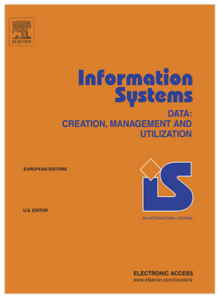
Accepted Manuscript



Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study

Michaël Verdonck, Frederik Gailly, Robert Pergl, Giancarlo Guizzardi, Beatriz Martins, Oscar Pastor

PII: S0306-4379(18)30372-7
DOI: https://doi.org/10.1016/j.is.2018.11.009
Reference: IS 1353

To appear in: *Information Systems*

Received date : 13 July 2018
Revised date : 15 October 2018
Accepted date : 21 November 2018

Please cite this article as: M. Verdonck, F. Gailly, R. Pergl et al., Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study, *Information Systems* (2018), https://doi.org/10.1016/j.is.2018.11.009

Comparing traditional conceptual modeling with ontology-driven conceptual modeling: an empirical study

Michaël Verdonck[1], Frederik Gailly[1], Robert Pergl[2], Giancarlo Guizzardi[3], Beatriz Martins[4] and Oscar Pastor[4]

[1] Faculty of Economics and Business Administration, Ghent University, Belgium; [2] Faculty of Information Technology, Czech Technical University Prague, Czech; [3] Free University of Bozen Bolzano, Italy; [4] ProS Research Center, Universitat Politècnica de València; Spain

{michael.verdonck, frederik.gailly}@ugent.be; perglr@fit.cvut.cz; giancarlo.guizzardi@unibz.it; bmartins@pros.upv.es; opastor@dsic.upv.es

Abstract. This paper conducts an empirical study that explores the differences between adopting a traditional conceptual modeling (TCM) technique and an ontology-driven conceptual modeling (ODCM) technique with the objective to understand and identify in which modeling situations an ODCM technique can prove beneficial compared to a TCM technique. More specifically, we asked ourselves if there exist any meaningful differences in the resulting conceptual model and the effort spent to create such model between novice modelers trained in an ontology-driven conceptual modeling technique and novice modelers trained in a traditional conceptual modeling technique. To answer this question, we discuss previous empirical research efforts and distill these efforts into two hypotheses. Next, these hypotheses are tested in a rigorously developed experiment, where a total of 100 students from two different Universities participated. The findings of our empirical study confirm that there do exist meaningful differences between adopting the two techniques. We observed that novice modelers applying the ODCM technique arrived at higher quality models compared to novice modelers applying the TCM technique. More specifically, the results of the empirical study demonstrate that it is advantageous to apply an ODCM technique over an TCM when having to model the more challenging and advanced facets of a certain domain or scenario. Moreover, we also did not find any significant difference in effort between applying these two techniques. Finally, we specified our results in three findings that aim to clarify the obtained results.

Keywords: Conceptual modeling; empirical study; ontology-driven conceptual modeling; experiment; OntoUML, Entity-relationship modeling

1 Introduction

Modeling, in all its various forms, plays an important role in representing and supporting complex human design activities. Especially in the development of information systems, their analysis, as well as in re-engineering, modeling has proved to be an essential element in achieving high performing information systems (Karimi, 1988). Conceptual models were introduced to increase understanding and communication of a system or domain among stakeholders. Some commonly used conceptual modeling techniques and methods include: Business Process Model and Notation (BPMN), entity relationship modeling (ER), object-role modeling (ORM), and the Unified Modeling Language (UML). We refer to these techniques and methods as conceptual modeling. Many of these early conceptual modeling techniques however lacked an adequate specification of the semantics of the terminology of the underlying models, leading to inconsistent interpretations and uses of knowledge (Gruninger, Atefi, & Fox, 2000). In order to overcome such issues, ontologies can be applied. Ontologies can be described as a set of things whose existence is acknowledged by a particular theory or system of thought (Honderich, 2006). Research on ontologies has become increasingly widespread in the computer science community, gaining importance in research fields such as knowledge engineering (Uschold & Gruninger, 1996), knowledge representation (Sowa, 1999) and information modeling (Ashenhurst, 1996). More specifically in the field of conceptual modeling, ontologies can be applied to aid in the reasoning on the contents of a conceptual model (Corea & Delfmann, 2017; Karagiannis & Buchmann, 2018) or to articulate and formalize the conceptual modeling grammars needed to describe the structure and behavior of the modeled domain (Wand & Weber, 1993). As for the scope of this paper, we will focus on the latter, where ontologies are applied to improve conceptual modeling languages by means of constraints and structuring rules to existing languages. Ontologies differ from conceptual models in several ways. For instance, while ontologies use an open-world assumption, a conceptual model embodies a closed-world assumption (Atkinson, Gutheil, & Kiko, 2006). Furthermore, ontologies can be characterized as descriptive, domain-relevant and static, while conceptual models idiosyncratically are more system-focused, do not require any shared understanding, nor do they model the whole domain (Henderson-Sellers, 2012).

As such, we can describe the utilization of these ontological theories, coming from areas such as formal ontology, cognitive science and philosophical logics, to develop engineering artifacts (e.g. modeling languages, methodologies, design patterns and simulators) for improving the theory and practice of conceptual modeling, as ontology-driven conceptual modeling (ODCM) (Guizzardi et al. 2015). On the other hand, the more ‘traditional’ form of conceptual modeling is generally described as the activity of representing aspects of the physical and social world for the purpose of communication, learning and problem solving among human users (Mylopoulos, 1992). Thus, in this paper we regard traditional conceptual modeling (TCM) as the utilization of modeling techniques and languages (e.g. BPMN, UML, EER) according to the modeling standards and specifications in which they have been introduced and proposed. In contrast, we regard ODCM as the utilization of these conceptual modeling techniques, but with the important difference that they are extended or supported by ontological theories that further articulate and formalize the conceptual modeling grammars of these languages.

The benefits of ODCM are presumed to be the most substantial when applied for the design, analysis and re-engineering of rather large and complex information systems. Their use would lead to various system engineering benefits such as increased re-usability and reliability (Uschold & Gruninger, 1996). Additionally, ODCM would aid in the development of a more sophisticated representation of the domain being modeled, and a higher level of domain understanding by its modelers and users (Gemino & Wand, 2005). These benefits can be obtained through adopting a wide variety of techniques and practices. For example, ontologies were used for the development of new conceptual modeling languages (Opdahl, Berio, Harzallah, & Matulevičius, 2012), for adding structuring rules to existing languages (Evermann & Wand, 2005b), and for proposing conceptual modeling patterns and anti-patterns (Guizzardi, 2014). Several techniques in ODCM go further than only evaluating or supporting a conceptual modeling technique, by providing a proper conceptual modeling technique themselves, as such adopting modelers to an ontological way of thinking. Such techniques (e.g. OntoUML, O^3) are often founded on existing modeling notations and enhance the metamodel of this notation by incorporating formal ontological constraints that correspond to the ontology's axiomatization. By applying these ontology-driven techniques, modelers are encouraged to reason more carefully about the domain and the classification of its elements in terms of ontological concepts (e.g. inferring types, properties, events etc.) and their underlying relationships.

However, while many of these ontology-driven techniques have demonstrated to be beneficial compared to the traditional conceptual modeling practices, the added value of their application is not always straightforward and there is no clear distinction when it is actually desirable to adopt these techniques. Understanding the philosophical concepts and structures of an ontology (e.g. theory of parthood, types and instantiations, identity, dependency, unity etc.) requires time and encompasses a certain degree of complexity. As asserted by (Guizzardi, Das Graças, & Guizzardi, 2011), this complexity posed a stark issue for novice modelers who were using the ontologically founded conceptual modeling language OntoUML – note: by novice we refer to modelers that have limited experience in conceptual modeling and have not yet developed a holistic understanding of adopting a particular modeling technique to translate problem descriptions into standard abstractions (Shanks, 1997).

Additionally, while it is generally assumed that ontology-based modeling can indeed enhance the development of a conceptual model, the study of (Soffer & Hadar, 2007) obtained less promising results, acknowledging that the overall effect of ontology-based modeling rules were not significant. They observed that the utilization of ontologies does not significantly improve model variation, which according to them means that ontologies do not sufficiently support the decision making during the conceptual modeling process. Thus, while ODCM aids in the construction of ontologically sound conceptual models, it appears that it also can increase the complexity of developing a conceptual model. This observation causes us to question the benefit of adopting an ODCM technique over a TCM technique. As to the knowledge of the authors, no research efforts have yet been made that assess the actual impact of adopting an ODCM technique compared to that of a TCM technique.

Therefore, it is the goal of this paper to conduct a study that investigates and compares the differences between TCM and ODCM. More specifically, we would like to differentiate between modelers that are trained in a TCM approach and modelers that have been taught an ODCM approach. These two groups of modelers then have to model a scenario that encompasses certain modeling challenges. Through our study, we then compare the two modeling approaches by investigating the quality of the resulting conceptual models, and the amount of effort a modeler had to spend in order to compose these models. To properly measure these effects, we intend to conduct an empirical study. Therefore, as the foundation for the further development of this paper, we formulate our research question as follows: *Are there meaningful differences in the resulting conceptual model and the effort spent to create such model between novice modelers trained in an ontology-driven conceptual modeling technique and novice modelers trained in a traditional conceptual modeling technique*. In section 2 of this paper, we formulate our testing hypothesis and meanwhile discuss previous related empirical research. Next, we draft our experimental design to test these hypotheses in section 3. We then present the results of our experiment in section 4 and discuss their outcome on the hypothesis. Next, in section 5, we interpret the results of our experiment, and discuss their consequences and implications. In section 6 we present our conclusion and future research opportunities. Finally, in section 7 we discuss the internal and external validity of this paper.

2 Hypothesis development

Based upon our research question, we formulate our testing hypotheses. In order to do so properly, we first investigate the different kinds of empirical studies that have been performed in the field and take a closer look at earlier comparisons between ODCM and TCM.

Over the years, the adoption of ontologies and ODCM as a modeling practice materialized steadily and in different trends or phases. Originally, ontologies were introduced in the field of conceptual modeling as a way to *evaluate the ontological soundness of a conceptual modeling language*. Or in other words, ontologies aimed to improve conceptual modeling languages by evaluating (1) Domain Appropriateness, i.e. the expressivity of a modeling language and its adequacy in being truthful to a set of phenomena in reality and (2) Comprehensibility Appropriateness, i.e. how easy it is for users of the language to unambiguously recognize the phenomena being represented by the language's models and primitives.

Ontologies proved quite useful in assessing whether different conceptual modeling procedures are likely to lead to good representations of real-world phenomena. Therefore, the first empirical research efforts concerning ODCM and TCM examined whether the semantic analysis offered by ontologies actually benefited the grammars of conceptual modeling languages. For example the paper of (Recker, Rosemann, Green, & Indulska, 2011b) investigated how users of the BPMN conceptual modeling grammar perceived existing ontological deficiencies, and how ontologies could aid in identifying such deficiencies. Other empirical studies such as (Poels, Gailly, Maes, & Paemeleire, 2005; Shanks, Tansley, Nuredini, Tobin, & Weber, 2008) performed similar research, where they measured the existing ontological deficiencies of conceptual modeling languages, studied their impact on users' perceptions and how techniques involving ontologies were applied to analyze such languages to identify these deficiencies.

Gradually however, a second trend for the usage of ontologies emerged, in the sense that an *ontology would express the fundamental elements of a domain*, and therefore becoming the theoretical foundations of a conceptual modeling language (Guarino, 1998). This new way of applying ontologies led to a growing interest in the role that they can fulfill in the improvement of conceptual modeling languages (Opdahl et al., 2012), by adding structuring rules to existing languages (Evermann & Wand, 2005a), and by proposing conceptual modeling patterns and anti-patterns (Ruy et al. 2017). In accordance, various empirical studies examined these enhanced ways of conceptual modeling. For instance, the study of

(Bera, 2012) tested the effect of ontological modeling rules on the development of conceptual models. Their results revealed that modelers face cognitive difficulties when developing conceptual models and that ontological modeling rules can alleviate these difficulties. Additionally, the study indicated that ontological rules could help modelers to commit fewer modeling errors and help them to develop conceptual models in a systematic way. Other studies such as (Evermann & Wand, 2006; Recker et al. 2011a) reached similar conclusion, supporting the use of ontological theories and rules in conceptual modeling. However, not all studies acknowledged the same results. For instance, (Soffer & Hadar, 2007) performed an explorative study to investigate the effect of applying ontological modeling rules to the modeling process on model variations. More specifically, their results expressed that difficulties were experienced in the adoption of the ontological concepts and rules underlying an ontology, especially with large sets of these rules. As a conclusion, they expressed their belief that further improvements may be achieved by adopting modelers to an ontological way of thinking, learning them to perceive and interpret the world in ontological concepts.

This is where the third trend in ODCM aims to deliver a solution, in the form of not only evaluating or supporting a conceptual modeling technique, but instead by evolving into a proper conceptual modeling technique itself, as such adopting modelers to the ontological way of thinking. These new techniques are often founded on existing modeling notations and enhance the metamodel of this notation by incorporating formal ontological constraints that correspond to the ontology's axiomatization. Examples of these new techniques are OntoUML and the O^3 language. The O^3 language (Pastor & Molina, 2007) can be considered as a natural language that fuses various ontological concepts based upon the Bunge-Wand-Weber (BWW) ontology together with the object-oriented paradigm, with the purpose to facilitate automatic development of information system applications. OntoUML (Guizzardi, 2005) on the other hand, is a modeling language that reflects the ontological distinctions prescribed by UFO (Unified Foundational Ontology) by incorporating the axiomatization of the UFO ontology by means of formal constraints in the UML metamodel. With these techniques, modelers adopt an ontological way of thinking, by learning them to perceive and interpret the world in ontological concepts and rules. However, this requires a modeler to understand the philosophical elements and structures from the underlying ontology, where its formal axiomatization and constructs can pose a significant challenge to novice modelers (Guizzardi et al., 2011).

Hence, it would seem that the ODCM technique at the one hand facilitates the development of ontologically sound conceptual models, while on the other hand it appears this practice can increase the complexity of developing a conceptual model. However – as to the knowledge of the authors – no empirical research has yet measured the actual impact of adopting an ODCM technique to develop a conceptual model and observe the resulting models and effort to create these models. Furthermore, no research study has yet compared the difference in modeling between ODCM and TCM techniques. Most of the empirical studies described above did compare ODCM to TCM, although this comparison was often either partial or incomplete, meaning that only certain aspects of an ontology or a limited set of ontological concepts or rules were compared. Additionally, subjects were either briefly introduced to the ontology or received only minor training in applying the ontology in the process of conceptual modeling. This results in modelers that are not fully competent with the respective ontology. It is our perception that modelers should be more intensively trained into an ontological way of thinking, by learning them to perceive and interpret the world in ontological concepts.

As such, the objectives of our study are (1) to have a complete comparison between an OCDM and a TCM technique, meaning that both techniques are taught in their full scope, and not only certain aspects of it; (2) to compare subjects that have been properly trained in both techniques, over a period of several months; (3) to require subjects not only to comprehend these techniques, but also have them apply the technique in order to construct a conceptual model and (4) to compare both the resulting models of each technique, as well as the effort required to construct these models. As the advantages of ODCM are presumed to be the most beneficial when applied to rather large and complicated modeling tasks and designs, we assume ODCM to deliver better results when applied to a more complex modeling task. Thus, based upon previous research efforts, and the assumptions given above, we formulate our hypotheses as followed:

1. *Novice modelers applying an ODCM technique will arrive at higher quality models compared to novice modelers applying a TCM technique – given a thorough understanding of the respective technique and a sufficiently complex modeling task.*
2. *Novice modelers applying an ODCM technique will experience more effort in the process of developing a conceptual model compared to novice modelers applying a TCM technique – given a thorough understanding of the respective technique and a sufficiently complex modeling task.*

In other words, we believe that the more complex a modeling task becomes, the more semantically correct conceptual models will be when adopting an ODCM technique over an TCM technique. However, we do expect that adopting an ODCM approach will also require more effort compared to a TCM approach. Finally, this experiment will also compare the model quality between both techniques. While many definitions and approaches exist to measure model quality, we believe that the quality of a model cannot be detached from the purpose it intends to fulfill. Therefore, we will measure model quality by the degree in which a model fulfills its purpose – which we will refer to as effectiveness. The next section further specifies how we set up our experimental design, based upon these hypotheses.

3 Experiment Design

A careful planning and design prepares for how the experiment is conducted and is essential in achieving validated experimental results. Due to the lack of a random assignment of subjects between our testing groups (infra Section 3.3), we would like to emphasize that we perform a quasi-experiment since key characteristics between subject treatments may differ. As such, when referring to the term 'experiment' in the further development of this chapter, we refer to a quasi-experiment. We base ourselves upon the experimental design described in Wohlin et al. (2012), where the design of an experiment can be divided into several steps. Based upon our hypothesis, the selection of the independent and dependent variables takes place. Next, the selection of subjects is carried out. The experiment design type is chosen based on the hypothesis and variables selected. Next the instrumentation prepares for the practical implementation of the experiment. Finally, the validity evaluation aims at checking the validity of the experiment. After the planning process is iterated, we can conduct the actual experiment, and collect the data in order to either accept or reject the testing hypotheses.

3.1 Variable development

Before any experimental design, the dependent, independent and control variables should be selected beforehand. Both the independent and dependent variables are derived from our hypotheses, and consequently from our research question.

Independent Variable

In our study, the independent or affecting variable constitutes of the two different modeling techniques or approaches our subjects can apply to construct a conceptual model. In other words, in our experimental setting we can control if we either assign our test subjects with a traditional modeling technique or with an ontology-driven technique. More specifically, we compare the enhanced entity relationship (EER) modeling technique against the ontology-driven OntoUML modeling technique. The entity-relationship (ER) approach – initially proposed by Chen (1976) – still remains the premier model for conceptual design (Fettke, 2009). It is used to represent information in terms of entities, their attributes, and associations among entity occurrences called relationships. The EER modeling technique can be applied in combination with several notations. The UML notation – more specifically class diagram notation – is a widely accepted notation, both in academics (Elmasri & Navathe, 2015) as in practice by analysts and software developers (Gornik & IBM, 2003). By enhancing the EER approach with the UML notation, the conceptual model gains significant benefits, including easier communication and a more truthful representation of a particular domain. Similarly, OntoUML is a well-known technique in the domain of ODCM and has been frequently adopted for various purposes. Additionally, OntoUML also applies the UML notation – class diagrams – but with the UFO ontology as an underlying foundational theory. More specifically, the purpose of OntoUML is to provide sound ontological foundations for various domains (domain appropriateness) and conceptual clarity (comprehensibility appropriateness) of modeling languages (Guizzardi & Wagner, 2005). As such, both techniques have been primarily developed to deliver conceptual models that offer sound representations of a particular domain. Additionally, both techniques apply the same UML notation, but are grounded in two different underlying theories – the EER approach and the UFO ontology.

Dependent variables

The purpose of our experiment is to measure the difference of the resulting conceptual model – i.e. the model quality – and the effort required to create such a model, when applying either a traditional modeling technique or an ontology-driven modeling technique. Therefore, to properly measure and compare such differences, we rely on the research of (Grüninger & Fox, 1995; Krogstie, 2012; Moody, 2003), where we make a distinction between the *effectiveness and efficiency* of our two techniques. While effectiveness is defined on how well a particular technique achieves its objectives, efficiency is viewed as the effort required to apply the technique. The former can be measured by output measures evaluating the quantity and/or quality of the results; the latter can be measured by a variety of input measures such as time, cost or perception.

Effectiveness

We are going to measure the effectiveness of the TCM and OCDM methods by evaluating the quality of the resulting models created by the participants. In the context of software engineering, quality is often described as the fitness for purpose. Therefore, we measure model quality by the degree in which the participants' models fulfill their purpose. Since both TCM and ODCM have been developed to represent a domain and its truthfulness to reality, we evaluate the resulting conceptual model in its capacity to represent a domain as truthfully as possible. In order to have a domain that is recognizable for our participants to model – all our participants are students – we have opted for the domain of a university.

In order to objectively assess the suitability and truthfulness of a model to represent a domain, we rely on the use of competency questions. Originally, competency questions were applied in ontology development (Grüninger & Fox, 1995), where a particular ontology was found adequate to represent a certain domain providing that the ontology could represent and answer a specific set of competency questions. In our experiment, we construct several domain requirements that are defined in a set of competency questions, to which the resulting conceptual models should be able to provide an answer in order to be deemed a good representation of the domain. Furthermore, we differentiate between two sets of competency questions. One set of questions measures if subjects adequately represented the domain as described in the assignment. The second set of questions measures if subjects were able to deal with certain 'complications' described in the case, which required subjects' to improve their model beyond the literal description given in the case. This corresponds to the work of (Daga et al., 2005; de Cesare & Partridge, 2016), where they make a distinction between competency questions that measure Content Interpretation (CI) and Content Sophistication (CS). While the former is defined as the identification of the entities that exists in the domain by an applicant or modeler, the latter can be seen as the process of gradually improving the model such that it provides a more precise representation of the world. Tailored to our experiment, participants receive a case that describes the university domain. When modeling the domain, they have to identify the necessary constructs, relationships and cardinalities that govern this domain – i.e. content interpretation. However, the case (deliberately) contains ambiguous descriptions or certain complications. Content sophistication can then take place if a participant responds by improving their model so that it provides a more precise representation of the university domain – and overcome these complications or ambiguities. As such, the competency questions allow us to evaluate the participants' models in a rather objective way, by distinguishing between the 'completeness' of the model (i.e. content interpretation), and the more innovative aspects of their models (i.e. content sophistication). These competency questions are adopted by the authors to assign scores to the models of the participants.

Efficiency

Based upon previous findings in the literature such as the one from (Soffer & Hadar, 2007), we expect that modelers adopting to an ontological way of thinking, and perceiving and interpreting the world in ontological concepts and rules require more effort – and hence achieve a lower efficiency — compared to modelers adopting a TCM technique, since they do not have to concern themselves with such rules and ontological concepts.

The efficiency of the ontology-driven models is measured by: (1) assessing the amount of time needed to develop the models, and (2) assessing the usage beliefs of each modeling technique. More specifically, we measure perceived usefulness and perceived ease of use, as introduced by the Technology Acceptance Model (TAM), and which are key to understanding modeling usage beliefs (Davies, Green, Rosemann, Indulska, & Gallo, 2006). Similar to (Moody, 2003), we will be measuring the effectiveness and efficiency of a modeling technique instead of a technology or tool. The TAM model is a popular information systems theory that describes how users come to accept and use information system artifacts. The model suggests that when users are presented with such an artifact, two primary factors influence their decision about how and when they will use it: (1) *Perceived ease of use*, which is determined by the degree to which a person believes that using a particular technique would be free of effort; and (2) *Perceived usefulness*, referring to the degree in which a person believes that a technique will be effective in achieving the intended modeling objective. Perceived usefulness can therefore also be seen as a way to measure the actual effectiveness of the technique (Moody, 2003), but since it is determined by its perceived ease of use we categorize it under efficiency. In our experiment, we will apply this model and the associated measures to measure how our subjects will perceive the TCM and ODCM modeling techniques. More specifically, participants have to answer several questions after completing the modeling task – using multiple-item scales, with five-point Likert scales – which measures both the perceived usefulness and the perceived ease of use. The reliability and validity of these questions has already been proven in several research efforts (Davis, 1989; Recker et al., 2011a).

Control Variables

Since we test participants modeling with a TCM and an ODCM technique, we need to ascertain that all subjects have an equal understanding of each technique they are modeling with. Therefore, we apply a control variable to test every subject's knowledge and understanding of the modeling technique, before the start of the experiment. The results from the subjects that failed the knowledge test are not incorporated into the results of the experiment. Next, to provide a complex enough modeling case as required in our hypotheses, we have selected a modeling case that served as an assignment of a modeling course given at the University Ghent. The feedback and the final results of the assignment that applied the modeling case confirmed that the modeling case is of a rather complex degree. Additionally, we have presented the modeling case at the OntoCom workshop at the 36th International Conference on Conceptual Modeling. During this workshop, the case has been given to several experts in the domain of conceptual modeling and ontology. Each of these experts have then created a conceptual model – often also based upon an ontological theory – according to their interpretation of the case. Afterwards, the different models were discussed for their completeness and how they dealt with the challenges or ambiguities that could be found in the case. During this workshop, many of the competency questions for both the content interpretation and especially the content sophistication were derived from the models of the workshop and the feedback from the different experts. Additionally, the experts who have modeled the case themselves also have labeled the case as sufficiently complex to be applied in an experimental setting.

3.2 Subject Selection

The subjects in our study all were novice conceptual modelers and were attending two different courses on conceptual modeling at the University of Ghent (Belgium) and the Technical University of Prague (Czech). While the subjects at the University of Ghent were taught how to adopt a TCM technique to construct a conceptual model, the course at the Technical University of Prague taught their students the ODCM technique. As stated by (Falessi et al., 2017), using students as participants remains a valid simplification of reality needed in laboratory contexts. It is an effective way to advance software engineering theories and technologies but, like any other aspect of study settings, should be carefully considered during the design, execution, interpretation, and reporting of an experiment. Consequently, we decided to select students as our test subjects since they have no prior knowledge of conceptual modeling and can thus be seen as novice modelers who can be trained in either TCM or ODCM. Hence, our selection of students enabled us to train subjects without having prior experience in another modeling technique. Consequently, we could measure the full impact of the modeling technique that is being taught.

At Ghent University, students have been taught the EER conceptual modeling technique through both theoretical classes and practical sessions. In these practical sessions, students were required to solve modeling assignments of certain scenarios. Additionally, students were required to submit a rather extensive group assignment, where they had to design and implement an information system. An important aspect of this assignment was to develop a sound EER conceptual model that forms the foundation of their database. Similarly, students at the Czech Technical University in Prague (CTU) received both theoretical classes as well as practical sessions on a weekly basis. Furthermore, they also had to complete a work assignment that required them to create sound OntoUML models, to serve as a foundation for a software system specification. Thus, both students at Ghent University and the CTU received assignments on a regular basis that required them to apply conceptual modeling for different scenario’s and domains. These assignments were evaluated and marked for their correctness, and students also had to complete an exam on conceptual modeling to evaluate their overall progress.

Moreover, all subjects have the same age (i.e. early-twenties) and the majority of our subjects have a business/technical-oriented background. Concerning motivation, students were asked to participate with the experiment out of self-interest and as an opportunity to improve their skills in conceptual modeling. There was no reward-based incentive. As such, students who participated in our experiment were essentially self-motivated based on the inclination to learn more and to improve their skillset. Thus, the specific selection and the education program leads to a controlled sample of subjects, all being novice modelers, who are properly trained in the respective modeling technique and with no prior knowledge of any other modeling technique.

Finally, in order to determine the number of subjects for our empirical study, we base ourselves on the differences in the averages in the model comprehension scores from the study of (Verdonck & Gailly, 2016). Based upon the sample size formula of Shao, Wang, and Chow (2008), assuming a Type I error (α) of 5% and a Power ($1-\beta$, where β is Type II error) of 0.8, we require a total number of 43 subjects per treatment group. In total 100 subjects participated in the study, of which 50 in each treatment. Hence, the number of participants in our experiment is sufficient with regard to the required statistical minimum.

3.3 Experimental Design Type

An experiment consists of a series of tests of different treatments (Wohlin et al., 2012). To get the desired results to answer our research question, the series of tests must be carefully planned and designed. Based on our hypotheses, we can derive two treatments: an UML treatment and an OntoUML treatment. The assignment in each treatment constitutes of a case study that has to be modeled by the participants of the respective treatment. We have assigned the participants to these treatments according to the balancing design principle. By balancing the treatments, we assign an equal number of subjects to each separate treatment, to arrive at a balanced design. Balancing is desirable since it both simplifies and strengthens the statistical analysis of the data. However, due to practical limitations we could not balance the students of the two different universities between the two treatments, e.g. half of the students of Ghent University being trained in TCM and ODCM and vice versa for the students at CTU in Prague. As such, one group may differ from the other – e.g. due to the students' specific profile or the teaching method of the respective professor. Hence, our type of experiment is a quasi-experiment (see section 7 on Validity). The design type of our quasi-experiment is a two-factors with two-treatments design, meaning that we compare the two treatments against each other with two dependent variables – the quality of the conceptual model (i.e. effectiveness) and the effort in constructing the model (i.e. efficiency). Each subject also takes part in only one treatment. Most commonly, the means of the dependent variables for each treatment are compared. We thus assign scores to the different measures of the dependent variables in order to compare our two different treatments objectively. This aspect is discussed in more detail in the following section.

3.4 Instrumentation

The instruments of an experiment provide means for performing the experiment and to monitor it, without affecting the control of the experiment. Below, we describe in detail the different phases a subject goes through when participating in our experiment, and the kinds of instruments we apply in each of these phases. We would like to note that all materials – the assignments per treatment, the case description, knowledge assessments, competency questions etc. – that have been applied in this experiment can be found in our online repository at Open Science Framework (OSF)[1].

Assessment of subjects' knowledge

In order to assess if the subjects clearly understood the respective modeling technique, we evaluate each subject's understanding with several written statements. Each of these statements describe a certain phenomenon or scenario, to which the subject has to choose the correct corresponding element of the modeling technique. The subjects can choose from four different multiple-choice answers. In total, six statements were given for each treatment (see OSF repository). Each of these statements was derived from examples from existing literature or exercises related to the techniques. If a student failed the assessment, their results were not included into the experiment.

[1] osf.io/w7mh2

Modeling Assignment

After the completion of the knowledge assessment, subjects could complete the modeling assignment. The assignment describes a company that desires to develop a software system for universities. As part of the development process, a conceptual model is required, that should be applicable to multiple universities. As means of a reference case, a description is given of a university. Subjects are given specific instructions that the concepts and entities of the university should be modeled, but that their model should also be accessible for representing the structure of other universities. The purpose of the task is thus of a rather businesslike nature, with the objective to deliver a 'complete' representation of the case, and which should at the same time be adaptive enough to apply to the structure of other universities. For example, the case describes that a professor can only work at one department of a faculty. However, this structure is specific for the university in the case. An adaptive model should also allow a professor to work at different faculties and/or work at different universities.

Additionally, we would like to emphasize that both groups did not have any tools, interfaces or any other modeling support at their disposal. The models were created with pen and paper, and when completed submitted to the supervisor of the experiment. As such, subjects relied solely on the respective modeling technique that they have been taught – i.e. the EER and the OntoUML technique – to derive and classify the information from the modeling scenario and to conceptualize this information into a model. This straightforward setup allows us to measure if the application of an ontology-driven technique increases the ontological thinking on the modeler's side.

Usage belief and perception

As a last phase in the experiment – after completing the modeling assignment – the participants are asked to fill in a set of 8 questions, which measure both the perceived usefulness and the perceived ease of use (see OSF repository). As a summary of this section, Figure 1 gives a more comprehensive overview of the different aspects of our experimental design.

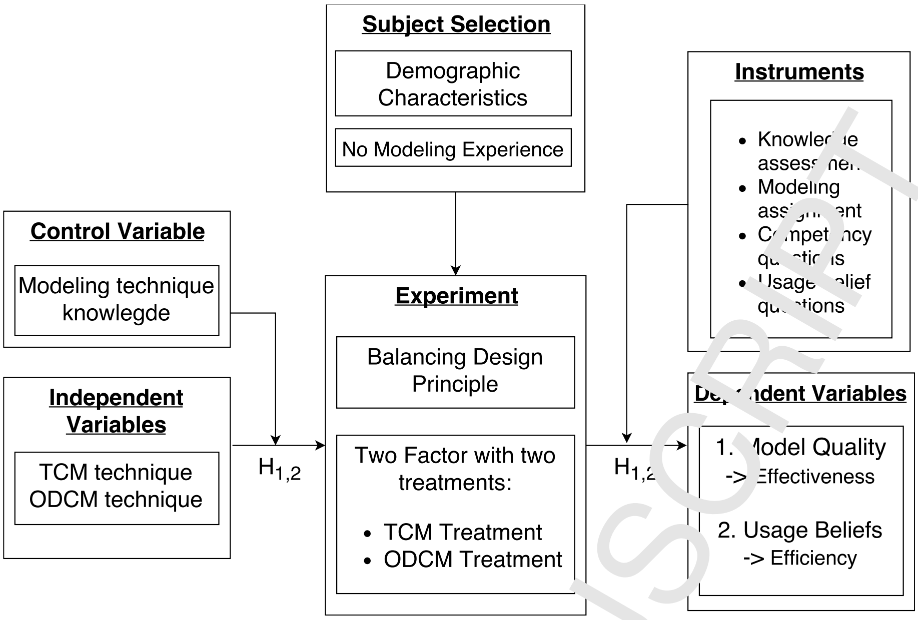


Figure 1: Overview of Experimental Design

4 Results

Below we first discuss the descriptive results related to the knowledge assessment, the effectiveness and the efficiency of each treatment. By regarding and discussing the descriptive results we can get a first indication of the differences that exists between the treatments. However, based upon the descriptive statistics we cannot conclude if the treatments are significantly different from one another. Therefore, we perform further statistical testing to test the hypotheses as formulated above and examine if significant differences can be deducted.

4.1 Descriptive statistics

Knowledge Assessment

The results of the knowledge assessment test – which was our control variable – indicate that all subjects gained a reasonable understanding of the respective technique's structure and concepts, with an average score of 97,6% for the TCM treatment and 94,3% for the ODCM treatment. None of the participating subjects gained a score lower than 50%, which would have excluded the results of the particular subject in the experiment.

Effectiveness of the treatments

As for the effectiveness of the treatments, we have demonstrated the average results of the competency questions in Table 1. More specifically, we have distinguished the total average scores for both the content interpretation questions as for the content sophistication questions. The very last column in this table then displays the total average scores for each separate treatment. As the table demonstrates, the scores for the ODCM concerning content interpretation are somewhat higher compared to the averages scores for the TCM treatment, although the difference is not substantial. Concerning the average scores of the content sophistication questions however we observe a much stronger difference. In total, the ODCM treatment obtained an average score of 46% while the TCM treatment achieves a considerable lower score of 24%. It would thus seem that adopting the ODCM approach enables subjects to better deal with the challenges and ambiguities that are confined in the modeling assignment compared to subjects adopting the TCM approach. Consequently, due to this rather substantial difference in results between the content sophistication scores, the total average scores of the ODCM treatment are also higher compared to the TCM scores. Hence, based upon the descriptive results of the effectiveness for each treatment, it would appear that the ODCM treatment was more effective in representing the domain as truthfully as possible compared to the TCM treatment – especially concerning content sophistication, which deals with the more challenging or ambiguous aspects of the case assignment.

Table 1: Average results corresponding to effectiveness

Average Results	Content Interpretation	Content Sophistication	Total Score
TCM	83.40%	24.00%	53.70%
ODCM	87.50%	46.00%	66.75%

Efficiency of the treatments

The average results for the measurements of the efficiency of each treatment can be found in Table 2. Here, the results are less straightforward in comparison with the results of the effectiveness. The first column displays the average time required for each subject to complete the modeling assignment. As we can observe, the average time is a little higher for the TCM treatment (38 minutes) compared to the ODCM treatment (36 minutes). However, this difference in time is rather small. We would like to note that this time measurement only involves the time required to complete the modeling assignment, meaning that it does not incorporate the time needed to complete the knowledge assignment or the completion of the perception questions. It is strictly the time required to read the modeling assignment and develop the conceptual model corresponding to this assignment. Additionally, the table demonstrates the average results per treatment. The average reduces the impact of outliners in your data – for example subjects that completed the modeling assignment rather fast or just the opposite, very slow. When we calculate the median of both treatments, we arrive at a greater difference, i.e. 37 minutes for the TCM treatment and 32 minutes for the ODCM treatment.

Next, we have calculated the total average score of the perceived usefulness and the perceived ease of use questions for every treatment. Since the questions correspond to a five-point Likert scale – 1 indicating strongly agree and 5 strongly disagree – this means that that the lower the score, the more the subject perceived the modeling technique as useful or easy to use (i.e. strongly agree with the statement concerning perceived usefulness or ease of use). As the table indicates, there is no clear difference between both treatments. The total score for the perceived usefulness is higher for the ODCM treatment, meaning that subjects perceived the technique as less useful compared to subjects of the TCM treatment. On the other hand, the perceived ease of use is slightly higher for the TCM treatment, indicating that this technique was perceived as less easy compared to the subject adopting the ODCM technique. These scores do not correspond to our second hypothesis, which expected subjects adopting the ODCM technique would perceive the technique as less easy to apply compared to the treatment adopting the TCM technique.

Table 2: Average results corresponding to efficiency, Note: lower values represent higher perceived usefulness and ease of use

Average Results	Time (hours: minutes)	Total Average Score Perceived Usefulness	Total Average Score Perceived Ease of Use
TCM	00:38	2.37	3.03
ODCM	00:36	2.60	2.91

4.2 Hypotheses Testing

Effectiveness of the treatments

In order to test our hypotheses, we are going to compare if the scores of the competency questions between the two treatments differ significantly. To determine which kind of test we have to apply, we first examine the distributions of our data – the total individual scores per subject, categorized per treatment. In order to identify if our data is normally distributed, we performed the Shapiro-Wilk test (p-value: 0.000), revealing that both the data of the ODCM and TCM treatment follow a non-normal distribution – indicating that we have to analyze our hypotheses with non-parametric tests. Additionally, we have performed the Kolmogorov-Smirnov test where we also obtained a significant p-value of 0.000. To compare the differences between our treatments, we have chosen the Mann-Whitney U test (McKnight & Najab, 2010). This test sets the following data limitations: (1) dependent variable should be measured at the ordinal or continuous level; (2) independent variable should consist of two categorical, independent groups; (3) independence of observations and (4) not-normally distributed data. Since our data is not normally distributed and consists out of two independent groups, with no relationship between the observations in each group or between the groups themselves, our data answers the assumptions of the Mann-Whitney U test (McKnight & Najab, 2010). In Table 3 and Table 4 we have displayed the results related to the Mean-Whitney U test. While Table 3 expresses the mean ranks and the sum of ranks for each assignment for both the TCM and ODCM treatment, Table 4 displays the outcome of the test and the associated p-values. We test our hypotheses on the 95% confidence interval. Additionally, since our hypotheses are directional – we test if one treatment scores higher than the other treatment – we regard the one-tailed significance level.

In line with our first hypothesis, we predict that the total score of the competency questions of the ODCM treatment will be higher compared to the scores of the TCM treatment. In order to gain more insight into the results, we have also tested for the total scores for both the content interpretation and the content sophistication questions. From Table 3 we can deduct that the mean rank and the sum of ranks are all higher for the ODCM treatment compared to the TCM treatment – i.e. for the interpretation, sophistication and total score of the competency questions. These ranks are in line with our observations of the descriptive results (supra). When we regard the results of the Mann-Whitney U test of the content interpretation questions, we observe a p-value equal to 0.161 – meaning that no significant difference can be acknowledged at the 95% confidence interval between the scores of the content interpretation questions of the ODCM and TCM treatment. However, when we retrieve the p-values of the content sophistication questions, we now obtain a significant result between the two treatments, with a corresponding p-value of 0.00. Finally, when we regard the total score for the competency questions, we again notice a significant difference at the 95% confidence interval between the ODCM and the TCM treatment, with a p-value of 0.00. In other words, we **accept H1**, and therefore confirming – at the 5% significance level – that novice modelers applying an ODCM technique arrive at higher quality models compared to novice modelers applying a TCM technique.

Table 3: Mann-Whitney U Ranks of Effectiveness Treatments

Ranks	**Group**	**Mean Rank**	**Sum of Ranks**
Content Interpretation Questions	TCM	47.67	2383.5
	ODCM	53.33	2666.5
Content Sophistication Questions	TCM	32.84	1642

Total Score Competency Questions	ODCM	68.16	3408
	TCM	38.82	1941
	ODCM	62.18	3109

Table 4: Mann-Whitney U Test of Effectiveness Treatments

Test Statistics	Content Interpretation Questions	Content Sophistication Questions	Total Score competency questions
Mann-Whitney U	1108.5	367	666
Wilcoxon W	2383.5	1642	1941
Z	-0.99	-6.127	-4.043
Asymp. Sig. (2-tailed)	0.322	0.000	0.000
Asymp. Sig. (1-tailed)	0.161	0.000	0.000

Efficiency of the treatments

Similar to the section above, we are going to compare if the perceived effort of developing a conceptual model with an ODCM technique is significantly higher compared to modelers applying a TCM technique. In other words, we examine if there exist significant differences in the time needed to develop the model, and the answers given by our modelers concerning the perceived usefulness and the ease of use of each respective technique. Similarly, we first investigate the distribution of our data with the Shapiro-Wilk test (p-value: 0.000), revealing again that our data – time required to complete the model and the efficiency questions – are non-normally distributed. Consequently, we can apply the non-parametric Mann-Whitney U test to compare our two treatments with each other.

In Table 5, we have displayed the ranks of the Mean-Whitney U test, while Table 6 displays the Mann-Whitney U results, for both the time and the two different types of efficiency questions. Since we are performing a one directional test – effort is higher for ODCM than for TCM – we have to regard the one-tailed asymptotic significance. First, when viewing the sum of ranks of the time measurement per treatment, we can see that the sum for the TCM treatment (2711.5) is substantially higher than for the ODCM treatment (2041.5). When we observe the results of the Mann-Whitney U Test, the p-value is equal to 0.0295, indicating that there is a significant difference between the TCM treatment and the ODCM treatment in time required to develop the model, on the 5% significance level. However, opposite to the hypothesis, it would seem that modelers of the ODCM treatment needed less time compared to modelers of the TCM treatment.

Next, when we regard the sum of ranks for both types of efficiency, we can observe that the difference between the sums are relatively smaller compared to the time measurement. Again, we would like to emphasize that the Mann-Whitney U test has been performed on scores related to the Likert scale – meaning that the lower the mean rank, the more the subject perceived the modeling technique as useful or easy to use (i.e. strongly agree with the statement concerning perceived usefulness or ease of use). When we examine the p-values of both types of efficiency questions, we observe a p-value of 0.0575 for the perceived usefulness questions and a p-value of 0.2425 for the questions corresponding to the perceived ease of use. Our results therefore do not confirm – at the 5% significance level – that the perceived usefulness and the perceived ease of use for the ODCM is lower compared to the TCM treatment. Since both these tests are not significant, and we obtain a significant difference (at the 95% confidence interval) with our time measurement in the opposite assumption of our hypotheses, we therefore **reject H2**, and cannot confirm – at the 5% significance level – that novice modelers applying an ODCM technique experience more effort in the process of developing a conceptual model compared to novice modelers applying a TCM technique.

Table 5: Mann-Whitney U Ranks of Efficiency Treatments

Ranks	**Group**	**Mean Rank**	**Sum of Ranks**
Time	TCM	54.23	2711.5
	ODCM	43.44	2041.5
Total Perceived Usefulness	TCM	45.53	2276.5
	ODCM	54.56	2673.5
Total Perceived Ease of Use	TCM	51.98	2599
	ODCM	47.98	2351

Table 6: Mann-Whitney U Test of Efficiency Treatments

Test Statistics	**Time**	**Total Perceived Usefulness**	**Total Perceived Ease of Use**
Mann-Whitney U	913.5	1001.5	1126
Wilcoxon W	2041.5	2276.5	2351
Z	-1.889	-1.574	-0.698
Asymp. Sig. (2-tailed)	0.059	0.115	0.485
Asymp. Sig. (1-tailed)	0.0295	0.0575	0.2425

5 Discussion

In the introduction of this article, we asked ourselves the question – the principal research question of this study – if there exist any meaningful differences in the resulting conceptual model and the effort spent to create such model between novice modelers trained in an ontology-driven conceptual modeling technique and novice modelers trained in a traditional conceptual modeling technique. The findings of our empirical study can now confirm that there do exist meaningful differences. More specifically, we found that novice modelers applying the ODCM technique arrived at higher quality models compared to novice modelers applying the TCM technique. On the other hand, we did not find any significant difference in effort between applying these two techniques. Below, we list various findings that are derived from the results of this study:

Finding 1. *Novice modelers applying an ODCM technique have no additional benefit over an TCM technique when modeling the foundational aspect of a domain.*

In our study, we composed a set of competency questions – content interpretation questions – that measured if the essential domain requirements of the scenario were met by the developed model of a subject. More specifically, these questions assessed if all the necessary concepts, relationships and multiplicities were adequately represented by the model conform to the description of the assignment. As indicated by our descriptive results in Table 1 the results of the content interpretation questions were somewhat higher for the ODCM technique (87,50%) compared to the TCM technique (83,40%). However, the following hypothesis testing in Table 4 designate that this difference is not significant (on the 5% significance level). Therefore, we can conclude that there exists no additional benefit in employing an ODCM technique over a TCM technique in the case where we have to model the basic requirements of a certain scenario or domain. These results were to be expected and are in line with the existing literature. As mentioned by (Gemino & Wand, 2005), the benefits of ODCM are presumed to be the highest when developing a more sophisticated representation of the domain being modeled, and should aid by achieving a higher level of domain understanding by its modelers and users. This assertion leads us to our second finding.

Finding 2. *Novice modelers applying an ODCM technique have a significant benefit over an TCM technique when modeling the advanced aspect of a domain.*

A second set of competency questions – the content sophistication questions – were also composed with the aim to measure how the models of a certain technique dealt with the more challenging and ambiguous facets of the case description. In order to score high on the content sophistication questions, subjects were required to respond beyond following the literal description of the case and improve their model so that it would provide a more precise representation of the domain. The descriptive results in

Table 1 already give a first indication that the ODCM technique amplifies content sophistication, since the average results of the competency questions were 46% compared to a total average score of 24% of the TCM treatment. The hypothesis testing displayed in Table 4 confirmed that the results of the content sophistication questions were higher for the ODCM technique compared to the results of the TCM technique. As such, the results of the empirical study demonstrate that it is advantageous to apply an ODCM technique over an TCM when having to model the more challenging and advanced facets of a certain domain or scenario. This clear difference in techniques can most probably be explained by the way modelers are adopted to an ontological way of thinking when learning and applying an ODCM technique. Idiosyncratically, modelers have to interpret and recognize a domain that they wish to model in the ontological concepts and rules that correspond to this technique. These ontological rules and concepts are governed by the axiom's, constraints and patterns of the underlying ontology. In other words, these patterns and constraints aid modelers in recognizing and coping with certain modeling pitfalls, to which modelers that adopt a non-ontological modeling technique are less well protected against. An example of such a pattern is displayed in Figure 2. In this figure, a typical pattern of the UFO ontology is displayed (Ruy et al., 2017). Without going into much details about the specific structure of the UFO ontology, a Kind can be seen as a 'rigid type', meaning that it supplies a principle of identity for its instances (Guizzardi, 2005). A Phase is always a specialization of a rigid type – in our case a Kind – where the specialization condition is always an intrinsic one. For instance, a child can be seen as a phase of a person, where the specific range of categorizing someone as a child can be specifically determined. Hence, modelers adopting the UFO ontology, and therefore also OntoUML, will model concepts such as childhood, adolescence and adulthood as phases of a person. Similar to the case description of our empirical study, modelers applying the OntoUML technique will have the tendency to model the different states of a course, i.e. 'Active' and 'Inactive', as phases of a course, and consequently as specializations of a course itself. Another way of modeling this description would be to simply assign active/inactive as a property of a course. However, when we then relate other concepts such as exam or exam date to a course, then we can have the conflicting situation where an exam and an exam date is scheduled for an inactive course. Therefore, the impact of the ontological pattern to recognize active and inactive states as further specializations of a course prompts modelers to more carefully consider the structure and order of their concepts and the intertwining relationships. We can find the impact of such patterns also clearly in the answers to the competency questions. For instance, when regarding the tenth content sophistication question – “Can exams and exam dates be associated only to active courses?” – the ODCM treatment scored a total of 74% on this question, compared to a 45% of the TCM treatment.

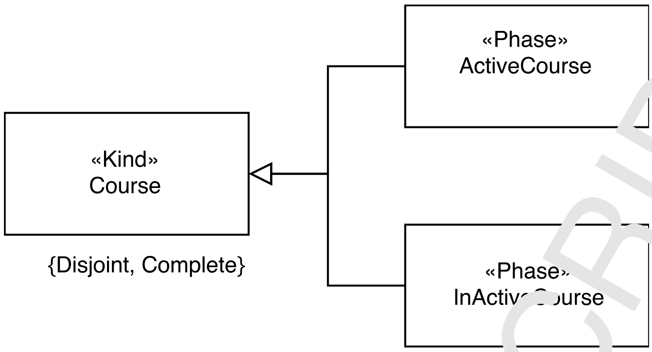


Figure 2: ODCM Pattern - Case Description Example

Finding 3. *Novice modelers applying an ODCM technique do not experience more effort in the process of developing a conceptual model compared to novice modelers applying a TCM technique – given a thorough understanding of the respective technique.*

Our last finding is contrary to our proposed hypothesis, by which we assume that applying an ODCM result in more effort due to the additional philosophical rules and concepts that have to be applied in the process, compared to a TCM where this is not the case. One possible explanation is that this effect can be attributed to the fact that these additional rules and structures lead to a well-designed modeling language, consequently increasing the conceptual clarity, which eventually does not result in a higher perceived effort to apply this modeling language. We would like to emphasize that this perception involves the application of the modeling technique by our subject – e.g. the EER and OntoUML modeling technique. The subjects received no modeling support in the form of a tool or interface. The purpose of this hypothesis was also not to measure how subjects perceive ontologies, but rather how ontology-driven modeling techniques are perceived in the task of creating a conceptual model.

This finding is also contrary to previous research efforts, such as the one of Soffer & Hadar (2007), where they found that difficulties were experienced in the adoption of the ontological concepts and rules underlying an ontology, especially with large sets of these rules. However, a key difference from this empirical study compared to previous research effort investigating ODCM, is that the subjects adopting the ODCM approach were trained and taught in this technique over a period of several months. Previous studies did also train their subjects in the ODCM technique, but this training occurred mostly in a rather short period of time. Presumably, when a modeler has been sufficiently familiarized with the ODCM technique, the different philosophical terms and rules no longer feel strenuous when developing a conceptual model. In fact, our results even expressed a significant difference in the time required to model the assignment, with a median time of 37 minutes for the TCM treatment and 32 minutes for the ODCM treatment. Of course, the significance in the difference of time spent by the subjects is intuitively less important than the perceived usefulness and perceived ease of use. When examining the descriptive results in Table 1 concerning the efficiency questions, we can observe that subjects from the TCM treatment slightly perceived their technique as more useful to apply compared to the subjects applying the ODCM technique. On the other hand, subjects from the ODCM treatment rated their technique as easier to use compared to subjects from the TCM treatment. These results feel quite contrary to the findings related to the effectiveness from each technique. The ODCM technique clearly assists a modeler in tackling the more challenging aspects of a domain, but at the same time it would seem that the modeler does not perceive the technique therefore as more useful.

One could argue that perhaps modelers are still unaware of the potential benefit an ODCM technique can have in achieving higher quality models. Perhaps even more surprising is that when we regard the specific questions (i.e. PU2 and PEU2) related to the effort of learning the technique, their results indicated that subjects of the ODCM treatment perceived their technique as easier to learn than the TCM treatment. On the other hand, subjects from the TCM treatment did find their technique more useful to learn compared to the subjects of the ODCM treatment. The results to these questions are quite surprising since one would expect that the ODCM technique would be perceived as more difficult to learn compared to the TCM treatment. Perhaps, when subjects are being taught the ODCM technique over a longer period of time, with regular practice and proper instructions, the difference in effort between learning a TCM technique and a ODCM technique fades. The results of the hypothesis testing in Table 4 also confirmed these observations, indicating that no significant difference – on the 5% significance level – can be found between the effort spent to construct a model between novice modelers trained in an ontology-driven conceptual modeling technique and novice modelers trained in a traditional conceptual modeling technique.

6 Conclusion

While many ontology-driven techniques have demonstrated to be beneficial compared to the traditional conceptual modeling practices, the added value of their application is not always straightforward and there is no clear distinction when it is actually desirable to adopt these techniques. Therefore, this paper conducted an empirical study that investigated the differences between adopting a TCM technique and an ODCM technique with the objective to understand and identify in which modeling situations an ODCM technique can prove beneficial compared to a TCM technique. More specifically, we trained two groups of novice modelers in each technique respectively and assigned these groups with an identical case description that had to be modeled with the corresponding technique. We then compared the two modeling approaches by investigating the quality of the resulting conceptual models, and the amount of effort a modeler had to spend in order to compose these models. The findings of our empirical study can now confirm that there do exist meaningful differences. However, since we are performing a quasi-experiment– meaning that key characteristics may differ between our treatments – we would like to emphasize that other effects such as the professor teaching the specific course or subject-specific characteristics can influence the outcome of our experimental results. Taking into account these limitations, our results revealed that novice modelers applying the ODCM technique arrived at higher quality models compared to novice modelers applying the TCM technique. More specifically, the results of the empirical study found that it is advantageous to apply an ODCM technique over an TCM when having to model the more challenging and advanced facets of a certain domain or scenario. This additional benefit can most probably be explained through the ontological rules and patterns that are incorporated into an ontology-driven modeling language. OntoUML is a pattern language, which means that the modeling primitives of the language are actually ontology design patterns and building a model in OntoUML is actually done by combining patterns (Ruy et al., 2017). In traditional languages, all the modeling choices are left to the modeler, who can combine low-granularity primitives (e.g., class, relation, attribute) in large variety of possible ways – including a number of undesirable ways. In a pattern language such as OntoUML, modeling construct come in well-defined clusters whose structure reflect the underlying ontological theories. These clusters provide a context for the interpretation and reuse of modeling constructs, decreasing the level of freedom in the use of these primitives and, hence, alleviating the modeler from the burden of making all modeling choices. For these reasons, our hypothesis is that the difference between ODCM and traditional conceptual modeling found in this paper would be even more profound in a setting in which students are educated in ODCM by fully exploring these pattern-based features. Testing this hypothesis, however, is something we intended to explore in future work. Finally, future research can also focus on comparing specific ODCM tools or technologies against existing conceptual modeling tools. These tools (e.g. Menthor[2]) hide the complexity of an ontology by incorporating the structuring rules and constraints in the tool itself. Additionally, also the machine-readability and machine-understandability of ODCM generated models could be compared against existing conceptual modeling technologies.

7 Validity

Internal Validity

In order to avoid any threats to validity, we have carefully designed and monitored the conduct of this experiment. Several experimental standards were also implemented to strengthen the validity of the experiment: (1) We applied the balancing design principle in order to balance between our treatments. However, due to practical limitations we could not balance the students of the two different universities between the two treatments, e.g. half of the students of Ghent University being trained in TCM and ODCM and vice versa for the students at CTU in Prague. As such, one group may differ from the other – e.g. due to the students' specific profile or the teaching method of the respective professor. Hence, our type of experiment is a quasi-experiment. The most important consequence of this quasi-experimental design is that our study may suffer from increased selection bias, meaning that other factors instead of our dependent variable may have influenced the outcome of our results. As a result, this also impacts the internal validity of our study; (2) subjects were selected from a 'controlled' environment, meaning that they all share a similar background and were novice modelers in the field of conceptual modeling; (3) neither of the subjects had any prior knowledge of either of the modeling techniques that were applied in the treatments; (4) we inserted a control variable in the experiment to assert that subjects had a similar understanding of the techniques before commencing the experiment; (5) our modeling task has been evaluated by a large amount of students before the actual experiment took place, in order to assure the modeling task was complex enough; and finally (6) the correction of the competency questions – although already rather objective by themselves – has been conducted by several authors of this article, where also the results between these authors were discussed and adjusted to ascertain that the models were rated as objectively as possible.

External Validity

[2] http://www.menthor.net/

Concerning external validity, we are well aware that by conducting our experiment on students, we limit the overall generalizability of our results. However, as stated by (Falessi et al., 2017), using students as participants remains a valid simplification of reality needed in laboratory contexts. It is an effective way to advance software engineering theories and technologies but, like any other aspect of study settings, should be carefully considered during the design, execution, interpretation, and reporting of an experiment. Consequently, we decided to select students as our test subjects since they have no prior knowledge of conceptual modeling and can thus be seen as novice modelers who can be trained in either TCM or ODCM. Furthermore, although we have balanced our number of subjects across our treatments, an even better approach would be to also balance subjects of the different universities over each treatment. In our current setup, only one type of technique was taught at each university. This was due to the practical organization of the classes given at the universities. We therefore acknowledge that dividing students over the different treatments per university would have increased the external validity of this study. We would like to remark however, that the nature of our results quite accurately follows the distinctions that exist between the techniques that have been applied in this study. For instance, the results of some competency questions can be clearly attributed to the existence of the ontological patterns that exist in the ODCM technique. Finally, we have presented the modeling case at the OntoCom workshop at the 36th International Conference on Conceptual Modeling in order to evaluate our case and the related competency questions by several experts in conceptual modeling and ontology. During this workshop, many of the competency questions –both for the content interpretation and especially the content sophistication –were derived from the feedback from these different experts. We deliberately also choose our assignment to deal with the university domain since students are well aware of this domain and so that there would not exist an additional advantage in modeling between the students.

Bibliography

Ashenhurst, R. (1996). Ontological aspects of information modeling. *Minds and Machines*, *6*(3), 287–394. http://doi.org/10.1007/BF00729802

Atkinson, C., Gutheil, M., & Kiko, K. (2006). On the Relationship of Ontologies and Models. *WoMM*, 96, 47–60.

Bera, P. (2012). Analyzing the Cognitive Difficulties for Developing and Using UML Class Diagrams for Domain Understanding. *Journal of Database Management*, *23*(3), 1–29. http://doi.org/10.4018/jdm.2012070101

Chen, P. P.-C. (1976). The Entity-Relationship Unified View of Data Model - Toward a Unified View of Data. *ACM Transactions on Database Systems*, *1*(1), 9–36. http://doi.org/http://doi.acm.org/10.1145/320434.320440

Corea, C., & Delfmann, P. (2017). Detecting Compliance with Business Rules in Ontology-Based

Process Modeling. *Proceedings Der 13. Internationale Tagung Wirtschaftsinformatik (WI)*, (April), 226–240. Retrieved from http://www.wi2017.ch/de/proceedings

Daga, A., de Cesare, S., Lycett, M., Partridge, C., Zhang, Z., Wang, S., … Partridge, C. (2005). An Ontological Approach for Recovering Legacy Business Content. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, *00*(C), 1–9. http://doi.org/10.1109/HICSS.2005.94

Davies, I., Green, P., Rosemann, M., Indulska, M., & Gallo, S. (2006). How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, *58*(3), 358–380. http://doi.org/10.1016/j.datak.2005.07.007

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of. *Information Technolog MIS Quarterly*, *13*(3), 319–340.

de Cesare, S., & Partridge, C. (2016). BORO as a Foundation to Enterprise Ontology. *Journal of Information Systems*, *30*(2), 83–112. http://doi.org/10.2308/isys-51428

Elmasri, R., & Navathe, S. B. (2015). *Fundamentals of database systems*. Pearson.

Evermann, J., & Wand, Y. (2005a). Ontology based object-oriented domain modelling: fundamental concepts. *Requirements Engineering*, *10*(2), 146–160. http://doi.org/10.1007/s00766-004-0208-2

Evermann, J., & Wand, Y. (2005b). Toward formalizing domain modeling semantics in language syntax. *IEEE Transactions on Software Engineering*, *31*(1), 21–37.

Evermann, J., & Wand, Y. (2006). Ontological Modeling Rules for Uml: an Empirical Assessment. *Journal of Computer Information Systems*, *47*, 14–29.

Falessi, D., Juristo, N., Wohlin, C., Turhan, B., Münch, J., Jedlitschka, A., & Oivo, M. (2017). Empirical software engineering experts on the use of students and professionals in experiments. *Empirical Software Engineering*, 1–38. http://doi.org/10.1007/s10664-017-9523-3

Fettke, P. (2009). How conceptual modeling is used. *Communications of the Association for Information …*, *25*. Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3494&context=cais

Gemino, A., & Wand, Y. (2005). Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. *Data & Knowledge Engineering*, *55*(3), 301–326. http://doi.org/10.1016/j.datak.2004.12.009

Gornik, D., & IBM. (2003). *Entity Relationship Modeling with UML*. *IBM*. Retrieved from http://www.ibm.com/developerworks/rational/library/content/03July/2500/2785/2785_uml.pdf

Grüninger, M., Atefi, K., & Fox, M. M. S. (2000). Ontologies to support process integration in enterprise engineering. *Computational & Mathematical Organization*, *6*(4), 381–394. http://doi.org/10.1023/A:1009610430261

Grüninger, M., & Fox, M. S. (1995). The role of competency questions in enterprise engineering. In *Benchmarking—Theory and practice* (pp. 22–31). Springer.

Guarino, N. (1998). Formal ontology and information systems. In *FOIS conference* (pp. 3–15).

Guizzardi, Das Graças, A. P., & Guizzardi, R. S. S. (2011). Design patterns and inductive modeling rules to support the construction of ontologically well-founded conceptual models in OntoUML. In *Advanced Information Systems Engineering Workshops* (Vol. 83 LNBIP, pp. 402–413). http://doi.org/10.1007/978-3-642-22056-2_44

Guizzardi, G. (2005). *Ontological Foundations for Structural Conceptual Models*. CTIT, Centre for Telematics and Information Technology. http://doi.org/10.1007/978-3-642-31095-9_45

Guizzardi, G. (2014). Ontological patterns, anti-patterns and pattern languages for next-generation conceptual modeling. In *International Conference on Conceptual Modeling* (pp. 13–27).

Guizzardi, G., & Wagner, G. (2005). Towards Ontological Foundations for Agent Modelling Concepts Using the Unified Fundational Ontology (UFO). *Agent-Oriented Information Systems II*, *3508*, 110–124. http://doi.org/10.1007/11426714_8

Guizzardi, G., Wagner, G., Almeida, J. P. A., & Guizzardi, R. S. S. (2015). Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story. *Applied Ontology*, *10*(3–4), 259–271. http://doi.org/10.3233/AO-150157

Henderson-Sellers, B. (2012). *On the mathematics of modelling, metamodelling, ontologies and modelling languages*. Springer Science & Business Media.

Honderich, T. (2006). *The Oxford Companion to Philosophy*. Oxford University Press.

ISO/IEC. (2001). *Software Engineering-Product Quality: Quality model* (Vol. 1). ISO/IEC.

Karagiannis, D., & Buchmann, R. A. (2018). A Proposal for Deploying Hybrid Knowledge Bases: the ADOxx-to-GraphDB Interoperability Case, *9*. Retrieved from https://scholarspace.manoa.hawaii.edu/handle/10125/50399

Karimi, J. (1988). Strategic Planning for Information Systems: Requirements and Information

Engineering Methods. *Journal of Management Information Systems*, *4*(4), 5–24.

Krogstie, J. (2012). *Model-based development and evolution of information systems: A Quality Approach*. Springer Science & Business Media.

McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. *Corsini Encyclopedia of Psychology*.

Moody, D. L. (2003). The Method Evaluation Model : A Theoretical Model for Validating Information Systems Design Methods. *Information Systems Journal*, 1327–1336.

Mylopoulos, J. (1992). Conceptual modeling and telos. In P. Loucopoulos & R. Zicari (Eds.), *Conceptual Modelling, Databases and CASE: An Integrated View of Information Systems Development*. Wiley.

Opdahl, A. L., Berio, G., Harzallah, M., & Matulevičius, R. (2012). An ontology for enterprise and information systems modelling. *Applied Ontology*, *7*(1), 49–92. http://doi.org/10.3233/AO-2011-0101

Pastor, O., & Molina, J. C. (2007). *Model-driven architecture in practice: a software production environment based on conceptual modeling*. Springer Science & Business Media.

Poels, G., Gailly, F., Maes, A., & Paemeleire, R. (2005). Object Class or Association Class? Testing the User Effect on Cardinality Interpretation. In *Proceedings of the 24th International Conference on Perspectives in Conceptual Modeling* (pp. 33–42). Berlin, Heidelberg: Springer-Verlag. http://doi.org/10.1007/11568346_5

Recker, J., Rosemann, M., Green, P. F., & Indulska, M. (2011a). Do ontological deficiencies in modeling grammars matter? *MIS Quarterly*, *35*(1), 57–79. Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=2942&context=misq

Recker, J., Rosemann, M., Green, P., & Indulska, M. (2011b). Do ontological deficiencies in modeling grammars matter? *MIS Quarterly*, *35*(1), 57–79.

Ruy, F. B., Guizzardi, G., Falbo, R. A., Reginato, C. C., & Santos, V. A. (2017). From reference ontologies to ontology patterns and back. *Data & Knowledge Engineering*, *109*, 41–69.

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

Shanks, G. G., Tansley, E., Nuredini, J., Tobin, D., & Weber, R. (2008). Representing Part-Whole Relations in Conceptual Modeling: An Empirical Evaluation. *MIS Quarterly*, *32*(3), 553–573.

Shao, J., Wang, H., & Chow, S.-C. (2008). *Sample size calculations in clinical research*. Chapman & Hall/CRC.

Soffer, P., & Hadar, I. (2007). Applying ontology-based rules to conceptual modeling: a reflection on modeling decision making. *European Journal of Information Systems*, *16*(5), 599–611. http://doi.org/10.1057/palgrave.ejis.3000683

Sowa, J. F. (1999). *Knowledge Representation: Logical, Philosophical, and Computational Foundations* (1st ed.). Course Technology.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, *11*(2), 93–136.

Verdonck, & Gailly, F. (2016). An Exploratory Analysis on the Comprehension of 3D and 4D Ontology-Driven Conceptual Models. In *Conceptual Modeling - ER 2016, Lecture Notes in Computer Science, vol. 9975* (Vol. 9975, pp. 163–172). http://doi.org/10.1007/978-3-642-33999-8

Wand, & Weber, R. (1993). On the ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal*, *3*(4), 217–237.

Wohlin, C., Runeson, P., Host, M., Ohlsson, M. C., Regnell, B., & Wesslen, A. (2012). *Experimentation in software engineering*. *Experimentation in Software Engineering* (Vol. 9783642290).

Highlights

i. This research article conducted an empirical study that investigated the differences between adopting a traditional conceptual modeling (TCM) technique and an ontology-driven conceptual modeling (ODCM) technique with the objective to understand and identify in which modeling situations an ODCM technique can prove beneficial compared to a TCM technique. We discuss previous empirical research efforts that have been previously made in this area and distill these efforts into two hypotheses.

ii. The purpose of our experiment is to measure the difference of the resulting conceptual model – i.e. the model quality – and the effort required to create such a model, when applying either a traditional modeling technique or an ontology-driven modeling technique. To measure these differences, we make a distinction between the effectiveness and efficiency of the two techniques. Effectiveness defines on how well a particular technique achieves its objectives, efficiency is viewed as the effort required to apply the technique.

iii. We trained two groups of novice modelers in each technique respectively and assigned these groups with an identical case description that had to be modeled with the corresponding technique. The hypotheses that were composed in the beginning of the article were then tested in a rigorously developed experiment, where a total of 100 students from two different Universities participated.

iv. The findings of our empirical study confirm that there do exist meaningful differences between both techniques. More specifically, our results revealed that novice modelers applying the ODCM technique arrived at higher quality models compared to novice modelers applying the TCM technique. More specifically, the results of the empirical study found that it is advantageous to apply an ODCM technique over an TCM when having to model the more challenging and advanced facets of a certain domain or scenario.