

# How quickly do we learn from conceptual models?

## ABSTRACT

In organizations, conceptual models are used for understanding domain concepts. Learning the domain from models is crucial for the analysis and design of information systems that are intended to support the domain. Past research has proposed theories to structure conceptual models in order to improve learning. It has, however, never been investigated how quickly domain knowledge is acquired when using theory-guided conceptual models. Based on theoretical arguments, we hypothesize that theory-guided conceptual models expedite the initial stages of learning. Using the REA ontology pattern as an example of theoretical guidance, we show in a laboratory experiment how an eye-tracking procedure can be used to investigate the effect of using theory-guided models on the speed of learning. Whereas our experiment shows positive effects on both outcome and speed of learning in the initial stages of learning, the real contribution of our paper is methodological, i.e. an eye-tracking procedure to observe the process of learning from conceptual models.

## Keywords

REA Ontology Pattern, Conceptual Model, Domain Knowledge, Learning, Eye-Tracking

## INTRODUCTION

A conceptual model describes a domain conceptualization that represents an abstraction of the domain for some purpose. In science, conceptual models are used to guide research by providing a visual representation of constructs or variables and their theorized relationships. In information systems engineering, the main purpose of conceptual models is sharing the domain knowledge that is required for systems analysis and design (Hoffer, Prescott, & McFadden, 2008). Conceptual models created using techniques and languages like Entity Relationship Modelling (ERM), Unified Modeling Language (UML), Event-Driven Process Chains (EPC), and Business Process Model and Notation (BPMN), are developed by analysts to foster shared domain understanding with the other stakeholders in systems development (Maass, Storey, & Kowatsch, 2011). These other stakeholders (e.g. future system users, system architects, project leaders) acquire domain knowledge by comprehending and interpreting the models developed by the analysts. In other words, they *learn* the domain from the model.

To improve learning from conceptual models, theories have been proposed that offer guidance on how to structure conceptual models. By structuring a conceptual model, we mean deciding how to use a modelling language or technique to represent domain concepts, relationships and constraints (i.e. choice of abstract syntax reflecting particular semantics) and/or how to present the model (i.e. choice of concrete syntax and, in case of a diagrammatic representation, choice of layout).

Predictions of improved learning using theory-guided conceptual models have been empirically demonstrated and explained using various theories of cognition (Bera, Burton-Jones, & Wand, 2011). Empirical studies have measured the acquisition of domain knowledge from using models by measuring model users' understanding of the model. Such measurement typically involves evaluating users' answers to semantic comprehension questions (i.e. assessing factual domain knowledge) (Khatri, Vessey, Ramesh, Clay, & Park, 2006). Whereas the focus of most of the research on theory-guided conceptual models was on the *outcome of learning*, in this paper we focus on the *speed of learning*. We define speed of learning as the rate at which domain knowledge is acquired.

A focus on speed of learning requires approaching learning from conceptual models as a process. A few studies have attempted to gain insights into the process of learning from models by means of perception-based measures requiring subjects to express their perception of the model's ease (or difficulty) of understanding or usefulness (Burton-Jones & Meso, 2008; Figl, Mendling, & Strembeck, 2013). Other studies have approximated speed of learning as the time taken to perform understanding-related tasks (Bodart et al., 2001; Mendling, Strembeck, & Recker, 2012). Both types of measurement are taken after having users interacting with models and as such they provide only indirect insights into the process of learning.

As the speed of learning from a conceptual model has not been adequately measured before, we propose a way of observing and measuring the process of learning using an eye-tracking procedure. We further propose *how to* evaluate the effect on speed of learning from models when theory-guided conceptual models are used. Hence, the research questions addressed in this paper are:

- (1) How to measure the speed of learning from conceptual models using eye-tracking?

(2) How to use eye-tracking to investigate whether conceptual models structured in a particular way expedite learning?

To demonstrate our proposed eye-tracking procedure, we conduct an empirical study using the REA ontology pattern (Dunn, Cherrington, & Hollander, 2005) as an example of theoretical guideline for structuring a conceptual model. Poels et al. (2011) showed that users of Entity Relationship Diagrams (ERDs) constructed using the REA ontology pattern acquire a more accurate understanding of policies governing business processes. In a follow-up study, Poels (2011) showed that such ERDs were also perceived as easier to interpret by users that are novices in modeling. However, these studies did not investigate the speed of learning from the models. Hence it is known whether learning from such models is also faster.

We emphasize that the intended contribution of our paper is the proposed procedure to observe and measure the process of learning. Although we develop theoretical arguments for effects on speed of learning from theory-guided conceptual models, we do not generalize the results of our empirical study on structuring models using the REA ontology pattern to other types of theory-guided models.

The paper is structured as follows: In the Background section, we introduce our terminology of ‘good representation’ and ‘poor representation’ models, depending on whether theoretical guidance for structuring conceptual models was used. We also present our theory-based hypotheses for the effects of ‘good representation’ models on the outcome and speed of learning. Further, we review how past research has measured learning from conceptual models. Then in the Eye Tracking section, we present our proposed procedure for observing and measuring the process of learning from conceptual models. This procedure is demonstrated in the Empirical Study section by comparing learning from models that follow or do not follow the REA ontology pattern. The second to last section discusses our findings and their implications for research and practice. The final section is the conclusion.

## **BACKGROUND**

### *Good and Poor Representation Models*

Several theories have been proposed that offer guidance for structuring conceptual models. Examples of such theories include the theory of ontological expressiveness (Wand & Weber,

1993), the Unified Foundational Ontology (Guizzardi, 2005), and the Physics of Notations (Moody, 2009). For instance, the theory of ontological expressiveness states that models are ontologically unclear if they contain instances of modelling language constructs that represent optional properties (Wand & Weber, 2017). Based on the theory, a prediction can be made that a model, for example an ERD, that shows optional attributes and relationships (i.e. using zero cardinalities) suffers from ontological clarity, hence will impede understanding of the domain represented through the model.

To explain in general why structuring of conceptual models affects learning, we refer to Cognitive Fit Theory (CFT) (Shaft & Vessey, 2006; Vessey, 1991). Figure 1 shows how we apply this theory to theory-based structuring of conceptual models.

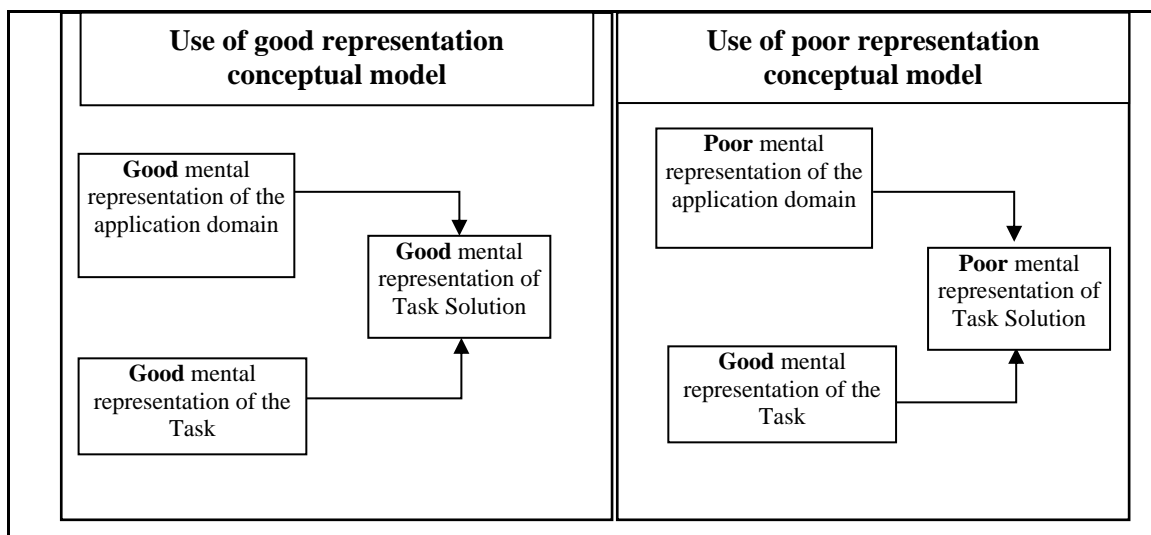


Figure 1: Effect of theory-based structuring of conceptual models in terms of cognitive fit

To perform a task requiring human information processing, humans create a mental representation of the task solution in their working memory (i.e. mental representation of Task Solution boxes in Figure 1) (Vessey, 1991). Problem-solving involves two distinct activities that need to be performed simultaneously to perform the task well - one related to the understanding of the task (i.e. what information is required for task resolution?) and the other related to the understanding of the application domain to which the task is performed (i.e. where and how to get the information required for task resolution?) (Vessey, 1991). CFT suggests that a good mental representation of the task solution is created if the mental representation of the task (i.e. mental representation of the Task boxes in Figure 1) and mental representation of the application

domain (i.e. mental representation of the Application Domain boxes in Figure 1) emphasize the same type of information (i.e. a situation of *cognitive fit*) (Vessey, 1991). Cognitive fit is highly relevant as the better the mental representation of the task solution, the better the task performance.

CFT has been used extensively in empirical studies on conceptual modeling (Bera, Burton-Jones, & Wand, 2014; Bera & Evermann, 2014; Bodart, Patel, Sim, & Weber, 2001; Burton-Jones & Meso, 2008; Evermann & Wand, 2006; Gemino & Wand, 2005; Shanks, Tansley, & Weber, 2004). Tasks considered in such studies relate to understanding a domain as modelled (i.e. *learning* a domain from a model). For instance, a task may involve using the conceptual model to verify the truth about some statement about the domain. In this context, understanding which information is required to evaluate the statement and retrieving that information from the model are two activities to be performed for task resolution. For a task that requires domain understanding, cognitive fit depends on the quality of the mental representation of the application domain (i.e. whether the information required for task resolution can be easily and accurately inferred from the conceptual model). If model users experience difficulties in comprehending and interpreting a conceptual model, then they face a situation known as *cognitive overload* (Mayer & Moreno, 2003). In this situation, individuals are faced with a task that demands more from their cognitive resources (such as working memory) than they can sustain.

The focus of much of the empirical research in conceptual modelling is to demonstrate that a particular structuring of conceptual models eases model comprehension and interpretation such that a good mental representation of the application domain is created (i.e. the **Good** mental representation of the Application Domain box in the left-hand side of Figure 1), resulting in cognitive fit and better task performance than when using models structured differently (i.e., the **Poor** mental representation of the Application Domain box in the right-hand side of Figure 1) (e.g. Bera, Burton-Jones, & Wand, 2014). In the remainder of this paper we refer to such theory-guided models as *good representation models* and contrast them with *poor representation models*, which are not structured following a particular theory-based guideline.

## *Hypothesis Development*

Past studies (e.g. Khatri et al. (2006) and Burton-Jones et al. (2009)) suggest that good representation models result in better performance in tasks related to domain understanding compared to bad representation models. Theory-guided structuring of conceptual models thus affects the *outcome of learning*, i.e. the extent to which domain knowledge was acquired. Past studies have, however, not approached learning from conceptual models as a *process*. What these studies have in common is that the effect of using a good representation model on domain understanding has been measured only after the exposure to the model, i.e. after the tasks have been performed. This means that a dynamic view of domain knowledge acquisition by means of conceptual models is absent in the past research.

Learning is, however, not instantaneous. In cognitive psychology research, it has been demonstrated that items are better learned when they are repeated sequentially with time intervals (Pavlik & Anderson, 2005) rather than presented all at once (Challis, 1993). This phenomenon is called the *spacing effect*. The *theory of encoding variability* (Young & Bellezza, 1982) is often used to explain why learning is enhanced when spacing is provided. As per this theory, the more spaced two items are, the more likely it is that they will be encoded in learners' mind differently and this encoding variability facilitates in providing more context and thus provides more retrieval cues (Richardson, 1973). Hence, referring to the CFT conceptual framework, when users get engaging tasks that require them to be exposed repeatedly to a model, then their mental representation of the application domain gets affected (i.e. the spacing effect) and so does the mental representation required for task solution, i.e. users understand the domain better and perform the tasks better. Repeated exposure to the model will thus eventually result in a situation of cognitive fit, regardless whether a good or poor representation model is used.

Based on CFT, we thus hypothesize that the use of a good representation model will lead to an *initial advantage* in learning from a conceptual model. In later phases of model use, it is plausible that any initial difference in learning from good and poor representation models will fade away. We thus formulate the first set of hypotheses related to the *outcome of learning*.<sup>1</sup>

---

<sup>1</sup> Note that these and the next set of hypotheses are not readily testable as they require further operationalization, which is done in the Empirical Study section.

H1: In the initial stages of model use, the *extent* to which domain knowledge is acquired from using a good representation model is higher than from a poor representation model.

H2: In the later stages of model use, the *extent* to which domain knowledge is acquired from using a good representation model is not different than from a poor representation model.

In cognitive psychology, learning speed is defined as the number of trials required for a learner to reach a state that enables recall or recognition of the learned material – the less trials, the faster the learning (Sandvik, 2013). Consistent with hypotheses 1 and 2, we hypothesize that related to the *speed of learning*, the use of a good representation model leads to an initial advantage only. We refer to the concept of learning curve to suggest how the speed of learning changes over time. The learning curve is a depiction of learning in a temporal setting where incremental learning improvements occur over a period of time (Linton & Walsh, 2013). Based on CFT, we expect that in the initial stages of model use, learning is faster using a good representation model as the theory-guided structuring will make the domain concepts that need to be understood for task resolution easier to locate and/or easier to interpret. In terms of the learning curve this means that we expect a steeper (or shorter) learning curve with the good representation model.<sup>2</sup> With repeated exposure of the models (i.e. the spacing effect), at a certain point of time, we expect that also users of poor representation models will be able to internalize the locations and interpretations of the modelled domain concepts that are needed for task resolution. Hence, after a while, in later stages of model use, users of poor representation models will catch up with the learning (i.e. the learning curves of good and poor representation models converge). Accordingly, we formulate a second set of hypotheses related to *speed of learning* from conceptual models.

H3: In the initial stages of model use, the *rate* at which domain knowledge is acquired from using a good representation model is higher than from a poor representation model.

---

<sup>2</sup> A steep or short learning curve does not indicate a difficult initial learning process as common language use of the phrase ‘steep learning curve’ usually seems to imply. On the contrary, a shallow or long learning curve indicates a more difficult initial learning process, which is what we expect when using a poor representation model.

H4: In the later stages of model use, the *rate* at which domain knowledge is acquired from using a good representation model is not different than from a poor representation model.

### *Measuring Learning from Conceptual Models*

In past conceptual modeling studies, subjects were required to answer questions about the domain that is represented in a model in order to assess their learning from the model. Khatri et al. (2006) differentiate semantic comprehension questions (e.g. must an X always be related to an Y?) from syntactic comprehension questions (e.g. what are the attributes of X?). While syntactic comprehension questions merely assess the user's comprehension of the model (depending on the user's knowledge of the used syntax), semantic comprehension questions require a deep engagement with the model and require an interpretation of the model's content in terms of domain concepts. Answering the latter type of questions requires knowledge of the semantics of the language or technique used to create the model.

A comprehension task does not allow for direct observation of the cognitive process of learning from models. To examine directly the underlying cognitive process of readers during learning from conceptual models, we propose the use of eye-tracking, which is a process-tracing technique. Another process-tracing technique that has been used before in empirical research with conceptual models is verbal protocol analysis (Burton-Jones & Meso, 2006), which requires subjects to verbalize their thought processes while performing tasks with the model. Another alternative to get direct observations of cognitive processing is the use of NeuroIS data collection techniques for measuring either central nervous system signals (e.g. PET, DMS, TDCS, EEG, fMRI and fNIRS for measuring brain activity) or instruments like electrocardiogram (heart beat), galvanometer (skin response), electromyography (facial muscular movement) and oculometry (pupil dilation) for measuring peripheral nervous system signals (Muller-Putz, Riedl, & Wriessnegger, 2015; Riedl & Léger, 2016; vom Brocke & Liang, 2014).

Compared to eye-tracking, verbal protocol analysis is a more intrusive process-tracing technique. Furthermore, coding verbal protocol data is challenging, prone to error, and requires subjective interpretation (Burton-Jones & Meso, 2006). To investigate the mental processes of humans, self-reporting is insufficient as people are often unable and/or unwilling to self-report (Dimoka, Pavlou, & Davis, 2011). NeuroIS methods on the other hand provide a viable



alternative to eye-tracking but are costly in terms of equipment and effort to apply (Riedl & Léger, 2016). Therefore, we developed an eye-tracking procedure to observe and measure the process of learning.

## EYE TRACKING

To address our research questions, we developed an eye-tracking procedure. The specific choice for using eye-tracking is that we are interested in what a user looks at when performing a task that requires understanding a domain as modelled and whether viewing patterns are different between the use of good and poor representation models. According to the eye-mind relationship (Petrusel, Mendling, & Reijers, 2017), we can only accurately perceive something when we pay *attention* to it, meaning fixate it with our eyes and focus our minds on it. Eye-tracking is capable of detecting when we focus on and pay attention to specific model areas and when we integrate information from different model areas. Its strength is in objectively, accurately, and unobtrusively observing the cognitive process in terms of attention to information elements and integration of information elements. With eye-tracking we can observe whether a user effectively pays attention to the parts of the model that are relevant for performing some task.

For the type of task to be performed we choose for semantic comprehension questions. This task requires a user to deeply engage with the model and make a conscious effort to understand the modelled domain concepts in order to locate and interpret the relevant information required for answering the questions. Our eye-tracking procedure for observing learning from conceptual models thus combines eye-tracking measurements (explained in the next sub-section) with performance data of a semantic comprehension task, which allows measuring both the outcome of learning and the speed of learning. Furthermore, by collecting eye-tracking measurements and performance data at regular intervals during the execution of the semantic comprehension task (e.g. after every  $x$  questions - see Empirical Study section for the operationalization used with the REA ontology pattern), we can observe the process of learning and compare the outcome and speed of learning between alternative models.

### *Eye-Tracking Measurements*

Eye-tracking offers a window into how individuals read and scan information that is displayed to them (Rayner, 1998). During decision-making tasks where users view information relevant to the decision, eye movements reveal the distribution of attention (Glaholt & Reingold, 2011). By

relating eye movements with decision-making data, one can obtain a picture of the decision-making process.

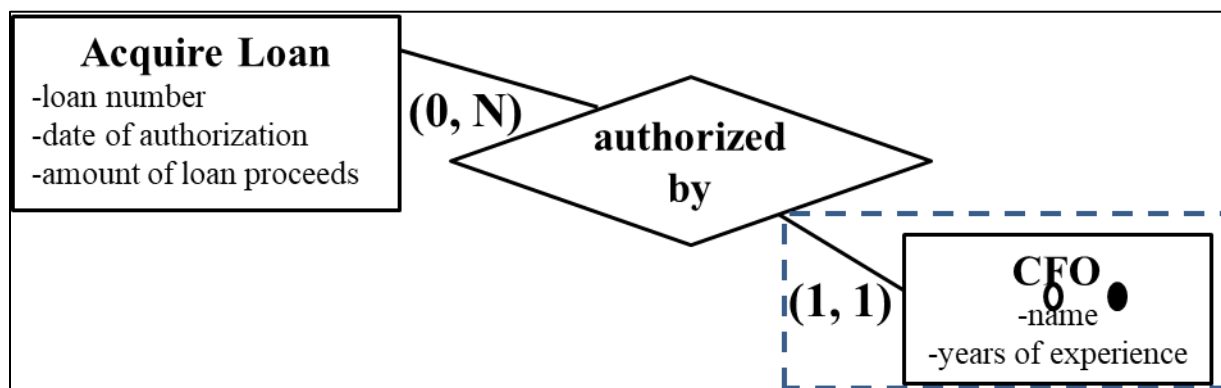
Two common eye movement observations are: eye fixations and eye saccades (Sharif & Maletic, 2010). Eye movements are made up of short bursts of stationary visual display termed *fixations* and are filled up with rapid and continuous movements termed *saccades* (Jacob, 1995). During fixations, eyes remain almost motionless, whereas saccades are movements from one fixation to another. A typical fixation lasts approximately 200-300 milliseconds and is generally understood to indicate where a viewer's attention is directed (Rayner, 1998).

When eyes fixate on a certain area, the brain starts to process the visual information received from the eyes (Rayner, 1998). Using eye-tracking it is possible to identify how much time a user has spent on a specific area of the model (referred to as the Area Of Interest (AOI)) and how quickly a user views a particular AOI. Especially the latter aspect is of interest for addressing our research questions. Lai et al. (2013) analyzed eye-tracking research in the education domain. They found that most educational research on eye tracking focuses on how users process information and that temporal metrics are the most commonly used eye tracking measures. Hence for our study, a relevant eye-tracking metric is the Time for First Fixation (TFF) (Poole & Ball, 2006). This metric indicates how quickly a user converges his/her eyes on a specific AOI in the model. The TFF is usually measured starting from the time a user is exposed to the model.

Figure 2 provides a fragment of an example ERD where TFF is measured with reference to an AOI termed CFO - a dotted rectangle created around the entity "CFO" including the cardinality that specifies to how many "CFO" instances an "Acquire Loan" instance is minimally and maximally related via an "authorized by" relationship. Consider that a user is exposed to Figure 2 at time  $t_0$ . Consider also that this user needs to evaluate the semantic comprehension question "*Is it possible to acquire a loan without the authorization of a CFO?*" (question 6 in our study, see Appendix B). To answer this question, the user needs to locate the "CFO" entity (possibly after first locating the "Acquire Loan" entity and next following the direction indicated by the "authorized by" relationship, although other reading sequences are possible) and interpret the cardinality specification within the CFO AOI. If the user's eyes fixate (i.e. eyes remain stationary for a minimum amount of time, e.g. 200 milliseconds) at any place inside the CFO

AOI, then the eye-tracker detects a fixation inside this AOI. The hollow dot and full dot in Figure 2 are examples of fixations within the CFO AOI. If the full dot in Figure 2 presents the first fixation within this AOI since the model was shown, and this fixation occurs at time  $t_1$ , then TFF for the CFO AOI is  $t_1 - t_0$ . For the example question, a second AOI is drawn around the “Acquire Loan” entity. If the first fixation within this AOI occurs at time  $t_2$  (which may be sooner or later than  $t_1$ ), then  $t_1 - t_0$  and  $t_2 - t_0$  are the TFFs to consider for answering the example question. Should the user’s eyes fixate again on one of these AOI’s (e.g., the hollow dot in the CFO AOI), then this time is not taken into account for TFF.

The shorter the TFFs for all AOIs relevant to answering a semantic comprehension question, the better a user was in locating and interpreting the relevant domain information for answering the question. If we observe at a certain moment during a semantic comprehension task (e.g. after  $x$  questions) lower TFFs for question  $x + 1$  when using model A than when using model B, then less trials to locate relevant model elements were needed with model A (i.e. less/shorter saccades and possibly less/shorter fixations outside the AOIs relevant to question  $x + 1$ ), signifying better recall or recognition of the learned material. Lower TFF values with model A than with model B thus indicate a higher rate of acquiring domain knowledge till that moment in the learning process. In other words, the TFF values indicate that at that moment in the learning process the learning curve when using model A is steeper than when using model B. Shorter TFFs thus indicate a higher speed of learning.



**Figure 2.** Example of fixations within the CFO AOI

## EMPIRICAL STUDY

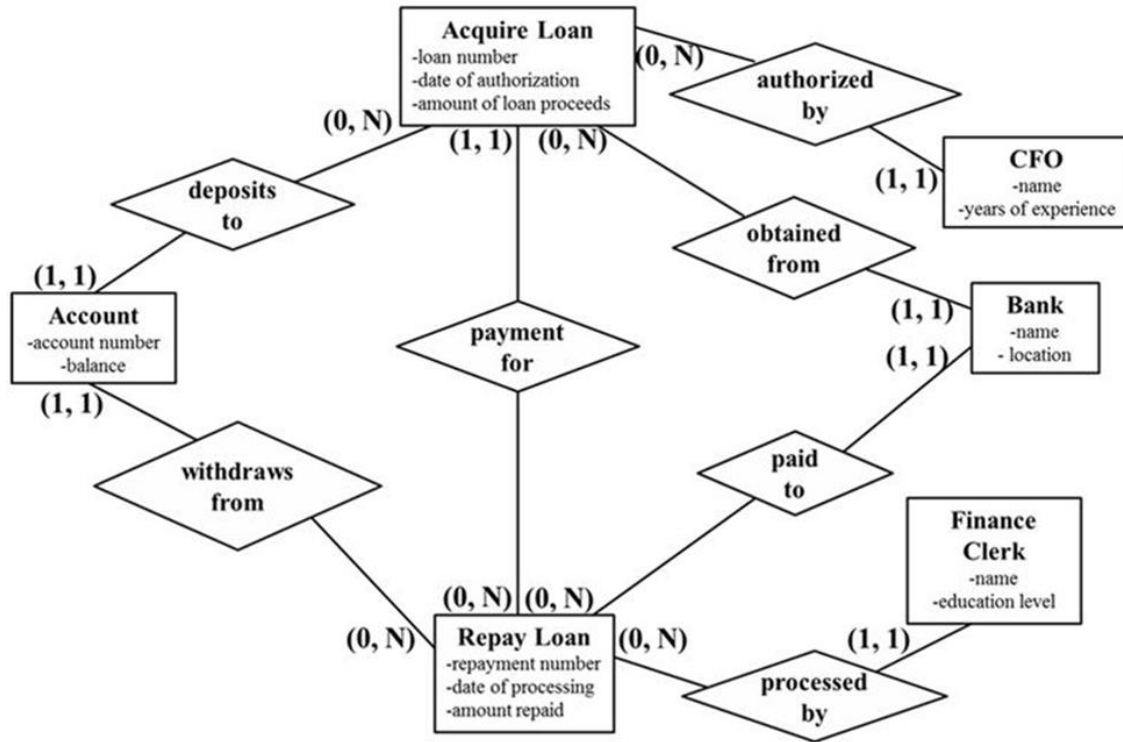
The empirical study that we present in this section aims at demonstrating how the eye-tracking procedure described in the previous section can be used to measure the speed of learning from conceptual models, and also how it can be investigated whether, as hypothesized (see the Hypothesis Development sub-section), good representation models expedite learning of domain concepts, at least in the initial stages of model use. For this demonstration, we use the REA ontology pattern (Dunn et al., 2005) as an example of a theory-derived guideline for structuring conceptual models. By applying the REA ontology pattern to ERDs we obtain good representation REA models. We also explain how we develop poor representation REA models as ERDs that do not follow the REA ontology pattern. We then use our eye-tracking procedure to compare the outcome and speed of learning from both types of model.

### *Structuring Conceptual Models: The REA Ontology Pattern*

The REA ontology is a *theory for conceptual modelling* originating in a semantic data model for accounting. This semantic model conceptualizes different real business domain situations related to economic exchange (e.g. selling, buying, renting, employing) using a common conceptual structure of three ontological categories: resources, events, and agents (McCarthy, 1982). The model was developed inductively by investigating massive amounts of transactional data, in which recurring sequences of resource-event-agent constellations were discovered (McCarthy, 1982). Later, the model was formalized as an ontology and was grounded in business process and value chain theories (Geerts & McCarthy, 2002).

The core concepts and relationships of the REA ontology have been specified as an *ontology pattern* (Ruy, Guizzardi, Falbo, Reginato, & Santos, 2017), called the REA ontology pattern, which can be represented using a template for ERDs that has a predefined layout for positioning domain concepts classified according to the REA ontological categories (Dunn et al., 2005). The predefined REA ontology pattern layout is one where resources, events, and agents are positioned in left, central, and right diagram regions respectively, such that the events that determine the essential ‘give’ and ‘take’ nature of the economic exchange form a central axis connecting the exchanged resources indirectly with the agents that exchange them. This layout possesses a number of properties that correspond well with principles from aesthetics-based computing research which, when adhered to, reduce the chance of cognitive overload (Petre,

2006; Purchase, 2014; Ware, Purchase, Colpoys, & McGill, 2002) (see Table 1). The REA ontology pattern (layout) thus serves as a guideline for structuring conceptual models as to obtain cognitive fit and creating, what we call in the paper, good representation REA models. Figure 3 shows the good representation REA model used in our study.



**Figure 3.** Good representation REA model used in the study

To instantiate the poor representation REA model needed for our study, we broke the conventional arrangement of the REA ontology pattern layout. To do so we deliberately violated the principles of aesthetics-based computing research. As there could be many ways of violating the principles, we focus on three perceptual segregation principles – proximity, symmetry, and orientation (Bennett, Ryall, Spalteholz, & Gooch, 2007). These principles are based on the idea that objects or figures can be memorable if they are differentiated from the background. Table 1 explains how we created the poor representation REA model used in the study (Figure 4). By not adhering to the REA ontology pattern, we expect that users of the poor representation REA model face a higher chance of suffering from cognitive overload. It is likely that their mental representation of the task solution is of lower quality than that of good representation REA model users, which according to the CFT conceptual framework will result in worse

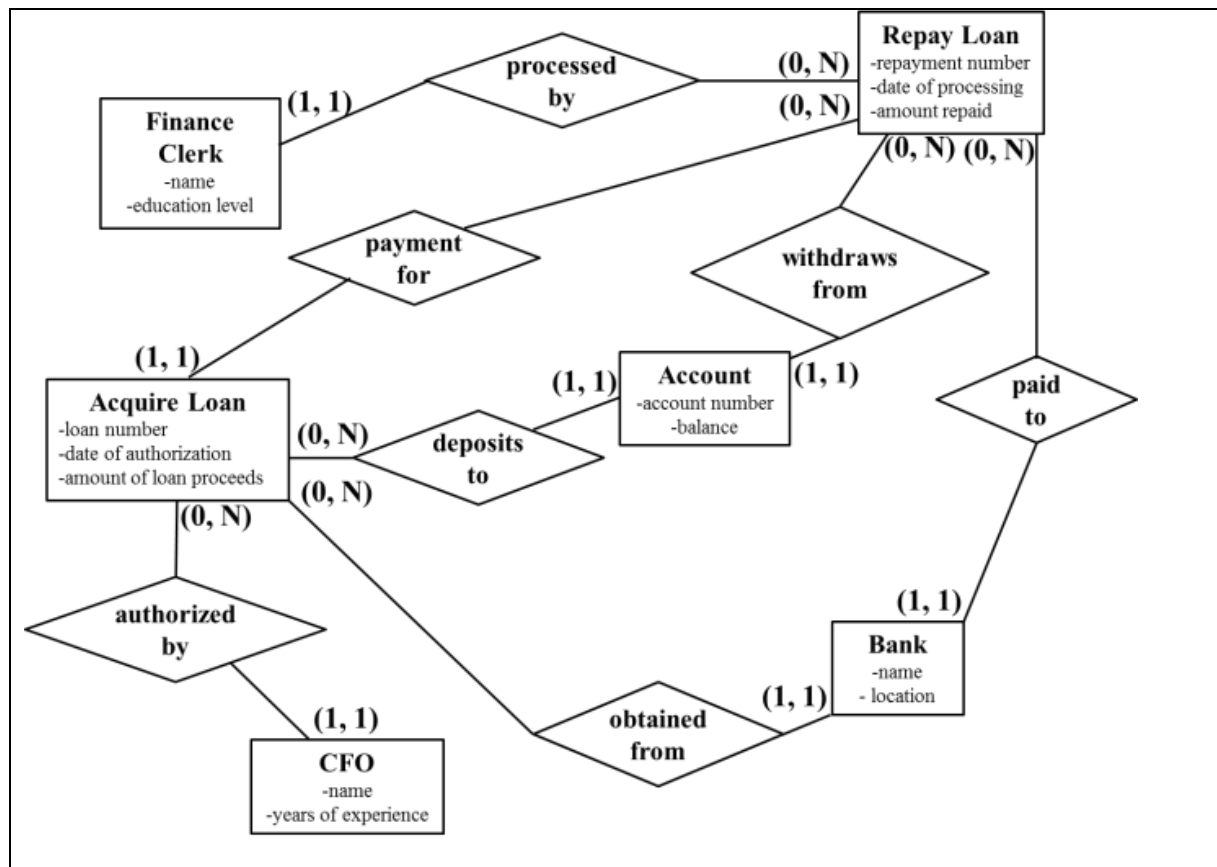
performance, i.e. a lower learning outcome and a more shallow (or long) learning curve (as we elaborated in the Hypothesis Development sub-section). However, as argued before, with repeated exposure to the models, the positive effect on learning from a good representation model will fade away and the learning curves of both groups will converge. Hence, we expect that the use of a good representation REA model has a positive effect on learning in the initial stages of model use only.

It is important to note that the information content of the models shown in Figures 3 and 4 is the same, implying that the models are informationally equivalent (Gemino & Wand, 2005). Only the arrangement of the model elements has changed. This means that the same set of semantic comprehension questions could be used for both models, allowing for a direct comparison of the outcome and speed of learning from the models.

**Table 1:** Principles of aesthetics-based computing

Concept	Details	Effect on cognitive overload
Symmetry	Graph-based layouts where elements are placed more symmetrically have been found to be easier to read. The symmetry that can be observed in the good representation REA model (i.e. a spatial organization where elements on the left are resources, elements at the middle are events, and elements on the right are agents) is broken in the poor representation REA model by positioning the elements in the diagram without any particular order.	The fixed position of REA ontological categories provides visual cues that help in identifying resources, events, agents, and their relationships. When these visual cues are missing then model users will have a higher chance of suffering from cognitive overload.
Use of locality (proximity)	When graphical elements that are logically related to each other are placed in close proximity, then it is easier to understand their connection. In the good representation REA model, all resource-event-agent constellations (i.e. which agent exchanges which resources with which other agent?) are modelled by entities located close to each other. Furthermore, these related entities are placed as rows in the diagram (as best as possible). In the poor representation REA model, we broke this row-oriented	By locating entities of resource-event-agent constellations in close proximity to each other, as ‘rows’ on the diagram, users are presented this information in perceptually and cognitively manageable chunks. These users will have less chance of cognitive overload when compared to users of the poor representation REA model, where the components of these constellations are spread out across the diagram.

	arrangement of the resource-event-agent constellations.	
Reading direction (orientation)	A good layout directs the reading of the elements in the model. In the good representation REA model, reading is directed from the middle where the events (as the ontological core of the transaction) are located to the right where the participating agents are found and to the left where the affected resources are found. But as the characteristic spatial organization of the REA ontology pattern no longer exists in the poor representation REA model, no particular reading direction is suggested.	The reading direction provides a visual cue for easily locating related resources, events, and agents. The absence of this cue affects cognitive overload negatively.



**Figure 4.** Poor representation REA model used in the study

### *Operational Hypotheses*

After choosing the theory-based guideline (i.e. REA ontology pattern), creating the models to be used in the study, and defining the measures for our dependent variables (i.e. number of correctly answered semantic comprehension questions for outcome of learning and TFF for speed of learning), we can now introduce the testable hypotheses:

H1: In the initial stages of model use, the *number of correctly answered semantic comprehension questions* when using the good representation REA model is higher than when using the poor representation REA model.

H2: In the later stages of model use, the *number of correctly answered semantic comprehension questions* when using the good representation REA model is not different from when using the poor representation REA model.

H3: In the initial stages of model use, the *TFF* when using the good representation REA model is lower than when using the poor representation REA model.

H4: In the later stages of model use, the *TFF* when using the good representation REA model is not different than when using the poor representation REA model.

What remains to be operationalized is the meaning of ‘initial’ and ‘later’ stages of model use. We show in the Results sub-section that this operationalization is robust with respect to the chosen interval of measuring our dependent variables.

### *Experimental design*

The study had a 1 x 2 between-subjects design where subjects were randomly assigned to one of two groups: good representation REA model and poor representation REA model. Forty-four graduate students (22 in each group) from a US university who enrolled in a graduate business analytics course over two terms participated in the study. For participation, these students received 2% of the course grade. We chose these students as study subjects as they learnt the basic concepts of conceptual modeling using ERDs. During the course, they were also introduced to the REA ontology, however, without showing them ERDs that followed the REA ontology pattern. Thus, when the subjects are provided with the ERDs used in the experiment (Figures 3 and 4), they are expected to have the required knowledge to be able to interpret the semantics of the model and thus perform a semantic comprehension task. We chose deliberately



for working with a relatively simple debt financing conceptual model constructed around a single business transaction (i.e. acquiring and repaying loans). A simple model combined with a large number of questions ensured that the same AOIs had to be revisited many times, allowing us to assess learning.

Each subject was placed in front of a computer fitted with an eye-tracker Tobii X2 60. At first, subjects' eyes were calibrated and validated (a standard procedure for eye-tracking) by asking them to follow a series of dots on the screen. After this procedure, the subjects accessed a web-based questionnaire to assess their familiarity with modeling and their knowledge of debt financing. Following this, they were shown a comprehension question (stage 1) and then the ERD depending on the group they belonged to (either good representation REA model or poor representation REA model) (stage 2). Once the subjects viewed the ERD carefully, they clicked on to the next screen to answer the question (stage 3). Subjects could click the back button to view the ERD again and then proceed to answer the question. A subject need not go back to stage 1 from stage 2 as the question is repeated in stage 3. Therefore, a subject can go back and forth only between stages 2 and 3. The model (stage 2) and the question (stage 3) were intentionally separated so that accurate eye measurement of subjects focusing on the model can be obtained. Appendix A shows the three stages for the good representation REA model with a specific semantic comprehension question example. The subjects were exposed to a set of 20 semantic comprehension questions, where for each question the 3-stage procedure explained before was repeated. The high number of questions asked for a relatively simple model ensured that the subjects get familiarized with the model after repeated exposure and thus the speed of learning from the model can be tracked across different phases of the experiment.

As investigating our research questions requires comparing learning at different phases of model use, we intentionally maintained the same level of difficulty of the semantic comprehension questions throughout the study. If compared to the initial stages, questions at the later stages of the study are easier or more difficult to answer, then an observed evolution in task performance and TFF measurements cannot be solely attributed to the type of the model used. Hence, we pilot-tested the questions with 5 PhD students with excellent knowledge of English from a European university who were familiar with the REA ontology. Based on their feedback on the clarity and difficulty level of the questions, some questions were modified for the study. Appendix B shows the final list of questions that we used (in the order shown).

## Results

As a preliminary note we state that in all statistical tests reported in this sub-section (Tables 2, 3, and 4) and Appendix E (Table E.1), the data is normally distributed for both groups and there is homogeneity of variance as assessed by Levene's Test for Equality of Variances. In all tables, 'Good' refers to the use of a good representation REA model and 'Poor' refers to the use of a poor representation REA model.

Before testing the hypotheses, it is important to verify the effectiveness of the randomization. Different levels of prior knowledge of the domain and familiarity with the modelling technique might affect subjects' performance. Therefore, subjects' prior domain knowledge and modelling familiarity was checked (see Appendix B for the measurement instrument used). Table 2 shows that the scores are not significantly different between the two groups.

**Table 2.** Analysis of the modeling familiarity and domain knowledge of the subjects

Groups	Modeling familiarity Mean (SD)	t- statistics	Domain knowledge Mean (SD)	t- statistics
Good	4.68 (0.90)	t = 0.16, p = 0.43	3.89 (1.14)	t = 0.62, p = 0.27
Poor	4.64 (0.95)		3.62 (1.72)	
Scores are on Likert scales ranging from 1 (lowest level of knowledge/familiarity) to 7 (highest level of knowledge/familiarity)				

Two types of analyses are done in this study, one with the task performance data (i.e. correctness scores) and the other with the eye movement data (i.e. TFF values) registered for answering each question. To distinguish between initial and later stages of model use, the twenty questions are divided into four phases where each phase consists of answering 5 questions. Thus, in the first phase subjects answer questions 1 to 5 and in the last phase they answer questions 16 to 20. To decrease the likelihood that this particular phasing decision has an impact on the results, we repeated all reported analyses for other groupings of questions. Appendix C shows that a similar pattern of results is obtained if questions were grouped per 2 or 3 (i.e. 7 phases) or per 4 (i.e. 5 phases).

To test the first set of hypotheses related to the *outcome of learning*, the analysis is done separately for each of the four phases. The sum of correct answers in each phase is compared between the two groups (Table 3). For example, the mean sum of correct answers on the first

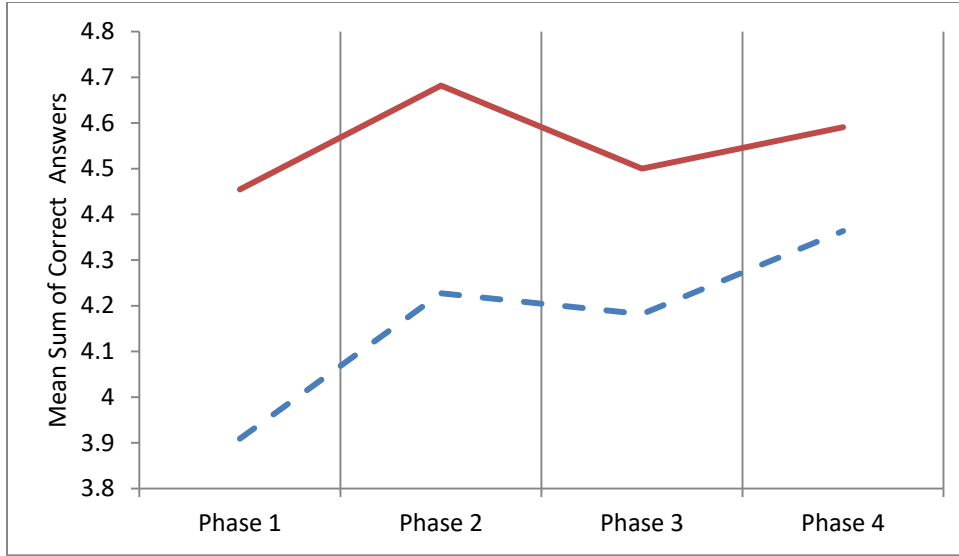
five questions is 4.45 for the good representation REA model group. The corresponding number for the poor representation REA model group is 3.91. Independent sample t-tests show that the mean differences between the two groups are significant in the first two phases but are not significant in the last two phases. Hence, based on these results we can accept hypotheses H1 and H2.

**Table 3.** Task Performance Analysis

<b>Groups</b>	<b>Phase 1 (Q1-Q5) Sum (SD)</b>	<b>t- statistics</b>	<b>Phase 2 (Q6-Q10) Sum (SD)</b>	<b>t- statistics</b>
Good	4.45 (0.59)	t = 2.24, p = 0.01*	4.68 (0.57)	t = 1.89, p = 0.03*
Poor	3.91 (0.97)		4.23 (0.97)	
<b>Groups</b>	<b>Phase 3 (Q11-Q15) Sum (SD)</b>	<b>t- statistics</b>	<b>Phase 4 (Q16-Q20) Sum (SD)</b>	<b>t- statistics</b>
Good	4.50 (0.74)	t = 1.37, p = 0.09	4.59 (0.50)	t = 1.13, p = 0.13
Poor	4.18 (0.79)		4.36 (0.79)	

\* p < 0.05

When all the questions are combined, the difference in mean sums of correct answers between the two groups is significant (i.e. the good representation REA model group performed better than the poor representation REA model group, t = 3.18, p = 0.001). A possible reason could be that, although the difference in mean sums of correct answers between the two groups is not statistically significant in the later phases, the values are still higher for the good representation REA model group. The following graph (Figure 5) demonstrates this where the mean sum of correct responses of both groups is plotted over the four phases. It can be noted that the gap between the two lines narrowed over the four phases, indicating that the advantage of working with the good representation REA model fades away after prolonged exposure to the model.



**Figure 5.** Comparison of task performance between good representation REA model users (solid line) and poor representation REA model users (dotted line). The maximum correct answer is 5 as each phase has 5 (Yes/No) questions.

To perform the TFF analysis (for testing hypotheses 3 and 4 related to the *speed of learning*), the models are divided into AOIs related to entities (along with the cardinalities as these cardinalities are also needed to answer the questions). These AOIs are of exact same size in both models. To answer a question, a subject must refer to specific AOIs in the model. Each question has two AOIs (except the 20<sup>th</sup> question which has 3). The list of AOIs is provided along with the questions in Appendix B. For example, to answer the first question “*Is it possible to acquire a loan from places other than a bank?*”, subjects need to refer to the AOIs “Acquire Loan” and “Bank”. The time it takes to look at these AOIs for the first time after the model is exposed (i.e. when a user moves from stage 1 to stage 2) is the TFF for these AOIs. If a subject navigates to the answer screen (i.e. moving from stage 2 to stage 3) without viewing the AOI and then comes back to view the AOI (i.e. moving back from stage 3 to stage 2) then the TFF is long as the entire duration since the model was exposed for the first time (i.e. moving from stage 1 to stage 2) till the subject has a fixation in the AOI is calculated as TFF. For each subject, the TFF values for the AOIs relevant to each question were registered.

The TFF analysis requires calculating the average TFF for each phase. To understand how the average TFF for the first phase is calculated, consider the AOIs “Acquire Loan” and “Bank” for the first question. For the good representation REA model group, the TFF averaged over all subjects is 1.92 seconds for the “Acquire Loan” AOI and 4.46 seconds for the “Bank”

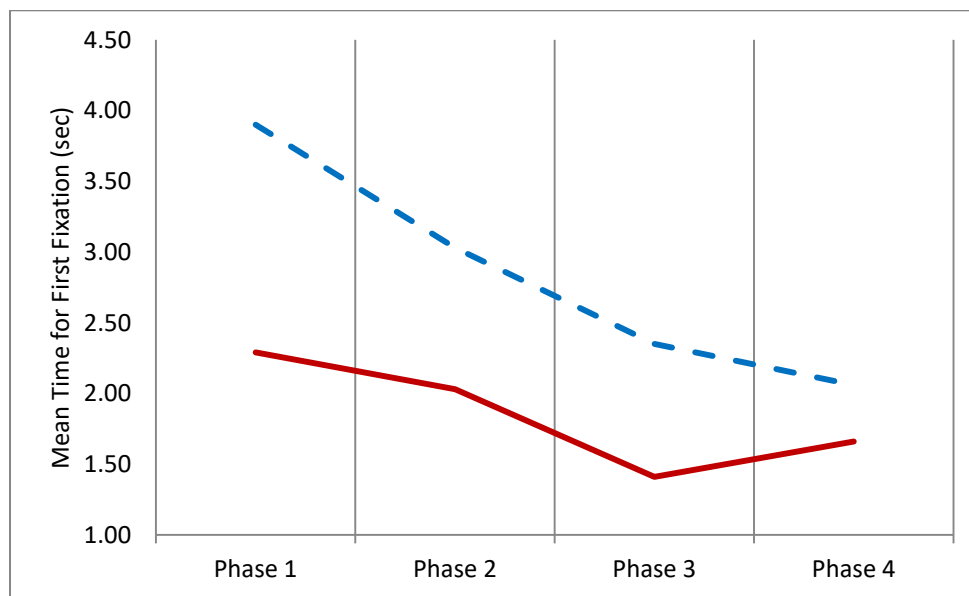
AOI. For the poor representation REA model group these values are 4.66 seconds and 5.61 seconds respectively. Now, the average TFF for the 10 AOIs (2 for each question and for the first five questions) is calculated. This comes to 2.29 seconds for the good representation REA model group and 3.90 seconds for the poor representation REA model group. Similar averages were calculated for the other three phases and are compared in Table 4. It is found that for each phase the average TFF was lower for those in the good representation REA model group than those in the poor representation REA model group. This allows accepting H3 but not H4.

**Table 4.** Time for First Fixation Analysis

Groups	Phase 1 (Q1-Q5) Mean in sec. (SD)	t- statistics	Phase 2 (Q6-Q10) Mean in sec. (SD)	t- statistics
Good	2.29 (0.45)	t = 5.41, p = 0.00*	2.03 (0.68)	t = 3.92, p = 0.00*
Poor	3.90 (0.66)		3.03 (0.89)	
Groups	Phase 3 (Q11-Q15) Mean in sec. (SD)	t- statistics	Phase 4 (Q16-Q20) Mean in sec. (SD)	t- statistics
Good	1.41 (0.50)	t = 4.26, p = 0.00*	1.66 (0.60)	t = 1.97, p = 0.03*
Poor	2.35 (0.90)		2.06 (0.74)	

\* p < 0.05

Between the two groups, the overall difference on average TFF for the AOIs for all the questions is also significant (t= 4.29, p = 0.00). Although the mean difference between the two groups narrows down over the phases, the differences remain significant in all four phases. Figure 6 demonstrates this where the average TFF of both groups is plotted over the four phases.



**Figure 6:** Comparison of time for first fixation between good representation REA model users (solid line) and poor representation REA model users (dotted line)

To get additional insights, we performed three additional analyses using the collected eye-tracking data. Appendix D shows some sample heat maps which are visual representations of *fixation intensity* using different colors (red - the largest number of fixations and green - the least number of fixations) (Jacob & Karn, 2003). These maps suggest that as subjects answered more questions, they were able to focus better on the AOIs that were necessary to answer the questions. Compared to the poor representation REA model group, the good representation REA model group initially focused better on the relevant AOIs. However, towards the end of the study, users of both groups were able to quickly converge their eyes on the AOIs needed to answer the questions.

A second analysis we performed was *visit count* analysis (Table 5). Visit counts track the number of times a viewer goes back to the same AOI.

**Table 5.** Visit Count Analysis

Groups	Phase 1 (Q1-Q5) Frequency (SD)	t- statistics	Phase 2 (Q6-Q10) Frequency (SD)	t- statistics
REA	3.07 (0.78)	t = 4.06, p = 0.00*	2.79 (0.65)	t = 3.42, p = 0.01*
Non-REA	4.08 (0.86)		3.48 (0.68)	
Groups	Phase 3 (Q11-Q15) Mean in sec. (SD)	t- statistics	Phasev4 (Q16-Q20) Mean in sec. (SD)	t- statistics
REA	2.52 (0.55)	t = 3.36, p = 0.01*	2.21 (0.56)	t = 2.05, p = 0.04*
Non-REA	3.21 (0.79)		2.60 (0.69)	

\* p < 0.05

The visit count analysis showed that poor representation REA model users go back to the relevant AOI a greater number of times than the good representation REA model users. However, in line with the trend depicted in Figure 6, visit counts decreased for both groups when moving through the phases of the task. Overall for all 20 questions, the mean visit count for the good representation REA model is 2.65 (0.39) and for the poor representation REA model it is 3.34 (0.40), (p = 0.00, t = 5.75). Finally, Appendix E provides similar insights based on a *mouse click count* analysis.

## DISCUSSION

We recall that the goal of the empirical study was to demonstrate how to use our proposed eye-tracking procedure to measure, apart from outcomes, the speed of learning from conceptual models (i.e. our first research question) and how it can be used to investigate whether a particular proposed way of structuring conceptual models expedite learning (i.e. our second research question). We first discuss the results of the empirical study, followed by a more in-depth discussion of how our proposed eye-tracking procedure addresses our research questions.

### *Theory-based structuring of conceptual models: The REA ontology pattern*

The analyses of comprehension task accuracy (Table 3) and TFF values (Table 4) show that users of the good representation REA model have an initial advantage over users of the poor representation REA model in learning the domain from the model, both in terms of outcome and speed of learning. Contrary to our expectations, this advantage is persistent for speed of learning (i.e. rejection of H4). Although the differences in average TFF values remain significant over the course of the semantic comprehension task, the visual inspection of the TFF data (Figure 6), the heat map analysis (Appendix D), the visit count analysis and the mouse click count analysis (Appendix E) indicate that the advantage becomes smaller in the later stages of model exposure, which suggests that users of the poor representation REA model catch up with the learning (i.e. converging learning curves).

The benefits of structuring conceptual models according to the REA ontology pattern have been investigated before (Fuller, Murthy, & Schafer, 2010; Gerard, 2005; Jones, Tsay, & Griggs, 2005). In particular for learning from such models, the studies of Poels et al. (2011) and Poels (2011) showed benefits related to more accurate understanding and easier interpretation of model contents. Our empirical study confirms the finding of better domain understanding if models are structured following the REA ontology pattern. However, the acceptance of H2 shows that this advantage is only temporary. Furthermore, using a model that is structured following the REA ontology pattern not only improves domain understanding, but the learning is also more efficient, i.e. the REA ontology pattern structure leads to faster focusing on those model elements that provide the domain understanding needed to perform tasks requiring such understanding.

We believe that compared to previous studies, our study provides unique insights into the cognitive process of learning from models. What is different from other studies, is that we measure outcome and speed of learning at regular intervals during the learning exercise. We acknowledge that the demonstration of our eye-tracking procedure on models structured using the REA ontology pattern does not corroborate or falsify the general, theory-agnostic hypotheses on outcome and speed of learning that we developed based on CFT. Nevertheless, our empirical study suggests the possibility that the beneficial effects on learning of structuring conceptual models using theory-derived guidelines are not persistent. In other words, using theoretical guidelines to structure conceptual models might be advantageous when it comes to foster shared domain understanding amongst stakeholders in systems development, however, the benefit of using theory-guided models is relative. Deep engagement with the models will eventually result in domain understanding, regardless how the models are structured.

The managerial implication is that using theoretical guidelines for structuring conceptual models that have proven to be beneficial for learning, is a good choice as learning will be faster and more effective, however, the benefits on the long term of such models should not be exaggerated. Nevertheless, for companies like IT consulting firms, which often have their consultants involved in information systems projects at client organizations, the benefits of structuring conceptual models can be substantial as time is a scarce resource and a steep learning curve in acquiring domain knowledge frees up time for other activities. Such companies should insist on using theory-based guidelines for structuring conceptual models. The implication for researchers is that they should also evaluate proposed theoretical guidelines for efficiency, and not just for effectiveness, as is usually the case now. Enabling faster learning is a quality dimension of conceptual models that is not to be neglected.

#### *Eye-tracking to observe the process of learning from models*

In a review of six studies that employed eye-tracking to investigate multimedia learning, Mayer (2010) finds that in all studies where a manipulation had an effect on learning outcomes it also had a corresponding effect on eye fixation time. Our empirical study results indirectly corroborate this finding as in the initial stages of learning, the good representation REA model users were both more efficient in locating relevant model elements and more effective in solving the task that depended on domain understanding.



What makes our eye-tracking procedure special is that we can investigate whether these effects are persistent. Although many studies have investigated how theory-guided conceptual models help in acquiring domain understanding, the impact of using a good representation model on domain understanding has been measured only after model exposure. For instance, studies like (Burton-Jones & Meso, 2008; Figl, Mendling, & Strembeck, 2013; Poels, 2011) that have used *perception-based* measures require subjects to express their perception of the model's ease (or difficulty) of understanding or usefulness after task performance and as such they measure the process of learning only indirectly. Studies like (Bodart et al., 2001; Mendling, Strembeck, & Recker, 2012) that have approximated speed of learning as the time taken to perform understanding-related tasks include in this time also the time taken to form a mental representation of the task (i.e. understanding the question). Compared to time measurements, our eye-tracking procedure allows to observe learning more directly as it only considers the time taken for forming the mental representation of the application domain (i.e. understanding the domain as modelled) since  $t_0$  in the calculation of TFF is the moment at which the model is exposed to the user.

Recently, the use of *NeuroIS measurements* has gained traction in IS research. Physiological measures such as heart rate (variability), brain activity, galvanic skin response and eye activity (i.e. pupil dilation and blink rate) have been proposed as indirect, real-time and objective measures of the cognitive processing required for understanding conceptual models (Weber, Sadiq, & Wang, 2018). In particular, such measurements can provide new insights into specific effects on cognitive load (i.e. effort of the working memory to move knowledge to long-term memory) like the worked example effect, the split-attention effect, the expertise reversal effect and other effects predicted by Cognitive Load Theory (Sweller, 1988). While these measures may reveal the effect of theory-guided structuring of conceptual models on cognitive (over)load, they cannot be used to observe where a person looks at when learning from a model. In contrast, our eye-tracking procedure based on scan path measures (e.g. TFF, visit counts) and visual gaze analysis (e.g. heat maps) allows observing model viewing patterns and thus provides valuable insights on how users read models, which can be used to test predictions based on specific theory-guided structuring of models.

For researchers, the implication is that using our eye-tracking procedure, which combines task performance and eye-tracking measurements, a more comprehensive evaluation of theory-

derived guidelines for structuring conceptual models can be done. In particular, our eye-tracking procedure allows gaining insights into how the structuring of conceptual models, based on theory, helps in directing attention to those model elements that matter for acquiring the domain knowledge needed for the tasks to be performed using models.

### *Limitations and future research*

We formulate the implications of our research with caution as we are aware of its limitations. First, the results of the empirical study with a model structuring following the REA ontology pattern cannot be generalized to theory-guided models in general. Structuring models can be done with respect to use of abstract syntax, concrete syntax, layout, or combinations of any of these. The REA ontology pattern only deals with diagrammatic layout. Hence, in our empirical study, two model versions were compared that were informationally equivalent. Comparing informationally equivalent models allows using the same comprehension task for both models. If theoretical guidance results in models not being informationally equivalent, then our proposed eye-tracking procedures is not applicable. Further research is needed to figure out how in such cases learning from domain models can be compared.

Second, we acknowledge that we did not formalize the precise relationship between a trend in TFF values over the course of a learning task (as in Figure 6) and the learning curve. Figure 6 shows that during the entire comprehension task, users of the good representation model are able to locate the model elements relevant for answering a semantic comprehension question quicker than the users of the poor representation model, indicating a steeper learning curve with the good representation model. But it is hard to tell from our data whether (and when) the learning curves of the alternative models are converging. Our graph of TFF values (Figure 6) seems to corroborate the education literature on learning curves, as per this literature, the rate of learning generally slows down after initial learning (Linton & Walsh, 2013). The exact nature of the relationship of TFF as a measure of speed of learning and the learning curve is a topic for future research.

## **CONCLUSION**

We presented a novel eye-tracking procedure to investigate how quickly we learn from conceptual models. This question has not been addressed in previous research, which did not approach learning from models as a process. Our procedure requires study subjects to perform a

relatively lengthy semantic comprehension task that requires deep engagement with a conceptual model. By repeatedly measuring the time for first fixation on model areas that convey the domain information needed to perform the task, we can observe and track learning of the domain and evaluate the speed of learning. By also measuring comprehension task accuracy at regular intervals during the course of the comprehension task, we can further control for the outcome of learning.

We also reported on an empirical study in which we used our eye-tracking procedure to test the hypothesis that conceptual models that are structured following the REA ontology pattern expedite learning. Our study demonstrates that, using our research design and eye-tracking procedure, researchers can investigate the hypothesized impact of theoretical guidelines for structuring models on users' domain understanding during model use. Our experience shows that eye-tracking is a viable procedure to get unique insights into the process of learning from conceptual models. Such insights allow researchers to more comprehensively evaluate proposed theories and guidelines for structuring conceptual models.

## REFERENCES

- Bennett, C., Ryall, J., Spalteholz, L., & Gooch, A. (2007). *The Aesthetics of Graph Visualization*.
- Bera, P., Burton-Jones, A., & Wand, Y. (2011). Guidelines for Designing Visual Ontologies to Support Knowledge Identification. *MIS Quarterly*, 35(4), 883-908.
- Bera, P., Burton-Jones, A., & Wand, Y. (2014). How semantics and pragmatics interact in understanding conceptual models. *Information Systems Research*, 25(2), 401-419.
- Bera, P., & Evermann, J. (2014). Guidelines for Using UML Association Classes and their Effect on Domain Understanding in Requirements Engineering. *Requirements Engineering Journal*, 19(1), 63-80.
- Bodart, F., Patel, A., Sim, A., & Weber, R. (2001). Should Optional Properties be Used in Conceptual Modeling? A Theory and Three Empirical Tests. *Information Systems Research*, 12(4), 384-405.
- Burton-Jones, A., & Meso, P. (2006). Conceptualizing Systems for Understanding: An Empirical Test of Decomposition Principles in Object-Oriented Analysis. *Information Systems Research*, 17(1), 38-60.
- Burton-Jones, A., & Meso, P. (2008). The Effects of Decomposition Quality and Multiple Forms of Information on Novices' Understanding of a Domain from a Conceptual Model. *Journal of the Association for Information Systems*, 9(12), 748-802.
- Burton-Jones, A., Wand, Y., & Weber, R. (2009). Guidelines for empirical evaluations of conceptual modeling grammars. *Journal of the Association for Information Systems*, 10(6), 495-532.
- Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 389-396.

- Dimoka, A., Pavlou, P., & Davis, F. D. (2011). NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research. *Information Systems Research*, 22(4), 687-702.
- Dunn, C., Cherrington, J. O., & Hollander, A. S. (2005). *Enterprise Information Systems: A Pattern Based Approach*: McGraw-Hill.
- Evermann, J., & Wand, Y. (2006). Ontological Modelling Rules for UML: An Empirical Assessment. *Journal of Computer Information Systems*, 47(1).
- Figl, K., Mendling, J., & Strembeck, M. (2013). The Influence of Notational Deficiencies on Process Model Comprehension. *Journal of the Association for Information Systems*, 14(6), 312-338.
- Fuller, R., Murthy, U., & Schafer, B. (2010). The effects of data model representation method on task performance. *Information and Management*, 47, 208-218.
- Geerts, G., & McCarthy, W. E. (2002). An Ontological Analysis of the Primitives of the Extended-REA Enterprise Information Architecture. *The International Journal of Accounting Information Systems*, 3(1), 1-16.
- Gemino, A., & Wand, Y. (2005). Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. *Data and Knowledge Engineering*, 55, 301-326.
- Gerard, G. J. (2005). The REA Pattern, Knowledge Structures and Conceptual Modeling: Comparison of Mandatory and optional properties. *Data and knowledge Engineering*(55), 301-326.
- Glaholt, M. G., & Reingold, E. M. (2011). Eye movement monitoring as a process tracing methodology in decision making research *Journal of Neuroscience, Psychology, and Economics*, 4, 125-146.
- Guizzardi, G. (2005). *Ontological Foundations for Structural Conceptual Models*. (PhD), University of Twente, The Netherlands.
- Hoffer, J. A., Prescott, M. B., & McFadden, F. R. (2008). Modern Database Management. In (8th ed.). Upper Saddle River, N.J.: Pearson Prentice Hall.
- Jacob, R. (1995). *Eye Tracking in Advanced Interface Design, in Virtual Environments and Advanced Interface Design*. New York: Oxford University Press,.
- Jacob, R., & Karn, K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye: Cognitive and applied aspects of eye movement research*. In R. Hyona (Ed.). Oxford, England: Elsevier.
- Jones, R. A., Tsay, J. E., & Griggs, K. (2005). An Empirical Investigation of the Task Specific Relative Strengths of Selected Accounting and Information Systems Diagramming Techniques. *Journal of Computer Information Systems*(46), 99-114.
- Khatri, V., Vessey, I., Ramesh, V., Clay, P., & Park, S. (2006). Understanding Conceptual Schemas: Exploring the Role of Application and IS Domain Knowledge. *Information Systems Research*, 17(1), 81-99.
- Lai, M., Tsai, M., Yang, F., Hsu, C., Liu, T., Lee, S., . . . Tsai, C. (2013). A Review of Using Eye-Tracking Technology in Exploring Learning from 2000 to 2012. *Educational Research Review*, 10, 90-115.
- Linton, J. D., & Walsh, S. T. (2013). Extracting Value from Learning Curves: Integrating Theory and Practice. *Creativity and Innovation Management*, 22(1).
- Maass, W., Storey, V., & Kowatsch, T. (2011). *Effects of External Conceptual Models and Verbal Explanations on Shared Understanding in Small Groups*: Springer.
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20, 167-171.

- Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43-52.
- McCarthy, W. E. (1982). The REA Accounting Model: A Generalized Framework for Accounting Systems in a Shared Data Environment. *The Accounting Review*, 57(3), 554-578.
- Mendling, J., Strembeck, M., & Recker, J. (2012). Factors of Process Model Comprehension-Findings from a Series of Experiments. *Decision Support Systems*, 53, 195-206.
- Moody, D. (2009). The Physics of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. *IEEE Transactions on Software Engineering* 35(6), 756 - 779.
- Muller-Putz, G., Riedl, R., & Wriessnegger, S. (2015). Electroencephalography (EEG) as a Research Tool in the Information Systems Discipline: Foundations, Measurement, and Applications. *Communications of the AIS*, 37(46).
- Pavlik, P., & Anderson, J. (2005). Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect. *Cognitive Science*, 29, 559-586.
- Petre, M. (2006). Cognitive dimensions 'beyond the notation'. *Journal of Visual Languages & Computing*, 17(4), 292-301.
- Petrusel, R., Mendling, J., & Reijers, H. A. (2017). How visual cognition influences process model comprehension. *Decision Support Systems*, 96, 1-16.
- Poels, G. (2011). Understanding Business Domain Models: The Effect of Recognizing Resource-Event-Agent Conceptual Modeling Structures. *Journal of Database Management*, 22(1), 69.
- Poels, G., Maes, A., Gailly, F., & Paemeleire, R. (2011). The Pragmatic Quality of Resources-Events-Agents diagrams: an experimental evaluation. *Information Systems Journal*, 21(1), 63-89.
- Poole, A., & Ball, L. J. (2006). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. In C. Ghaoui (Ed.), *Encyclopedia of Human Computer Interaction* (pp. 211-219). Hershey, PA: Idea Group.
- Purchase, H. C. (2014). Twelve years of diagrams research. *Journal of Visual Language Computing*, 25(2), 57-75.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Richardson, J. (1973). Effect of speed of learning and degree of learning on cue selection. *Journal of Experimental Psychology*, 98(2), 396-403.
- Riedl, M., & Léger, P.-M. (2016). *Fundamentals of NeuroIS. Information Systems and the Brain*: Springer.
- 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.
- Sandvik, Ø. (2013). *The Interaction of Learning Speed and Memory Interference: When Fast is Bad*. (Masters of Philosophy), University of Oslo, Oslo.
- Shaft, T., & Vessey, I. (2006). The Role of Cognitive Fit in the Relationship Between Software Comprehension and Modification. *MIS Quarterly*, 30(1), 29-55.
- Shanks, G., Tansley, E., & Weber, R. (2004). Representing Composites in Conceptual Modeling. *Communications of the ACM*, 47(7), 77-80.
- Sharif, B., & Maletic, J. (2010). *An eye tracking study on the effects of layout in understanding the role of design patterns*. Paper presented at the IEEE International Conference on Software Maintenance.
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12(2), 257-285.

- Vessey, I. (1991). Cognitive Fit: A Theory-Based Analysis of the Graphs versus Tables Literature. *Decision Sciences*, 22(2), 219-240.
- vom Brocke, J., & Liang, T.-P. (2014). Guidelines for neuroscience studies in information systems research. *Journal of Management Information Systems*, 30(4), 211-234.
- Wand, Y., & Weber, R. (1993). On the Ontological Expressiveness of Information Systems Analysis and Design Grammars. *Journal of Information Systems*, 3, 217-237.
- Wand, Y., & Weber, R. (2017). Thirty Years Later: Some Reflections on Ontological Analysis in Conceptual Modeling. *Journal of Database Management*, 28(1).
- Ware, C., Purchase, H. C., Colpoys, L., & McGill, M. (2002). Cognitive measurements of graph aesthetics. *Information Visualization*, 1(2), 103-110.
- Weber, B., Sadiq, S., & Wang, W. (2018). *How to use Neuro-physiological Measurements in BPM Research and Practice*. Paper presented at the 16th International Conference on Business Process Management, Sidney, Australia.
- Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(6), 545-559.

## Appendix A: Procedure for answering semantic comprehension questions in the experiment

Stage 1: Subject exposed to the question	
<p style="text-align: center; color: #808080;">Is it possible to acquire a loan from places other than a bank?</p> <div style="text-align: center; margin-top: 10px;"> <span style="background-color: #4a7c9c; color: white; padding: 5px 15px; border: 1px solid #4a7c9c;">Continue</span> </div>	
Stage 2: Subject exposed to the model	
<pre> graph TD     AL["Acquire Loan -loan number -date of authorization -amount of loan proceeds"]     A["Account -account number -balance"]     RL["Repay Loan -repayment number -date of processing -amount repaid"]     CFO["CFO -name -years of experience"]     B["Bank -name -location"]     FC["Finance Clerk -name -education level"]      AL -- "(0, N)" --&gt; D1{authorized by}     D1 -- "(1, 1)" --&gt; CFO     AL -- "(0, N)" --&gt; D2{obtained from}     D2 -- "(1, 1)" --&gt; B     AL -- "(1, 1)" --&gt; D3{payment for}     D3 -- "(0, N)" --&gt; RL     AL -- "(0, N)" --&gt; D4{paid to}     D4 -- "(1, 1)" --&gt; FC     AL -- "(0, N)" --&gt; D5{processed by}     D5 -- "(1, 1)" --&gt; FC     A -- "(1, 1)" --&gt; D6{deposits to}     D6 -- "(0, N)" --&gt; AL     A -- "(1, 1)" --&gt; D7{withdraws from}     D7 -- "(0, N)" --&gt; RL     </pre> <div style="text-align: center; margin-top: 10px;"> <span style="background-color: #4a7c9c; color: white; padding: 5px 15px; border: 1px solid #4a7c9c;">Continue</span> </div>	
Stage 3: Subject provides the answer	
<p style="text-align: center; color: #808080;">Is it possible to acquire a loan from places other than a bank? Yes <input type="radio"/> No <input type="radio"/></p> <div style="display: flex; justify-content: center; gap: 20px; margin-top: 10px;"> <span style="background-color: #4a7c9c; color: white; padding: 5px 15px; border: 1px solid #4a7c9c;">Back to the Model</span> <span style="background-color: #4a7c9c; color: white; padding: 5px 15px; border: 1px solid #4a7c9c;">Continue</span> </div>	

## Appendix B: Semantic comprehension, domain knowledge and modeling familiarity

No	Semantic Comprehension Question	AOIs (number)
1	Is it possible to acquire a loan from places other than a bank?	Acquire loan (1), Bank (2)
2	Does a finance clerk process a loan repayment to a bank?	Repay loan (3), Finance Clerk (4)
3	Is loan acquisition processed by a finance clerk?	Acquire loan (5), Finance Clerk (6)

4	Does a CFO authorize a loan repayment?	CFO (7), Repay loan (8)
5	Is a loan repayment paid to a bank?	Repay Loan (9), Bank (10)
6	Is it possible to acquire a loan without the authorization of a CFO?	Acquire loan (11), CFO (12)
7	Are the loan proceeds deposited to a specific account?	Acquire loan (13), Account (14)
8	Can a bank be associated with no loan repayment?	Repay loan (15), Bank (16)
9	Is it possible to deposit a specific loan proceed to more than one account?	Acquire loan (17), Account (18)
10	Can a bank be associated with no loan acquisition?	Bank (19), Acquire Loan (20)
11	Is it possible to trace the total amount of loan repayments for a loan with loan number?	Acquire loan (21), Repay loan (22)
12	Is it possible to repay a loan over a period of time?	Acquire loan (23), Repay loan (24)
13	Is it possible to repay a loan other than by using funds from an account?	Repay loan (25), Account (26)
14	Can an account be associated with no loan repayment?	Repay loan (27), Account (28)
15	Is it possible to obtain more than one loan from the same bank?	Acquire loan (29), Bank (30)
16	For a specific loan repayment, is it possible to withdraw funds from more than one account?	Repay loan (31), Account (32)
17	Can an account be associated with no loan acquisition?	Acquire loan (33), Account (34)
18	Can a CFO authorize more than one loan acquisition?	Acquire loan (35), CFO (36)
19	Can a specific repayment number have more than one loan number?	Acquire loan (37), Repay loan (38)
20	Is it possible to repay a loan using funds from an account that was also used to deposit the loan proceeds?	Acquire loan (39), Repay loan (40), Account (41)

### Questions for assessing modeling familiarity and domain knowledge

To what extent do you know data modeling concepts (such as entities and relationships)?

To what extent do you have experience in using data modeling concepts (such as entities and relationships)?

To what extent are you familiar with the processes of obtaining and repayment of bank loans?

To what extent do you have experience with the processes of obtaining and repayment of bank loans?



### Appendix C: Data analysis for other phasings of the comprehension task

<b>Groups</b>	<b>Performance difference between two groups</b>	<b>TFF difference between two groups</b>
Q1-Q3	Yes	Yes
Q4-Q6	Yes	Yes
Q7-Q9	Yes	Yes
Q10-Q12	No	Yes
Q13-Q15	No	Yes
Q16-Q18	No	Yes
Q19-Q20	No	No
Q1-Q4	Yes	Yes
Q5-Q8	Yes	Yes
Q9-Q12	No	Yes
Q13-Q16	No	Yes
Q17-Q20	No	Yes
Q1-Q5	Yes	Yes
Q6-Q10	Yes	Yes
Q11-Q15	No	Yes
Q16-Q20	No	Yes

Note: As per the above analysis, it is difficult to pinpoint the exact question which the breaking point for the performance difference between the two groups is. Rather a range (Question 9-Question 10) is the probable point from where there is no difference in the performance between the two groups.

## Appendix D: Heat map analysis (for sample heat maps)

Question	Good representation REA model group	Poor representation REA model group
5		
12		
19		

## Appendix E: Mouse click count analysis

To formulate the hypotheses, we made an assumption that the internal representation of users gets modified by being exposed to an engaging task with the model (Shaft & Vessey, 2006). Learning takes place as this representation gets modified. If users have already created a good mental representation of the model, then they need not refer to the model frequently in answering the questions. This would be particularly true for those who used the good representation REA model. To check this assumption, we performed a *mouse click count analysis*. The Tobii eye tracker registered the number of times users click on a certain screen area (in this case the “back to the model” button- see Appendix A, stage 3).

After viewing the model (stage 2), users continued to view the question to answer it (stage 3). However, as the model was not visible while answering the question<sup>3</sup>, users had to click “back to the model” in order to see the model again. This process can be repeated till users form a good mental representation of the model required for task solution. If users did not form a good mental representation of the model then they would click on “back to the model” repeatedly.

Table E1 shows the resulting mouse click count analysis. The number 0.79 refers to the average frequency of clicking the “back to the model” button by the good representation REA model group for the first five questions. For example, subject 1 might have hit 3 times the back button for the first five questions (% hit is 0.6) and subject 2 might have hit 5 times the back button for the first five questions (% hit is 1.0). Thus, the average mouse click percentage for these two subjects is 0.8.

**Table E1.** Mouse click count on “back to the model” analysis

Groups	Phase 1 (Q1-Q5) Mean (SD)	t- statistics	Phase 2 (Q6-Q10) Mean (SD)	t- statistics
Good	0.79 (0.33)	t = 3.06, p =0.00*	0.72 (0.20)	t = 2.38, p =0.01*
Poor	1.08 (0.30)		0.89 (0.27)	
Groups	Phase 3 (Q11-Q15) Mean (SD)	t- statistics	Phasev4 (Q16-Q20) Mean (SD)	t- statistics

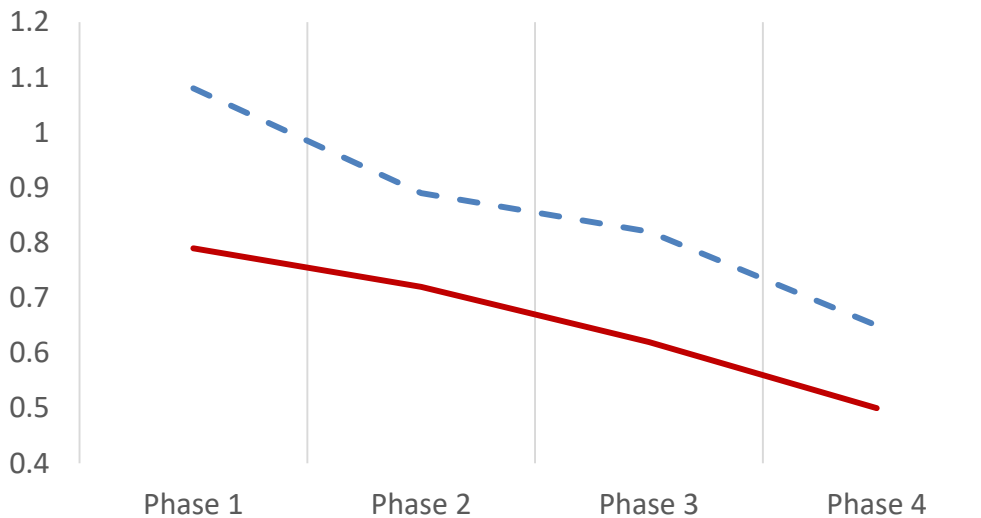
---

<sup>3</sup> The model and the question were intentionally separated into two different screens to get accurate eye movement data. If both question and model appeared in the same screen then it would have been difficult to track the eye movements over the model as the question will distract the user from viewing the model.

Good	0.62 (0.30)	$t = 2.01, p = 0.02^*$	0.50 (0.20)	$t = 2.06, p = 0.02^*$
Poor	0.82 (0.35)		0.65 (0.26)	

\*  $p < 0.05$

The analysis shows that the poor representation REA model group clicked on the “back to the model” button significantly more than the good representation REA model group in all phases. Figure E1 visually depicts this situation where the average click percentage is plotted for the two groups for all four phases.



**Figure E1:** Comparison of mouse click count between good representation REA model users (solid line) and poor representation REA model users (dotted line)

The mouse click count data indicate that initially more iterations between model and question were required to create a good mental representation of the model. Thus, at the beginning of the study, when the subjects were not familiar with the model, they need to refer to the model more frequently to create the mental representation of it. This was more prominent for the poor representation REA model group. But as the experiment progressed, the mental representation of the model was more stable in both groups and therefore the number of times to refer back to the model decreased.