, P14). The further four subjects use the markers to highlight possible model elements in the modeling task description when reading the entire modeling task before modeling (P11, P12, P15, P16).

The third study was conducted at a Spanish university in 2019 observing expert modelers (Study III): Regarding the sequencing of modeling activities, we find four participants who model strictly according to the sequence of statements in the modeling task description (P19, P20, P21, P22), as observed for a few medium-experienced modelers. Two participants, by contrast, start with modeling entity types with their corresponding attributes and data types in a first step and only then proceed with the relationship types and cardinalities simultaneously (P18, P24). One modeler (P23) exhibits a sequencing of modeling the different language concepts that is not observed in Study I and Study II: The modeler begins with modeling all the entity types first and then continues with modeling relationship types between the entity types. Subsequently, the modeler inserts all the attributes and data types into the model and finalizes it with determining the cardinalities of the relationship types. The remaining modeler (P17) does not show a specific sequencing with regard to modeling language concepts: The modeler starts with creating some of the entity types and relationship types and continues with the attribute types and data types of the modeled entity types. In a next step, further entity types and relationship types are added, and the modeler proceeds once again with modeling the attribute types. The model is finalized by determining the cardinalities. Furthermore, we observe that six of the eight expert modelers (P17, P18, P19, P21, P23, P24) turn their attention to reading the entire modeling task and highlighting specific parts of the task description before building their model.

In the aggregate view of all three studies, our findings reinforce our presumption that participants would choose different modeling styles: Regarding the sequencing of modeling entity types, relationship types, generalization hierarchies, attributes with their data types and cardinalities, our observations suggest different modeling styles that we explain in the following. Figure 2 shows an overview of the identified modeling styles as observed in the modeling processes in the three studies.

Entity type-focused modeling style: We observe eleven participants (P1, P2, P3, P4, P5, P6, P8, P11, P15, P18, P24) first creating all entity types of the aspired data model, together with the corresponding attributes and data types. Only in a next step, these participants create relationship types and determine cardinalities. The proceeding of the participants suggests a modeling style which evolves around the completion of entity types with their attributes and corresponding data types in a first step, hence an entity type-focused modeling style. With eleven out of 24 participants, this is the most frequently adopted modeling style. It is noteworthy that the vast majority of subjects adopting this modeling style, i.e., ten subjects (P1, P2, P3, P5, P6, P7, P11, P15, P18, P24), read the entire modeling task including thinking about all model elements before starting to model in the modeling tool and use colored markers provided during the study to mark text segments referring to model elements in the paper-based modeling task.

Linear modeling style: Our findings indicate that modelers adopt further styles on how to create the conceptual data model. We observe seven modelers who strictly follow the sequence of statements in the natural language description in modeling entity types, relationship types, generalization hierarchies, attributes with their data types and cardinalities—from beginning to end—to construct the data model accordingly (P10, P13, P16, P19, P20, P21, P22). This suggests a modeling style that linearly follows the modeling task and that we thus denote as linear modeling style. In the group of subjects adopting the linear modeling style, we observe only three subjects (P16, P19, P21) using the markers to highlight possible model elements in the modeling task description, of which all three subjects read the entire modeling task before starting to create the data model.

Structuring modeling style: Beside the two modeling styles explained above, we observe one subject (P23) modeling all entity types followed by creating the relationship types between the entity types in a first step. Only in a next step, attributes and corresponding data types are added to the entity types and cardinalities are determined for the relationship types. We identify this proceeding as structuring modeling style on the basis of the modeling decision to create the entity types and relationship types—hence, the structure of the data model—in a first step. This very participant reads the modeling task from beginning to end and uses the provided markers to mark text segments referring to model elements in the paper-based modeling task before starting to create the model in the modeling tool.

	Study I Non-experienced modelers	Study II Medium-experienced modelers	Study III Expert modelers	No. of observed modelers
Entity type-focused modeling style	P1 P2 P3 P4 P5 P6 P8	P11 P15	P18 P24	11
Linear modeling style		P10 P13 P16	P19 P20 P21 P22	7
Structuring modeling style			P23	1
No modeling style recognizable	P7	P9 P12 P14	P17	5

Figure 2. Overview of identified modeling styles with occurrences in Study I, II and III.

Moreover, the modeling processes of the remaining five modelers (P7, P9, P12, P14, P17) do not exhibit a traceable sequencing of processing the modeling task and modeling the different language concepts. In addition, we do not recognize similarities between the proceeding of the modelers. Hence, the modeling processes are not assigned to a specific modeling style. However, these modelers do not create the data model erratically—but we observe no specific sequencing, e.g., in the sense that they model entity types first or process the task in a linear manner, is recognizable. It is noteworthy that this applies to only one non-experienced modeler (Study I) and one expert modeler (Study III) but, with three subjects, to almost half of the medium-experienced modelers (Study II).

5 Discussion

Analyzing over 12 hours of visually-supported verbal protocols of non-experienced, medium-experienced and expert modelers combined with analyzing modeler-tool interactions and surveying the modelers, we identify three distinct modeling styles adopted by these modelers to create a conceptual data model starting from a textual description. In the following, we discuss the findings with respect to the modelers' stage of modeling experience and suggest fruitful paths for future research on modeling styles in conceptual data modeling.

It was our presumption that a modeling style depends on the used modeling language with its language concepts, here a variant of the ERM with its entity types, relationship types, generalization hierarchies, attributes with their data types and cardinalities. Our findings, in addition, indicate that specific modeling styles are preferred in the different studies: The vast majority of non-experienced modelers adopts the *entity type-focused modeling style* while we observe the medium-experienced modelers applying the *entity type-focused* and *linear modeling style* and observe all three identified distinct modeling styles in the expert modelers' modeling processes. The outliers regarding prior modeling experience, Participants 8 and 10, do not exhibit a modeling style that stands out recognizably from those of the other participants in the respective study. There is, furthermore, no specific modeling style that most medium-experienced or expert modelers apply in comparison to novice modelers though. The variety in modeling styles used by the individuals in these groups may stem from deliberate decisions of the subjects based on their individual modeling experience and preference. Hence, it is indicated that, with a higher level of modeling experience, modelers tend to develop a specific modeling style—possibly rather a modeling strategy—that is adopted for creating a data model—in contrast to beginning modelers who seem to adopt the approach to proceeding with a natural language modeling task commonly taught in data modeling courses at universities today, i.e., to start with creating entity types with the corresponding attributes and data types and only then continue with creating the relationship types and determining cardinalities. However, prior work suggests that novice modelers' modeling processes follow a linear, beginning-to-end processing of the modeling task (Batra and Antony, 1994; Batra and Davis, 1992). Our findings, on the contrary, suggest that the *linear modeling style* is not chosen by the observed non-experienced modelers, but more frequently adopted by the medium-experienced and expert modelers. One plausible thesis is that teaching of data modeling at the time of Batra et al.'s research emphasized a different modeling style (linear) compared to current prevalent teaching approaches (entity type-focused). A further observation is that with more modeling experience the observed modelers appear to turn to a modeling style that strictly follows the sequence of statements in the natural language modeling task rather than the different language concepts, the *linear modeling style*, or modelers do not exhibit a traceable sequencing of processing the modeling task and modeling language concepts. Prior work indicated that more advanced and expert modelers focus on developing a holistic understanding of the problem at hand in a first step (Batra and Davis, 1992; Venable, 1996). However, our observations indicate that the majority of all observed modelers in our studies read and annotate the entire modeling task including thinking about model elements before starting to model in the modeling tool, i.e., seventeen of 24—with seven of eight modelers this especially applies to the non-experienced modelers in Study I. Moreover, we do not find that the medium-experienced and expert modelers in our studies adopt a modeling style that indicates a more holistic processing of the natural language description and, hence, the problem at hand. On the contrary, the modeling processes of several of those subjects is characterized by a beginning-to-end processing of the modeling task.

Besides the modelers' level of modeling experience, our observations, furthermore, give indications that further modeler characteristics besides modeling experience are relevant for adopting a specific modeling style: Especially in Study II and III, we observe modelers with a similar level of modeling experience adopting different modeling styles working on the same data modeling task: Medium-experienced modelers in Study II as well as expert modelers in Study III split into modelers applying an *entity type-focused modeling style* and a *linear modeling style*, complemented with one modeler opting for a *structuring modeling style* and further modelers for whom no modeling style is discernible. This observation indicates that adopting a specific modeling style depends on further modeler characteristics.

It is expected that—in addition to modeler characteristics—task-specific characteristics are relevant for choosing a modeling style (e.g., Pinggera et al., 2015). The data modeling tasks employed for the studies constitute typical modeling tasks as natural language descriptions, e.g., used in teaching conceptual data modeling. Each task was deliberately designed in alignment with the level of modeling experience of the participants in the respective study—with the aim to show a certain complexity posing challenges on the participants, but in balance with the time to perform the task. However, future research refining the present findings should aim to investigate the influence of task-specific characteristics on data modeling styles.

Our observations encourage further studies into individual modeling processes tying in with the exploratory results of the present study: (1) A fruitful path lies in further investigating how following a linear, beginning-to-end processing of a modeling task is related to a modeler's level of modeling experience—and how it relates to prevalent teaching approaches for conceptual data modeling. (2) Regarding modelers developing a holistic understanding of a modeling problem, future studies could aim to focus on observing how modelers—non-experienced, medium-experienced and expert modelers—interact with a modeling task and how they proceed in processing the task. Following these two paths promises to shed light on the contradiction of our current, exploratory findings with findings of previous research (e.g., Batra and Antony, 1994; Batra and Davis, 1992; Venable, 1996). (3) A further fruitful path for future research lies in exploring further modeler characteristics in future studies: For example, further studies could expand the surveying of participants regarding their background, including more detailed questions on the modeler's studies of conceptual modeling and domain knowledge—in order to shed light on varying modeler characteristics besides modeling experience. In addition, interviewing participants post-modeling, i.e., after finishing the modeling task, offers the possibility to ask modelers about adopting a specific modeling style and thus to achieve in-depth insights into the modelers' reasoning (Creswell and Plano Clark, 2018; Myers and Newman, 2007).

6 Limitations

The sample size of 24 subjects in total and 8 subjects at each level of modeling experience aligns with the number of participants in related studies applying the think aloud approach (e.g., Batra and Davis, 1992; Nielsen, 1994; Haisjackl et al., 2016; Haisjackl et al., 2018). Think aloud protocol analysis is recognized to provide in-depth insights into cognitive processes. Studying an increased number of participants promises to lead to further insights into data modeling styles. With regard to characterizing modelers as non-experienced, medium-experienced and expert modelers, we observe that the subjects in each study have substantial differences of knowledge and experience in conceptual modeling. For follow-up studies, research efforts are needed into how (practical) experience in conceptual modeling, knowledge of theoretical foundations of conceptual modeling and modeling training interact with respect to characterizing modelers, e.g., as expert or novice modelers. In addition, general limitations apply to the think aloud method: Thinking aloud is considered to affect the cognitive resources of individuals while modeling potentially leading to a decreased modeling performance, especially as the modeling task includes a visual, non-verbal perceptual component

(Ericsson and Simon, 1980). Furthermore, differences in verbalization skills have long been discussed (e.g., Ericsson and Simon, 1993) with modelers having difficulties verbalizing their thoughts while modeling on principle accounts (Blech et al., 2019) or on modeling-related accounts—resulting in some verbal protocols not being as complete as others (van Someren et al., 1994).

The present study observes modelers constructing a conceptual data model starting from a natural language description. Assuming that data modeling starts from a natural language description of the modeling task primarily in a learning scenario, the modeling tasks used in the studies may be considered to be artificial—specifically in the light of imprecise representation of requirements in real-world contexts, e.g., as interview transcripts, notes, videos or in documents. However, the universe of discourse that is to be modeled is always represented and communicated through language (Frank, 2006) and, hence, a representation as natural language description is also possible to achieve in real-world applications. Therefore, we deem our set-up to be appropriate also with regard to real-world contexts as it reduces this complexity factor to enable gathering insights into the modelers' reasoning during modeling.

7 Conclusion

Integrating complementary modes of observation to identify modeling styles in 24 individual data modeling processes of modelers with different levels of modeling experience, the present work contributes to modeling style research by identifying and characterizing three distinct data modeling styles: (1) an entity type-focused modeling style, (2) a linear modeling style and (3) a structuring modeling style. Contrary to results from earlier studies (Batra and Antony, 1994; Batra and Davis, 1992; Venable, 1996), we find novice modelers predominantly to follow an entity type-focused modeling style rather than a linear modeling style, and we do not find (more) experienced modelers to adopt a modeling style that indicates a more holistic processing of the modeling task. Still, further research is needed to understand distinct styles of conceptual data modeling and if a specific modeling style is preferred by novice respectively (more) experienced modelers. Studying progressively more and more individual modeling processes and following a refined data analysis strategy focusing on the modelers' cognitive processes and taking into account the constructed data models promises to shed further light on these open questions, and, ultimately, to enable us to design style-specific (software tool) support for data modelers. Hence, we intend to complement the present studies with follow-up studies under varying conditions, i.e., online, in the field and the laboratory, observing subjects including practitioners with various backgrounds, e.g., regarding prior experience and knowledge of conceptual data modeling as well as their first language.

As next steps following the present work, we are planning for a series of future studies refining and extending the exploratory insights. Currently, we are conducting a follow-up study with non-experienced modelers: In a virtual setup, subjects work on a data modeling task in the web-based modeling tool while the modeling processes are recorded remotely, including the recording of think aloud protocols, tracking modeler-tool interactions and surveying modelers. In this setup, subjects are able to participate in the study from any place—seeming an appropriate way to continue our research on modeling processes despite the restrictions imposed by the COVID-19 pandemic. In addition, we are planning for a large-scale study with a considerably higher number of participants with different levels of modeling experience—also in a virtual setup—to expand our data set of modeling processes: The observation setup is adapted for a large-scale study by focusing the data collection on recording modeler-tool interactions and surveying modelers. Complementing our current data analysis strategy, data from the large-scale study is open for applying additional analysis techniques, e.g., cluster analysis (cf. Pinggera et al., 2015) to contribute to refining the exploratory insights into discernible modeling styles in conceptual data modeling.

References

Batra, D. and S. R. Antony (1994). “Novice errors in conceptual database design.” *European Journal of Information Systems* 3 (1), 57–69.

Batra, D. and J. G. Davis (1992). “Conceptual data modelling in database design: similarities and differences between expert and novice designers.” *International Journal of Man-Machine Studies* 37 (1), 83–101.

Batra, D., J. A. Hoffer, and R. P. Bostrom (1988). A Comparison of User Performance between the Relational and the Extended Entity Relationship Models in the Discovery phase of Database Design.” *In: Proceedings of the 9th International Conference on Information Systems* (ICIS 1988). Minneapolis, MN, USA.

— (1990). “Comparing Representations with Relational and EER Models.” *Commun. ACM* 33 (2), 126–139.

Bera, P., A. Burton-Jones, and Y. Wand (2014). “How Semantics and Pragmatics Interact in Understanding Conceptual Models.” *Information Systems Research* 25 (2), 401–419.

Bera, P., P. Soffer, and J. Parsons (2019). “Using Eye Tracking to Expose Cognitive Processes in Understanding Conceptual Models.” *MIS Quarterly* 43 (4).

Blech, C., R. Gaschler, and M. Bilalic´ (2019). “Why do people fail to see simple solutions? Using think-aloud protocols to uncover the mechanism behind the Einstellung (mental set) effect.” *Thinking & Reasoning 26* (4), 552–580.

Chaiyasut, P. and G. G. Shanks (1994). Conceptual Data Modeling Process: A Study of Novice and Expert Data Modellers. *In: Proceedings of the 1st International Conference on Object-Role Modeling* (ORM-1). Magnetic Island, Australia, pp. 310–323.

Chen, P. P.-S. (1976). “The Entity-Relationship Model—Toward a Unified View of Data.” *ACM Transactions on Database Systems* 1 (1), 9–36.

Claes, J., I. Vanderfeesten, F. Gailly, P. Grefen, and G. Poels (2015a). “The Structured Process Modeling Theory (SPMT) – A cognitive view on why and how modelers benefit from structuring the process of process modeling.” *Information Systems Frontiers* 17 (6), 1401–1425.

Claes, J., I. Vanderfeesten, J. Pinggera, H. A. Reijers, B. Weber, and G. Poels (2015b). “A visual analysis of the process of process modeling.” *Information Systems and e-Business Management* 13 (1), 147– 190.

Creswell, J. W. and V. L. Plano Clark (2018). *Designing and Conducting Mixed Methods Research*. 3rd Edition. Los Angeles, CA: Sage.

Ericsson, K. A. and H. A. Simon (1980). “Verbal reports as data.” *Psychological Review* 87 (3), 215–251.

— (1993). *Protocol analysis: Verbal reports as data*. 2nd Edition. Cambridge, MA: MIT Press.

Frank, U. (2006). *Towards a Pluralistic Conception of Research Methods in Information Systems Research*. ICB Research Report 7. University of Duisburg-Essen.

Frank, U., S. Strecker, P. Fettke, J. vom Brocke, J. Becker, and E. Sinz (2014). “The Research Field “Modeling Business Information Systems”: Current Challenges and Elements of a Future Research Agenda.” *Business & Information Systems Engineering* 6 (1), 39–43.

Gregor, S. (2006). “The Nature of Theory in Information Systems.” *MIS Quarterly* 30 (3), 611–642.

Haisjackl, C., I. Barba, S. Zugal, P. Soffer, I. Hadar, M. Reichert, J. Pinggera, and B. Weber (2016). “Understanding Declare models: strategies, pitfalls, empirical results.” *Software & Systems Modeling* 15 (2), 325–352.

Haisjackl, C., P. Soffer, S. Lim, and B. Weber (2018). “How do humans inspect BPMN models: an exploratory study.” *Software & Systems Modeling* 17 (2), 655–673.

Hoffer, J. A. (1982). An empirical investigation into individual differences in database models. *In: Proceedings of the 3rd International Conference on Information Systems* (ICIS 1982). Ann Arbor, MI, USA, pp. 153–168.

Hoppenbrouwers, S. J. B. A., L. Lindeman, and H. A. Proper (2006). Capturing Modeling Processes – Towards the MoDial Modeling Laboratory. *In: On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*. LNCS, Vol. 4278. Berlin, Heidelberg: Springer, pp. 1242–1252.

Hoppenbrouwers, S. J. B. A., H. A. Proper, and T. P. van der Weide (2005). A Fundamental View on the Process of Conceptual Modeling. *In: 24th International Conference on Conceptual Modeling* (ER 2005). LNCS, Vol. 3716. Berlin, Heidelberg: Springer, pp. 128–143.

Johnson, R. B. and A. J. Onwuegbuzie (2004). "Mixed Methods Research: A Research Paradigm Whose Time Has Come." *Educational Researcher* 33 (7), 14–26.

Kundisch, D., J. Muntermann, A. M. Oberländer, D. Rau, M. Röglinger, T. Schoormann, and D. Szopinski (2021). "An Update for Taxonomy Designers: Methodological Guidance from Information Systems Research." *Business & Information Systems Engineering*. DOI: 10.1007/s12599-021-00723-x.

Myers, M. D. and M. Newman (Jan. 2007). "The qualitative interview in IS research: Examining the craft." *Information and Organization* 17 (1), 2–26.

Nielsen, J. (1994). "Estimating the number of subjects needed for a thinking aloud test." *International Journal of Human-Computer Studies* 41 (3), 385–397.

Pinggera, J., P. Soffer, D. Fahland, M. Weidlich, S. Zugal, B. Weber, H. A. Reijers, and J. Mendling (2015). "Styles in business process modeling: an exploration and a model." *Software and Systems Modeling* 14 (3), 1055–1080.

Pretz, J. E., A. J. Naples, and R. J. Sternberg (2003). Recognizing, defining, and representing problems. *In: The psychology of problem solving*. New York, NY, US: Cambridge University Press, pp. 3–30.

Rosenthal, K., B. Ternes, and S. Strecker (2020a). Understanding individual processes of conceptual modeling: A multi-modal observation and data generation approach. *In: Modellierung 2020*. Vienna, Austria, pp. 77–92.

Rosenthal, K. and S. Strecker (2019). Toward a Taxonomy of Modeling Difficulties : A Multi-Modal Study on Individual Modeling Processes. *In: 40th International Conference on Information Systems* (ICIS 2019). Munich, Germany.

Rosenthal, K., S. Strecker, and O. Pastor (2020b). Modeling difficulties in data modeling: Similarities and differences between experienced and non-experienced modelers. *In: 39th International Conference on Conceptual Modeling* (ER 2020). LNCS, Vol. 12400. Berlin, Heidelberg: Springer, pp. 501–511.

Rosenthal, K., J. Wagner, and S. Strecker (2022). *Supplementary material for "Modeling styles in conceptual data modeling: Reflecting observations in a series of multimodal studies"*. https://doi.org/10.5281/zenodo.6387447.

Serral, E., J. De Weerdt, G. Sedrakyan, and M. Snoeck (2016). Automating Immediate and Personalized Feedback: Taking Conceptual Modelling Education to a Next Level. *In: 10th International Conference on Research Challenges in Information Science* (RCIS 2016). Grenoble, France: IEEE, pp. 1–6.

Shanks, G. (1997). "Conceptual data modelling: An empirical study of expert and novice data modellers." *Australasian Journal of Information Systems* 4 (2), 63–73.

Siau, K. L. and X. Tan (2005). "Improving the quality of conceptual modeling using cognitive mapping techniques." *Data & Knowledge Engineering* 55 (3), 343–365.

Srinivasan, A. and D. Te'eni (1995). "Modeling as Constrained Problem Solving: An Empirical Study of the Data Modeling Process." *Management Science* 41 (3), 419–434.

Strecker, S., K. Rosenthal, and B. Ternes (2021). Studying Conceptual Modeling Processes: A Modeling Tool, Research Observatory, and Multimodal Observation Setup. *In:* Gimpel, H., Krämer, J., Neumann, D., Pfeiffer, J., Seifert, S., Teubner, T., Veit, D. J. & Weidlich, A. (eds.) *Market Engineering: Insights from Two Decades of Research on Markets and Information*. Cham: Springer International Publishing, pp. 99–111.

Ternes, B., K. Rosenthal, H. Barth, and S. Strecker (2020a). TOOL - Modeling Observatory & Tool: An Update. *In: Short, Workshop and Tools & Demo Papers Modellierung 2020*, Vienna, Austria. CEUR-WS, Vol. 2542. Vienna, Austria, pp. 198–202.

Ternes, B., K. Rosenthal, S. Strecker, and J. Bartels (2020b). TOOL–A Modeling Observatory & Tool for Studying Individual Modeling Processes. *In: Demo Track at the 39th International Conference on Conceptual Modeling* (ER 2020). CEUR-WS, Vol. 2716. Vienna, Austria, pp. 178–182.

Topi, H. and V. Ramesh (2011). "Human Factors Research on Data Modeling." *Journal of Database Management* 13 (2), 3–19.

van Someren, M. W., Y. F. Barnard, and J. A. C. Sandberg (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*. London: Academic Press.

Venable, J. R. (1996). Teaching Novice Conceptual Data Modellers to Become Experts. *In: International Conference Software Engineering: Education and Practice*. Dunedin, New Zealand: IEEE, pp. 50–56.

Venkatesh, V., S. A. Brown, and Y. W. Sullivan (2016). "Guidelines for Conducting Mixed-Methods Research: An Extension and Illustration." *Information Systems* 17 (7), 435–494.

Wilmont, I., S. Brinkkemper, I. van de Weerd, and S. Hoppenbrouwers (2010). Exploring Intuitive Modelling Behaviour. *In: Enterprise, Business-Process and Information Systems Modeling*. Vol. 50. Berlin, Heidelberg: Springer, pp. 301–313.

Wilmont, I., S. Hengeveld, E. Barendsen, and S. Hoppenbrouwers (2013). Cognitive mechanisms of conceptual modelling. *In: 32th International Conference on Conceptual Modeling* (ER 2013). LNCS, Vol. 8217. Berlin, Heidelberg: Springer, pp. 74–87.

Zugal, S., C. Haisjackl, J. Pinggera, and B. Weber (2013). "Empirical evaluation of test driven modeling." *International Journal of Information System Modeling and Design* 4 (2), 23–43.