Participant	P1	P2	P3	P4	P5	P6	P7	P8	#
Completion time (in minutes)	23	20	35	18	15	17	35	9	172
Breakdowns (overall)	3	2	6	3	2	3	6	0	25
• Differentiate between entity types			1						1
• Decide between entity and relationship type	2		3	2					7
• Develop identifiers for relationship types	1	2	2			1	2		8
• Choose data type of attribute					1		1		2
• Determine cardinalities				1	1	2	3		7

Table 2. Completion times and numbers of breakdowns

For example, a passage from the think aloud protocol of P3 indicates this type of difficulties: "*I think I have too few entities. Hmm... I'm not sure right now […] or the loan. The users can borrow as many copies of a library item as they like. Hmm… Is it possible to create an entity loan? If it would be so, if I would do that, then... Um, I would have... Where would I do that? I have a user who borrows books, library items, he can borrow several books at the same time. I don't have that at the moment, I can't represent that. I only have the relationship between the user and the library item. So, I will put the loan between*" (P3). Difficulties of this type caused long and severe periods of uncertainty in the modeling processes (of up to 3 min) and, in this sense, are particularly remarkable—especially because this type of difficulties occurred seven times and because two of the respective three participants were not able to find a solution for the difficulty.

Develop identifiers for relationship types: Difficulties of this type occurred in five modeling processes (P1, P2, P3, P6, P7) constituting the most frequent type of difficulties in terms of the total number of occurrences (8) and number of participants concerned (5). This type of difficulties refers to a modeler who creates a relationship type and encounters a problem with finding a descriptive and sensible identifier for the model element. For example, P7 faces a difficulty of this type ("*In any case, we have a relationship type between user and loan here. Hmm… What is the best way to call it that it makes sense somehow... Um, how can we connect user and loan in sensible way? ... Ok, let's do it the other way around...*", P7) as well as P6 ("*Um... I call the relationship um... well, how do you call that? Um... I say, belongs to*", P6).

Choose a data type for an attribute: Encountered by two participants (P5, P7), this type of difficulties relates to the attributes of entity types (note that the chosen variant of the ER model and its notation does not allow for attributes of relationship types, to simplify the learning process for modeling beginners). These two participants faced the difficulty of choosing a data type for an attribute that is adequate in the context of the modeling task. Please note that a full list of predefined data types was included in the instructions and available to subjects throughout the entire modeling process. For example, P7 was in doubt about the data type for the attribute ISBN: "*We also need the ISBN. That is… ISBN is actually alpha-numeric, because there are minuses in it. So, let's take a string... however, you can also write them without minuses. Then it would be Integer... Um... I would say a matter of consideration... We make an integer out of it [...] That fits, right? … Well, no, moment...*" (P7). P5 encountered a difficulty with the attribute SHELFMARK: "*Each copy has a*

unique shelfmark... Hmm, could of course also be string, could also be... Uh, a library could also consider... There, you sometimes also have some... Well, maybe I'll change that to String" (P5).

Determine cardinalities: We identified difficulties with regard to determining cardinalities for relationship types in four modeling processes (P4, P5, P6, P7) with a total of seven occurrences. Remarkably, five of the seven occurrences of this type of difficulties pertain to a relationship type with one-to-one cardinalities. A text passage from the think aloud protocol of P5 illustrates this type of difficulties: "*A loan always refers to exactly one copy, that means 1 and 1. And a loan to one copy, and well, a copy can... yes... hmm… A copy can actually be as many... It can be borrowed or not, I would say now for the one example. It can be borrowed several times, but only once at the same time. So, I would say that now. Does that make sense?... I model it that way for now*" (P5). Also, P6 faced problems related to determining cardinalities that the participant was not able to resolve: "*From every copy... I just have to look again. To each item ... yes... hmm... no, that doesn't work like that... Hmm... Um. Then I would choose another approach first, so that we don't sit here for hours now*" (P6).

Further Observations

The observations indicate that the modeling processes of the eight participants differ in certain respects: First, six of the eight subjects take the opportunity to use the colored markers to mark text segments in the paper-based modeling task while two participants (P4, P8) do not use the markers. In addition, it is noteworthy that P1 deviates from the instructions by creating fragments of the model with pen&paper before interacting with the modeling tool. Second, in creating the conceptual data model with the modeling tool, we observe participants choosing different sequences of creating model elements. Four participants (P3, P4, P5, P6) start by creating entity types, attributes and data types and only then create relationship types and assign cardinalities at the same time. Participants 1 and 2 start by creating entity types as well, but continue by creating relationship types and, in a separate step, determine cardinalities for all modeled relationship types. The remaining two participants (P7, P8) do not exhibit a comparable, traceable sequencing of modeling steps. Third, regarding the length of the modeling processes, we observe a wide range from nine minutes for the outlier P8 and from 15 to 35 minutes for the other seven participants. The modeling process of P3 is noticeable different from the other subjects' modeling processes in terms of speed of tool interaction, in particular, model construction is performed much slower, and the resulting data model is incomplete as it misses several attributes described in the modeling task. Further exploring the modeler-tool interaction and the audio-video protocol reveals that the participant terminated working on the modeling task without finishing. We conclude that the modeling task—as intended—shows a certain complexity posing challenges on the participants.

Interestingly, the verbal protocols entail remarks on the modeling tool only by three modelers (P3, P4, P6). Participants 4 and 6 mentioned criticism of the modeling tool, e.g., regarding the visualization of attributes in entity types, but do not exhibit difficulties with respect to the modeling tool. The recordings of subject P3's

interactions with the graphical editor exhibit particular difficulties when adding attributes to entity types (four times). Regarding domain knowledge, i.e., knowledge of the library domain, the participants were asked to self-assess the statements "I understood what the modeling task was about" and "I am familiar with the domain of the modeling task" on a scale from 1 to 7 where 1 corresponds to "I do not agree at all" and 7 to "I agree entirely". Regarding the first statement, all eight participants entirely agreed. For the second statement, the answers ranged from 2 to 7 with a median of 5,5 which indicates that the participants understood the chosen modeling domain well enough to perform the task.

Analyzing the audio recordings of the eight individual data modeling processes led us to observe substantial differences in how well subjects are able to verbalize their cognitive processes while modeling. Differences in verbalization skills, especially the ease with which people verbalize thoughts, have long been discussed, resulting in the advice to offer think aloud training (e.g., van Someren et al. 1994, pp. 34f)—an advice the present study followed. In about three total hours of verbal protocols, we do not observe a single silent period of 30 seconds or more, and, therefore, conclude that the think aloud instructions were suitable to initiate the intended behavior. In the post-modeling survey, the participants were asked to self-assess the statement "I had difficulties to verbalize my thoughts" on a scale from 1 to 7 where 1 corresponds to "I do not agree at all" and 7 to "I agree entirely". The answers of seven participants ranged from 1 to 3, with only one participant choosing a 5. This very participant actually exhibited problems in verbalizing thoughts in the first few minutes of the modeling process, explicitly pointing to having two native languages as one reason. However, after a few minutes, the participant started to verbalize her/his thoughts in a comprehensible way, especially regarding modeling difficulties.

Discussion and Conclusion

Integrating complementary modes of observation of eight individual data modeling processes and an analysis using the concept of cognitive breakdowns leads us to identify five types of modeling difficulties these subjects face while performing the data modeling task. We discuss fruitful paths for future research on modeling difficulties and design science research on developing (tool) support for learners of conceptual modeling.

Our findings suggest that the majority of difficulties encountered by the participants in the modeling processes relates to modeling relationship types (difficulties of the types *Decide between entity type and relationship type*, *Develop identifiers for relationship types*, *Determine cardinalities*). This observation is in line with prior work on difficulties in conceptual data modeling (Batra 1993) and on cognitive complexity in data modeling (Batra 2007) suggesting that modeling problems are not experienced mainly in modeling entity types and attributes, but in modeling relationship types. Our findings suggest that—in addition to severe problems with deciding whether a relationship type warrants modeling—modelers especially faced difficulties with regard to developing sensible identifiers for relationship types and in determining cardinalities (see Table 2). The exploratory findings of the present study can serve as starting point for

future research on individual modeling processes. To better understand typical difficulties encountered in conceptual modeling, further studies into individual modeling processes are encouraged which tie in with the exploratory results of the present study: For example, modeling tasks could aim to induce specific challenges with respect to the identified modeling difficulties or, e.g., with regard to modeling concepts such as generalization/specialization in data or object modeling. The developed mixed methods research design integrates complementary modes of observation of modeling processes in an innovative way—providing a methodical basis for future studies on individual modeling processes.

Within the long-term research program on targeted (tool) support for learners of conceptual modeling, the exploratory results of the present study are intended as a starting point for developing a taxonomy of modeling difficulties over the course of multiple studies, in the sense of a classification or taxonomic theory (following, e.g., Gregor 2006): Such a taxonomy is expected to distinguish modeling difficulties that occur in individual modeling processes based on shared properties and to include decision rules to assign difficulties to the resulting types of difficulties (Gregor 2006, p. 619). The types of modeling difficulties identified in the present study provide a starting point for such a classification system: Further studies have to build on the preliminary classification of modeling difficulties for analyzing further individual modeling processes—refining the classification system by adding emerging types of difficulties on the basis of characteristics of the actual difficulties observed in the modeling processes (Gregor 2006, p. 619). The taxonomy, in turn, is intended to serve as theoretical foundation for design science research on developing (tool) support for learners of conceptual modeling: On the basis of distinctions of modeling difficulties following the taxonomy, support for learners is in prospect that systematically and deliberately targets modeling difficulties. However, a number of further studies is needed to deepen and substantiate our understanding of modeling difficulties.

The observed differences among subjects in the length of the modeling process is in line with earlier work on prior modeling knowledge and modeling experience of conceptual data modeling (e.g., Batra and Davis 1992, p. 94). It is not surprising that P8 exhibits the shortest modeling process with substantially less time spent to complete the modeling task—in the light of 20+ years of modeling experience. Regarding the sequencing of modeling activities (constructing entity types, relationship types, determining cardinalities) in the individual modeling processes, our findings reinforce our presumption that participants would choose different approaches. Hence, a potential path for future research lies in further investigating the modeling approaches exhibited in individual modeling processes. For this path of research, the exploratory findings of the present study can serve as starting point. We deem the distinct styles of modeling identified in Pinggera et al. (2015) and cognitive modeling techniques identified for process modeling in Claes et al. (2015) as further fruitful anchor points. As a subsequent step, we deem exploring the interdependencies between the approaches to modeling and specific modeling difficulties promising—contributing to better understanding modeling difficulties in their genesis and to, subsequently, develop targeted modeler support.

Limitations and Outlook

Principle limitations relate to analyzing think aloud protocols. Generally, it is assumed that thinking aloud does not interfere with thought processes—but as the modeling task includes a visual, non-verbal perceptual component, thinking aloud may slow down the thought processes and/or the modeling performance (Ericsson and Simon 1980). It is important to note that the scope of this study limits findings to conceptual *data* modeling processes—but this limitation is not by principle and the study design could be applied to object-oriented modeling and process modeling as well. Also, please note that we recruited all participants from one university. In future studies, we plan to complement the present study with follow-up studies observing subjects with various backgrounds, e.g., regarding prior modeling experience, and observing not only data modeling processes but also, e.g., process modeling processes—with the overarching aim to integrate all findings in a subsequent step elaborating on similarities and differences in data modeling, object-oriented modeling, and process modeling.

Acknowledgements

We would like to express our gratitude to Benjamin Ternes, Hagen Barth and Malika Omeragic for their support during data collection and to the eight participants. We would also like to thank the associate editor and reviewers for their constructive comments which helped us to improve the present article.

Appendix

As part of a project for the introduction of a new information system in the university library, you are asked to create a conceptual data model that reconstructs the following facts representing a simplified description of a library:
• The current stock of the university library includes library items (e.g. books). Library items are described by title and year of publication. In addition, the international standard book number (ISBN) is recorded for unique identification.
• There may be one or more copies of a library item. Each copy of a library item has a unique shelfmark. To determine the age of the library's holdings, the acquisition date is recorded for each copy. In order to be able to identify particularly valuable items, the acquisition price is also recorded for each copy.
• When registering as user, the first name and last name are recorded. The date of birth is also recorded in order to comply with any age restrictions.
• Users of the university library can borrow any number of copies of library items. A loan always refers to exactly one copy. For a loan, the date of the loan and the due date are recorded in order to be able to determine if the loan period is overdue.

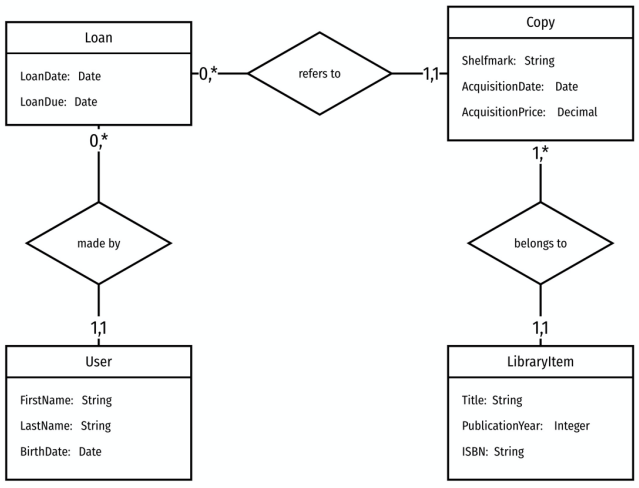

Figure 4. Main modeling task (translated from German into English, on the left) and reference solution (on the right)

References

Alavi, M., and Leidner, D. E. 2001. "Research Commentary: Technology-Mediated Learning–A Call for Greater Depth and Breadth of Research," *Information Systems Research* (12:1), pp. 1–10.

ACM/AIS. 2018. "Curricula Recommendations". (https://www.acm.org/education/curricula-recommendations/).

Batra, D. 1993. "A Framework for Studying Human Error Behavior in Conceptual Database Modeling," *Information & Management* (25:3), pp. 121–131.

Batra, D. 2007. "Cognitive Complexity in Data Modeling: Causes and Recommendations," *Requirements Engineering* (12:4), pp. 231–244.

Batra, D., and Antony, S. R. 1994. "Novice Errors in Conceptual Database Design," *European Journal of Information Systems* (3:1), pp. 57–69.

Batra, D., and Davis, J. G. 1992. "Conceptual Data Modelling in Database Design: Similarities and Differences between Expert and Novice Designers," *International Journal of Man-Machine Studies* (37:1), pp. 83–101.

Bazeley, P. 2012. "Integrative Analysis Strategies for Mixed Data Sources," *American Behavioral Scientist* (56:6), pp. 814–828.

Bera, P. 2011. "Situations That Affect Modelers' Cognitive Difficulties: An Empirical Assessment," in *5th Americas Conference on Information Systems (AMCIS)*, Detroit, MI, Paper 254.

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

Biggs, J. 1996. "Enhancing Teaching through Constructive Alignment," *Higher Education* (32:3), pp. 347–364.

Burton-Jones, A., and Meso, P. N. 2006. "Conceptualizing Systems for Understanding: An Empirical Test of Decomposition Principles in Object-Oriented Analysis," *Information Systems Research* (17:1), pp. 38–60.

Burton-Jones, A., and Meso, P. N. 2008. "The Effects of Decomposition Quality and Multiple Forms of Information on Novices' Understanding of a Domain from a Conceptual Model," *Journal of the Association for Information Systems* (9:12), pp. 748–802.

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

Claes, J., Vanderfeesten, I., Pinggera, J., Reijers, H. A., Weber, B., and Poels, G. 2014. "A Visual Analysis of the Process of Process Modeling," *Information Systems and E-Business Management* (13:1), pp. 147–190.

Claes, J., Vanderfeesten, I., Gailly, F., Grefen, P., and Poels, G. 2015. "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), pp. 1401–1425.

Claes, J., Vanderfeesten, I., Gailly, F., Grefen, P., and Poels, G. 2017. "The Structured Process Modeling Method (SPMM) What Is the Best Way for Me to Construct a Process Model?" *Decision Support Systems* (100), pp. 57–76.

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

Elmasri, R., and Navathe, S. 2017. *Fundamentals of Database Systems*, (7th ed.), Boston, MA: Pearson.

Ericsson, K. A., and Simon, H. A. 1980. "Verbal Reports as Data," *Psychological Review* (87:3), pp. 215–251.

Ericsson, K. A., and Simon, H. A. 1993. *Protocol Analysis: Verbal Reports as Data*, (3rd ed.), Cambridge, MA: MIT Press.

Gregor. 2006. "The Nature of Theory in Information Systems," *MIS Quarterly* (30:3), pp. 611–642.

Haisjackl, C., Barba, I., Zugal, S., Soffer, P., Hadar, I., Reichert, M., Pinggera, J., and Weber, B. 2016. "Understanding Declare Models: Strategies, Pitfalls, Empirical Results," *Software & Systems Modeling* (15:2), pp. 325–352.

Hirschheim, R. A., Klein, H.-K., and Lyytinen, K. 2008. *Information Systems Development and Data Modeling: Conceptual and Philosophical Foundations*, Cambridge, UK: Cambridge University Press.

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

Hoppenbrouwers, S. J. B. A., Lindeman, L., and Proper, H. A. 2006. "Capturing Modeling Processes: Towards the MoDial Modeling Laboratory," in *On the Move to Meaningful Internet Systems 2006: On OTM 2006 Workshops*, R. Meersman, Z. Tari and P. Herrero (eds.), Berlin, Heidelberg: Springer, pp. 1242–1252.

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

Mendling, J., Strembeck, M., and Recker, J. 2012. "Factors of Process Model Comprehension—Findings from a Series of Experiments," *Decision Support Systems* (53:1), pp. 195–206.

Miles, M. B., Huberman, A. M., and Saldaña, J. 2014. *Qualitative Data Analysis: A Methods Sourcebook*, (3rd ed.), Thousand Oaks, CA: Sage.

Newell, A., and Simon, H. A. 1972. *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall.

Nielsen, J. 1994. "Estimating the Number of Subjects Needed for a Thinking Aloud Test," *International Journal of Human-Computer Studies* (41:3), pp. 385–397.

Pastor, Ó., España, S., and Panach, J. I. 2016. "Learning Pros and Cons of Model-Driven Development in a Practical Teaching Experience," in *Advances in Conceptual Modeling, ER 2016 Workshops*, S. Link and J. C. Trujillo (eds.), Cham: Springer, pp. 218–227.

Pinggera, J., Soffer, P., Fahland, D., Weidlich, M., Zugal, S., Weber, B., Reijers, H. A., and Mendling, J. 2015. "Styles in Business Process Modeling: An Exploration and a Model," *Software & Systems Modeling* (14:3), pp. 1055–1080.

Pinggera, J., Zugal, S., Furtner, M., Sachse, P., Martini, M., and Weber, B. 2014. "The Modeling Mind: Behavior Patterns in Process Modeling," in *Enterprise, Business-Process and Information Systems Modeling*. Lecture Notes in Business Information Processing, Vol. 175. I. Bider, K. Gaaloul, J. Krogstie, S. Nurcan, H. A. Proper, R. Schmidt and P. Soffer (eds.), Berlin, Heidelberg: Springer, pp. 1–16.

Polanyi, M., and Sen, A. 2009. *The Tacit Dimension*, Chicago, IL, London: University of Chicago Press.

Pretz, J. E., Naples, A. J. and Sternberg, R. J. 2003. "Recognizing, defining, and representing problems," in *The Psychology of Problem Solving*. J. E. Davidson and R. J. Sternberg (eds.) Cambridge, UK: Cambridge University Press, pp. 3–30.

Rosenthal, K., Ternes, B., and Strecker, S. 2019 "Learning Conceptual Modeling: Structuring Overview, Research Themes and Paths for Future Research," in 27th *European Conference on Information Systems (ECIS)*, Stockholm, Sweden, Research Paper 137.

Schoonenboom, J., and Johnson, R. B. 2017. "How to Construct a Mixed Methods Research Design," *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* (69), pp. 107–131.

Sedrakyan, G., and Snoeck, M. 2017. "Cognitive Feedback and Behavioral Feedforward Automation Perspectives for Modeling and Validation in a Learning Context," *Communications in Computer and Information Science* (692), pp. 70–92.

Sedrakyan, G., Snoeck, M., and De Weerdt, J. 2014. "Process mining analysis of conceptual modeling behavior of novices – empirical study using JMermaid modeling and experimental logging environment," *Computers in Human Behavior* (41), pp. 486–503.

Sedrakyan, G., De Weerdt, J., and Snoeck, M. 2016. "Process-Mining Enabled Feedback: 'Tell Me What I Did Wrong' vs. 'Tell Me How to Do It Right'," *Computers in Human Behavior* (57), pp. 352–376.

Serral, E., De Weerdt, J., Sedrakyan, G., and Snoeck, M. 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)*, Grenoble, France, pp. 1–6.

Song M., and van der Aalst W. M. P. 2007. "Supporting process mining by showing events at a glance," in *17th Annual Workshop on Information Technologies and Systems (WITS)*, Montreal, Canada, pp. 139-145

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

Sweller, J. 1988. "Cognitive Load during Problem Solving: Effects on Learning," *Cognitive Science* (12:2), pp. 257–285.

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

Ternes, B., Strecker, S., Rosenthal, K., and Barth, H. 2019 "A browser-based modeling tool for studying the learning of conceptual modeling based on a multi-modal data collection approach," in *14th International Conference on Wirtschaftsinformatik (WI)*, Siegen, Germany, pp. 1998–2002.

Venable, J. R. 1996. "Teaching Novice Conceptual Data Modellers to Become Experts," in *International Conference Software Engineering: Education and Practice*, Dunedin, New Zealand: IEEE Computer Society Press, pp. 50–56.

Venkatesh, V., Brown, S. A., and Bala, H. 2013. "Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems," *MIS Quarterly* (37:1), pp. 21–54.

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

VERBI Software. 2018. "MAXQDA Standard 12," Berlin, Germany: VERBI. (https://www.maxqda.com/).

Vessey, I., and Conger, S. A. 1994. "Requirements Specification: Learning Object, Process, and Data Methodologies," *Communications of the ACM* (37:5), pp. 102–113.

Wand, Y., and Weber, R. 2002. "Research Commentary: Information Systems and Conceptual Modeling—A Research Agenda," Information Systems Research (13:4), pp. 363–376.

Wilmont, I., Brinkkemper, S., van de Weerd, I., and Hoppenbrouwers, S. 2010. "Exploring Intuitive Modelling Behaviour," in *Enterprise, Business-Process and Information Systems Modeling*. Lecture Notes in Business Information Processing, Vol. 50, I. Bider, T. Halpin, J. Krogstie, S. Nurcan, E. Proper, R. Schmidt, and R. Ukor (eds.), Berlin, Heidelberg: Springer, pp. 301–313.

Wilmont, I., Hengeveld, S., Barendsen, E., and Hoppenbrouwers, S. 2013. "Cognitive Mechanisms of Conceptual Modelling, How Do People Do It?" in *32nd International Conference on Conceptual Modeling (ER)*, Hong Kong, China: Springer, pp. 74–87.

Wilmont, I., Hoppenbrouwers, S., and Barendsen, E. 2017. "An Observation Method for Behavioral Analysis of Collaborative Modeling Skills," in *Advanced Information Systems Engineering Workshops. CAiSE 2017*. Lecture Notes in Business Information Processing, Vol. 286. A. Metzger and A. Persson (eds.), Cham: Springer, pp. 59–71.

Zugal, S., Haisjackl, C., Pinggera, J., and Weber, B. 2013. "Empirical Evaluation of Test Driven Modeling," *International Journal of Information System Modeling and Design* (4:2), pp. 23–43.