

Journal of the Association for Information Systems

JAIS 

Research Article

Ontological Clarity, Cognitive Engagement, and Conceptual Model Quality Evaluation: An Experimental Investigation

Simon K. Milton

The University of Melbourne
simon.milton@unimelb.edu.au

Jayantha Rajapakse

Monash University
jayantha.rajapakse@infotech.monash.edu.my

Ron Weber

Monash University
ron.weber@monash.edu

Abstract

When analysts build information systems, they document their understanding of users' work domains via conceptual models. Once a model has been developed, analysts should then check it has no defects. The literature provides little guidance about approaches to improve the effectiveness of conceptual model quality evaluation work. In this light, we propose a theory in which two factors have a material impact on the effectiveness of conceptual model quality evaluation work: (a) the ontological clarity of the conceptual models prepared, and (b) the extent to which analysts use a quality evaluation method designed to cognitively engage stakeholders with the semantics of the domain represented by a conceptual model. We tested our theory using an experiment involving forty-eight expert data modeling practitioners. Their task was to find as many defects as possible in a conceptual model. Our results showed that participants who received the conceptual model with greater ontological clarity on average detected more defects. However, participants who were given a quality evaluation method designed to cognitively engage them more with the semantics of the domain did not detect more defects. Nonetheless, during our analysis of participants' protocols, we found that those who manifested higher levels of cognitive engagement with the model detected more defects. Thus, we believe that our treatment for the level of cognitive engagement evoked by the quality evaluation method did not take effect. Based on our protocol analyses, we argue that cognitive engagement appears to be an important factor that affects the quality of conceptual model evaluation work.

Keywords: Conceptual Model, Conceptual Model Quality Evaluation, Ontological Clarity, Cognitive Engagement.

* Jeffrey Parsons was the accepting senior editor. This article was submitted on 28th April 2010 and went through three revisions.

Volume 13, Issue 9, pp. 657-694, September 2012

Ontological Clarity, Cognitive Engagement, and Conceptual Model Quality Evaluation: An Experimental Investigation

1. Introduction

A conceptual model represents (often graphically) someone's or some group's perceptions of certain semantics in some domain. They are usually constructed by information systems professionals (e.g., a systems analyst), and might be employed in various ways – for instance, to design databases, to evaluate an enterprise system's capabilities, or to formulate queries on a data warehouse.

Defects in conceptual models can have costly consequences (e.g., Maes & Poels, 2007; Moody, 2005). For instance, they can lead to misinterpretation of user requirements and to erroneous system designs. Ultimately, they can be a root cause of system failure. They might also lead to the formulation of incorrect queries against a database, the reliance on incorrect information obtained via these queries, and costly decision-making errors. Because of the deleterious outcomes that can occur when conceptual models contain defects, those who build them should evaluate their quality carefully.

Despite the importance of having high-quality conceptual models, to the best of our knowledge, little prior research exists on methods that can be used to evaluate conceptual models or the factors associated with the effectiveness of conceptual model evaluation work. In this regard, Connolly and Begg (2010, pp. 433-435) provide a more extensive discussion than most sources on some approaches that might be used to evaluate conceptual models. Presumably, certain characteristics of the approaches they advocate are associated with the extent to which they can be used effectively. To the best of our knowledge, however, such characteristics have not been researched systematically. Moreover, certain characteristics of conceptual models are likely to make them more or less amenable to evaluation. For instance, presumably, the extent to which a model can be easily understood is associated with how effectively it can be evaluated. In this regard, Genero, Poels, and Piattini (2008) provide evidence to show certain structural characteristics of a conceptual model are associated with how easily it can be understood. In addition, certain characteristics of the stakeholders who undertake conceptual model evaluation work are likely to be associated with the effectiveness of this work. For instance, Dunn, Gerard, and Grabski (2005) provide evidence to show that some types of human cognitive biases affect how well stakeholders undertake conceptual model evaluation work.

In light of the paucity of prior research that has been done, in this paper we describe research we undertook to try to find more effective ways to evaluate the quality of conceptual models. We assess effectiveness in terms of how well the quality evaluation work leads to the identification of defects in a conceptual model's representation of a domain's semantics. The research we describe is part of a longer-term program that we are undertaking in which we are (a) determining whether some approaches to preparing conceptual models result in models that are easier to evaluate, and (b) assessing the strengths and weaknesses of different approaches that stakeholders might use when evaluating conceptual models (so they can then make better choices).

The remainder of our paper proceeds as follows. Section 2 presents the theory and propositions that underpin our empirical work. Section 3 describes our research method. In Section 4, we indicate how we analyzed our data and present the results we obtained. In Section 5, we discuss our results, identify some implications of our results for research and practice, and conclude.

2. Theory and Propositions

The theory we propose is what Gregor (2006, pp. 626-628) calls a Type-IV theory – that is, a theory that has been developed for both explanatory and predictive purposes. Figure 1 provides a diagrammatic overview of our theory. The following subsections articulate the theory's components in more detail and the propositions that it motivates.

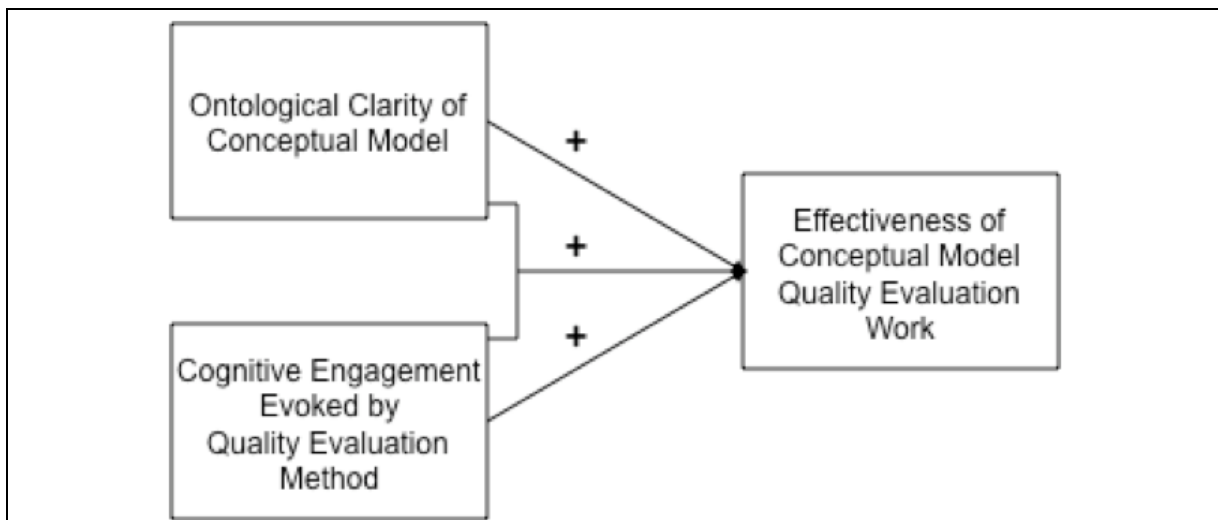


Figure 1. Theoretical Model of the Effects of Ontological Clarity and Cognitive Engagement on the Effectiveness of Conceptual Model Quality Evaluation Work

2.1. Constructs in the Theory

Our theory has three constructs. One is an endogenous (outcome) construct – namely, the effectiveness of the conceptual model quality evaluation work undertaken. Two are exogenous (causal) constructs – namely, (a) how clearly the conceptual model to be evaluated represents the semantics of a domain (the model's ontological clarity), and (b) the extent to which the quality evaluation method used leads stakeholders to engage cognitively with the semantics of the domain represented by the model.

2.1.1. Effectiveness of Conceptual Model Quality Evaluation Work

Moody (2005, p. 252) defines the quality of a conceptual model as “[t]he totality of features and characteristics of a conceptual model that bear on its ability to satisfy stated or implied needs”. His definition is consistent with the concept of quality used in international standards (e.g., International Standards Organisation, 2005).

We can evaluate the quality of a conceptual model in several ways. For instance, the following factors might be considered: whether or not the model contains syntactic errors (relating to the conceptual modeling grammar used), how well the model represents a particular user's perceptions of a domain's semantics (perceived semantic quality); whether the model was developed using a consultative approach; and whether the model is perceived to be easy to use and useful (e.g., Krogstie, Lindland, & Sindre, 1995; Lindland, Sindre, & Sølvsberg, 1994; Maes & Poels, 2007; Moody, 2005; Moody & Shanks, 2003). Ultimately, some type of summative judgment based on the values assigned to these types of factors must be made on a model's quality.

The stakeholders who evaluate a conceptual model must employ a set of normative semantics to assess the model (their perceptions of the semantics the conceptual model ought to represent). Burton-Jones, Wand, and Weber (2009) point out that stakeholders are likely to invoke two kinds of semantics. First, they will use denotational semantics. These are the *prima facie* or commonly accepted semantics associated with the focal domain. Second, they will use connotational semantics. These reflect how stakeholders actually interpret a model while they seek to ascribe meaning to a focal domain (the focus of pragmatics in linguistics). In this regard, they may use factors such as their prior experience, extant knowledge, and current work context to construct their interpretation of the semantics of the domain.

The information systems professionals who construct a conceptual model are expected to seek to represent the denotational semantics associated with the focal domain for two reasons. First, a conceptual model cannot accommodate conflicting perceptions of a domain's semantics. In this light,

the semantics represented in the model often reflect a compromise reached among stakeholders. Second, to some extent, connotational semantics may be unique to a particular stakeholder. Moreover, because of their sometimes implicit nature, connotational semantics may be inaccessible via the “standard” information-elicitation techniques used by information systems professionals during the development of an information system.

Nonetheless, both denotational and connotational semantics are important when stakeholders evaluate the quality of a conceptual model. Their primary concern is likely to be how clearly the conceptual model represents the denotational semantics associated with the focal domain (and it is this issue that is the primary focus of our research). If multiple, egregious violations of stakeholders’ connotational semantics exist in the conceptual model, however, the likely success of the system diminishes.

The effectiveness of quality evaluation work undertaken can be conceptualized in both a relative sense and an absolute sense. In a relative sense, one set of quality evaluation work is more effective than another set if it produces a higher defect count. In an absolute sense, the effectiveness of a set of quality evaluation work can be assessed in terms of the proportion of the total defect count that it achieves. In practice, an absolute measure of effectiveness is problematic to determine because the total defect count associated with a conceptual model is not known. In an experimental context, however, defects can be seeded deliberately in a model, which allows absolute measures to be determined based on the denotational semantics that the model is intended to represent. A more refined measure of effectiveness would also take into account (a) the materiality of the defects identified using a particular quality evaluation method, and (b) the extent to which a stakeholder’s connotational semantics are represented in the conceptual model.

2.1.2. Ontological Clarity of Conceptual Model

Wand and Weber (1993) have defined a characteristic of conceptual modeling grammars and conceptual models (sometimes called conceptual modeling scripts) that they call “ontological clarity”. While the focus of our theory is conceptual models, the ontological clarity of a model depends, in part, on the ontological clarity of the grammar used to generate it.

The ontological clarity of a conceptual modeling grammar is a function of the extent to which a one-one mapping exists between the set of modeling constructs in the grammar and the set of constructs in a particular theory of ontology (a theory that provides a taxonomy of constructs that can be used to describe different generic types of phenomena that occur in the real world). In this regard, Wand and Weber (1993, pp. 228-233) argue the mapping between grammatical constructs and ontological constructs highlights three situations that undermine the ontological clarity of a conceptual modeling grammar¹:

- Construct overload: A single modeling construct maps to two or more ontological constructs.
- Construct redundancy: Two or more modeling constructs map to a single ontological construct.
- Construct excess: A modeling construct does not map onto any ontological construct.

When instances of construct overload, redundancy, and excess exist in a conceptual modeling grammar, Wand and Weber (1993) argue that the meaning of the grammatical constructs will be unclear. As a result, the meaning of the models (usually expressed in the form of diagrams) generated via the grammar to represent some real-world domain may be unclear. Nonetheless, sometimes the problems that arise as a result of construct overload, redundancy, and excess can be mitigated (at least

¹ Wand and Weber (1993, pp. 226-227) highlight a fourth problematic mapping that they call “construct deficit” (or “incompleteness”) in which an ontological construct exists that is not the image of any grammatical construct. When construct deficit exists, some type of real-world phenomenon cannot be represented by the grammar. Thus, in some cases, models generated via the grammar will provide incomplete representations of a real-world domain. Wand and Weber (1993, p. 233) define the “ontological expressiveness” of a conceptual modeling grammar in terms of its ontological clarity (lack of construct overload, redundancy, and excess) and lack of construct deficit.

to some extent) through disciplined use of the grammar when generating a model (Wand & Weber 1993, p. 234; Recker, Rosemann, Indulska, & Green, 2009, pp. 353-354). For instance:

- Where construct overload exists, an overloaded construct can be annotated in different ways so that a particular annotation of the construct maps only to a single ontological construct.
- Where construct redundancy exists, only one of the redundant grammatical constructs might be used to represent the target ontological construct.
- Where construct excess exists, the excess construct is not used to represent any phenomena.

The ontological clarity of a conceptual model is a function, therefore, of the number of instances of grammatical constructs that suffer from overload, redundancy, or excess that appear in the model. As more instances of problematic grammatical constructs are used to represent the domain, the ontological clarity of the conceptual model is undermined. The extent to which problematic instances of constructs appear depends on (a) the ontological clarity of the grammar used to generate the model, and (b) whether instances of problematic grammatical constructs have been avoided through disciplined use of the grammar.

The usefulness of ontological clarity as a theoretical construct, however, is a function of the quality of the ontological constructs that are the target of the mapping applied to the grammatical constructs. The ontological constructs must make sense to the stakeholders associated with the focal domain (Milton & Kazmierczak, 2004). Otherwise, the concepts of overload, redundancy, and excess have little power as a means of understanding, predicting, or explaining how clearly a conceptual model represents a focal domain. Instead, they are simply names for the outcomes of a mapping process. In short, the ontology must resonate with stakeholders through either explicit training in the ontology or vicarious engagement with the ontology (e.g., because it has been used to inform the creation of conceptual models that stakeholders examine).

2.1.3. Cognitive Engagement Evoked by Quality Evaluation Method

A conceptual modeling quality evaluation method comprises the set of artifacts (e.g., checklists) and processes (e.g., interviews) used to detect a model's defects. Historically, the following types of methods have been proposed (e.g., Connolly & Begg, 2010; Elmasri & Navathe, 2011; Simsion & Witt, 2005).

- Independent study of conceptual model by a stakeholder other than the model creator: Using this method, the creator of a conceptual model gives the model to another stakeholder (e.g., a domain expert or systems analyst) who is asked to evaluate the model independently with the goal of identifying defects in the semantics represented by the model and then report their conclusions to the model creator.
- Joint examination of conceptual model by another stakeholder and the model creator: Using this method, another stakeholder and the model creator examine the conceptual model together. Both can raise questions with each other to identify defects in the model.
- Structured questioning of another stakeholder by model creator: Using this method, the model creator proceeds systematically through various components of a conceptual model. The model creator describes the semantics represented by each component, and asks another stakeholder whether they can identify any defects in the model's representation of the application domain semantics.
- Use-case/test data: Using this method, either the model creator or other stakeholders evaluate a conceptual model by examining how well it accommodates a set of sample data that are representative of the domain the model is intended to represent.

Recently, some new quality evaluation methods have been proposed in research (experimental evaluations of conceptual models (e.g., Shanks, Tansley, & Weber, 2003).

- Use of free-recall tasks: Using this method, the model creator gives another stakeholder a period of time to study the model. The model is then removed, and the other stakeholder is asked to reconstruct the model from memory. The goal of the recall task is to lead the other stakeholder to improve their understanding of the semantics represented by the model. The model creator and the other stakeholder then jointly compare the output of the recall task with the model. Differences between the two provide pointers to possible defects in the model.
- Use of comprehension tasks: Using this method, the model creator first formulates a set of questions about the model. The goal of the questions is to test another stakeholder's understanding of the semantics represented by the model. If the other stakeholder provides wrong answers to the questions, further discussion occurs between the model creator and the other stakeholder about whether the model is a defective representation of the domain's semantics.
- Use of problem-solving tasks: Using this method, the model creator formulates problems that another stakeholder must solve. The answers to the problems can then be used to motivate a discourse between the model creator and the other stakeholder about whether the model is an accurate and complete or a defective representation of the domain's semantics.

The construct of interest in our theory is the extent to which a quality evaluation method leads stakeholders to engage cognitively with the application-domain semantics represented by a conceptual model. As the level of cognitive engagement increases, stakeholders are more likely to develop an understanding of the domain semantics represented by the model.

The concept of cognitive engagement has been investigated in a number of contexts – for instance, student learning (e.g., Corno & Mandinach, 1983; Webster & Hackley, 1997) and work performance (e.g., Bakker, Schaufeli, Leiter, & Taris, 2008). Cognitive engagement is characterized in various ways: vigor, absorption, flow, dedication, focus, sustained attention, and energy in action. Moreover, in many contexts, cognitive engagement is associated with superior task performance (e.g., Radosevich, Radosevich, Riddle, & Hughes, 2008).

Our focus is the extent to which a quality evaluation method leads stakeholders to undertake cognitive activities that elicit a full understanding of the semantics represented in the conceptual model. In this regard, we believe two characteristics of a quality evaluation method promote more cognitive engagement with these semantics.

1. Structure: Stakeholders will engage more forcefully with the semantics of the domain that a conceptual model is intended to represent when a quality evaluation method provides a well-defined structure for evaluating the model. Structure promotes confidence among stakeholders that they can undertake the quality evaluation task competently (feelings of self-efficacy). Prior research has shown effective cognitive engagement is unlikely to occur when individuals feel the task they have been given is beyond their abilities (e.g., Bakker et al., 2008).
2. Task challenge: Stakeholders will engage more forcefully with the semantics of the domain that a conceptual model is intended to represent when a quality evaluation method invokes challenging tasks. Cognitive challenges pique the interest of humans, which leads them to engage more with a task (e.g., Marks, 2000), especially when they have a purposeful goal (such as validating a conceptual model).

A particular quality evaluation exercise might use several quality evaluation methods to leverage the strengths and compensate for the weaknesses of each method.

2.2. Associations in the Theory

Figure 1 shows that our theory has three associations. The first two are direct (main) effects. The third is an interaction effect between the two exogenous constructs.

2.2.1. *Ontological Clarity of Conceptual Model and Effectiveness of Quality Evaluation Work*

This first association is founded on the assumption that a high-quality ontology has been chosen to evaluate the ontological clarity of a grammar and any conceptual models generated using the grammar². In other words, the constructs in the ontology must resonate with stakeholders in terms of the meaning they ascribe to different types of real-world phenomena.

If a high-quality ontology has been chosen to evaluate ontological clarity, an ontologically clear conceptual model will elucidate the semantics associated with a domain in an unambiguous way. Specifically, it will assign each phenomenon in the focal domain to an ontological class. Each ontological class has certain semantics, which are defined by the ontology. Instances of the ontological class assume these semantics.

For instance, in Bunge's (1977) ontology, the ontological construct called "thing" has a particular meaning. Any phenomenon classified as a "thing" has an independent existence, has properties, might be a component of another thing, might interact with other things, and so on. If a conceptual modeling grammar uses one and only one grammatical construct to represent a "thing", and if a model creator complies with the rules of the grammar, instances of the construct in a model will always represent a thing (and only a thing). As a result, stakeholders who examine the model and know it complies with Bunge's ontology will be able to make reliable, additional inferences about the semantics of the focal domain (based on their knowledge of the semantics of the "thing" construct). Those who examine the model but have little or no knowledge of Bunge's ontology still ought to be able to make reliable, additional inferences about the semantics of the focal domain. They should see an underlying consistency in the way the model represents the semantics of the focal domain. As a result, they will either explicitly or implicitly educe the semantics represented by the grammatical construct in due course.

Nonetheless, we argue that an ontologically clear conceptual model assists its readers to educe both denotational and connotational semantics associated with the domain the conceptual model is intended to represent. In the case of denotational semantics, an ontologically clear model represents the focal domain in a more accurate and more consistent way. Because readers can educe the denotational semantics more easily, they then have more cognitive resources available to construct their connotational semantics. Moreover, variations among readers in relation to their connotational semantics may be reduced.

We also argue that an ontologically clear conceptual model will allow stakeholders to identify domain phenomena that have not been represented in the model (but should have been represented in the model). Because the meaning of the phenomena represented in the model is clear, stakeholders are less likely to assume incorrectly that missing phenomena have in fact been represented. Moreover, they will have more cognitive resources available to identify missing phenomena.

In this light, we hypothesize that a direct effect exists between the ontological clarity of the conceptual model to be evaluated and the effectiveness of the quality evaluation work undertaken. Specifically, as the ontological clarity of a conceptual model increases, the effectiveness of the quality evaluation work undertaken will be higher (see also Shanks et al., 2003). Because domain stakeholders will find ontologically clear conceptual models easier to understand and thereby have more cognitive resources available to evaluate the model, they will be better able to identify defects in the model's representation of the focal domain's semantics. In short, our theory motivates the following main effect:

² Recall that, even if a grammar is deemed to be ontologically clear using a high-quality ontology, models generated using the grammar might be unclear if the grammar is employed in undisciplined ways. Similarly, disciplined use of an ontologically unclear grammar might still enable ontologically clear models to be generated using the grammar.

Proposition 1: *Conceptual models that possess higher levels of ontological clarity will enable quality evaluation work to be more effective.*

2.2.2. Cognitive Engagement Evoked by Quality Evaluation Method and Effectiveness of Quality Evaluation Work

We hypothesize a direct effect exists between the extent to which a quality evaluation method results in a stakeholder engaging cognitively with the semantics represented in a conceptual model and the effectiveness of the quality evaluation work undertaken. As the level of cognitive engagement increases, stakeholders come to understand better the semantics the model represents. As a result, they are better able to evaluate whether these semantics correspond to the focal domain's semantics (including whether any domain semantics are missing from the model). In short, our theory motivates the following main effect:

Proposition 2: *Quality evaluation methods that result in stakeholders having higher levels of cognitive engagement with the semantics represented by a conceptual model will enable quality evaluation work to be more effective.*

2.2.3. Ontological Clarity of Conceptual Model, Cognitive Engagement Evoked by Quality Evaluation Method, and Effectiveness of Quality Evaluation Work

We hypothesize that an interaction effect exists between the ontological clarity of the conceptual model to be evaluated, the quality evaluation method used, and the effectiveness of the quality evaluation work undertaken. If a quality evaluation method evokes little cognitive engagement by a stakeholder in relation to the conceptual model to be evaluated, the fact that the model is ontologically clear is likely to have little or no effect on the number of defects identified. On the other hand, because a quality evaluation method evokes more cognitive engagement by a stakeholder with the conceptual model to be evaluated, ontological clarity is likely to more significantly influence the effectiveness of the quality of evaluation work undertaken. In short, our theory motivates the following interaction effect:

Proposition 3: *A quality evaluation method will be more effective when it is undertaken in conjunction with an ontologically clear conceptual model.*

2.3. Boundaries of the Theory

In four contexts, we believe the explanatory and predictive power of our theory will be weak or non-existent. These contexts circumscribe the boundaries of our theory.

The first is where the domain is simple. As a result, the conceptual model that represents the domain is also simple. Under these circumstances, the quality evaluation task is straightforward. Defects in the conceptual model should therefore be easy to identify. Neither ontological clarity nor the level of cognitive engagement evoked by the quality evaluation method used are likely to have much, if any, impact on the effectiveness of the quality evaluation work.

The second is where a poor-quality ontology is used to evaluate the ontological clarity of a conceptual model. Unless an ontology covers all phenomena of interest in a focal domain, provides clearly defined and meaningful ontological constructs, and articulates ways of mapping different phenomena of interest to their associated ontological constructs, concepts such as construct overload, redundancy, excess, and deficit have little or no import. They will not underpin understanding of or predictions or explanations about the behavior of stakeholders who use a conceptual modeling grammar or conceptual model.

The third is where stakeholders who execute the quality evaluation method have little knowledge of the focal domain. If they do not have at least some general understanding of the domain, they are unlikely to be capable of identifying defects in a conceptual model, even if the model is ontologically clear. Similarly, the use of a quality evaluation method that results in stakeholders engaging cognitively with the semantics represented by the conceptual model is likely to have little effect if the stakeholders do not sufficiently understand the focal domain's semantics to be able to make sense of the conceptual model.

The fourth is where stakeholders who execute the quality evaluation method lack competence in the method. If they do not know how to execute the method competently, the method is unlikely to be an effective means of engaging them cognitively with the conceptual model's semantics. Thus, they will remain with only a cursory understanding of the domain semantics represented by the conceptual model. Under these circumstances, defects in the model are unlikely to be identified.

3. Research Method

Siau and Rossi (2007) describe different empirical research techniques that have been used to evaluate the merits of conceptual modeling methods and the strengths and limitations of these techniques. Given the exploratory nature of our research and our focus on explanation and prediction rather than description, we chose an experimental technique to try to obtain preliminary evidence in support of our propositions. If support were to be forthcoming, we would then be better able to justify more costly field research (e.g., action research) to investigate our propositions.

3.1. Design

Our experimental design had two factors. The first, which we called "ontological clarity", had two levels: (1) high clarity, and (2) low clarity. The second, which we called "cognitive engagement of quality evaluation method", also had two levels: (1) high engagement, and (2) low engagement. The two factors were fully crossed (2×2); thus, we had four treatments in our design.

As we indicate in Section 2.1.1, an evaluation of a conceptual model's quality covers multiple dimensions, involves a number of different types of stakeholders, and engages various quality evaluation processes. In an experiment, we cannot investigate all dimensions, involve all types of stakeholders, or consider all types of quality evaluation processes. As such, we constrained the scope of our experiment in four ways.

First, we used a fairly well-defined quality evaluation task. Specifically, it involved having participants evaluate whether a conceptual model that they were given provided an accurate and complete representation of the semantics described in a narrative about a manufacturing domain. In a real-world context, a systems analyst might have prepared the narrative and conceptual model based on interviews with end users. We sought to reduce error variance in our experiment by using a domain and a task that we felt primarily would engage denotational rather than connotational semantics.

Second, we used a single quality outcome variable in our experiment – namely, the number of defects that participants in our experiment correctly identified in the conceptual model they were given. A defect was either an error or omission in the model. An error was an incorrect representation in the model of any item of domain semantics described in the narrative. An omission was any item of domain semantics described in the narrative that the conceptual model failed to represent.

Third, we had only one type of stakeholder act as participants in our experiment. Specifically, we used data modeling experts because we wanted to reduce error variance in our data that might arise because participants were unsure about and/or lacked experience with conceptual modeling grammars. In essence, our participants were playing the role of information systems professionals who had to evaluate the work of another information systems professional (e.g., a systems analyst).

Fourth, we examined the effect of ontological clarity on the quality evaluation process using only a single grammar – namely, the unified modeling language (UML) (Rumbaugh, Jacobson, & Booch, 2005). We chose UML because of its prominence as a modeling grammar. Moreover, Davies, Green, Rosemann, Indulska, and Gallo (2006) report that UML is one of the most frequently used modeling grammars among the practitioners they surveyed. We used the UML grammar in a disciplined way to generate an ontologically clear conceptual model and a somewhat undisciplined way to generate an ontologically unclear conceptual model³. By using only a single modeling grammar rather than two grammars (one having greater ontological clarity than the other), we sought to reduce error variance

³ We sought to avoid or minimize construct overload, redundancy, and excess by using the type of disciplined approach we discussed above in Section 2.1.2. In the case of our somewhat undisciplined approach, we still conformed to the rules of the grammar.

in our data that might arise because of differences in the grammars other than ontological clarity (e.g., the understandability of the different syntaxes they used).

3.2. Materials

We developed four sets of experimental materials. They instantiated the treatments in our experiment and enabled us to measure values for our outcome variable.

The first set of materials was developed for use in a warm-up task that participants undertook prior to beginning the primary experiment. The materials comprised (a) a brief narrative describing a human-resources domain, and (b) a conceptual model (drawn using UML constructs) that purportedly represented the semantics of the domain⁴. Two versions of the conceptual model were developed: one was ontologically clear, and the other was ontologically unclear.

The second set of materials was developed for use in the primary experiment. The materials comprised (a) a case description of a manufacturing domain with ontologically clear (Figure 2) and ontologically unclear (Figure 3) conceptual models (also drawn using UML constructs), and (b) a data dictionary describing each object and attribute shown in the associated conceptual model (ontologically clear or ontologically unclear).

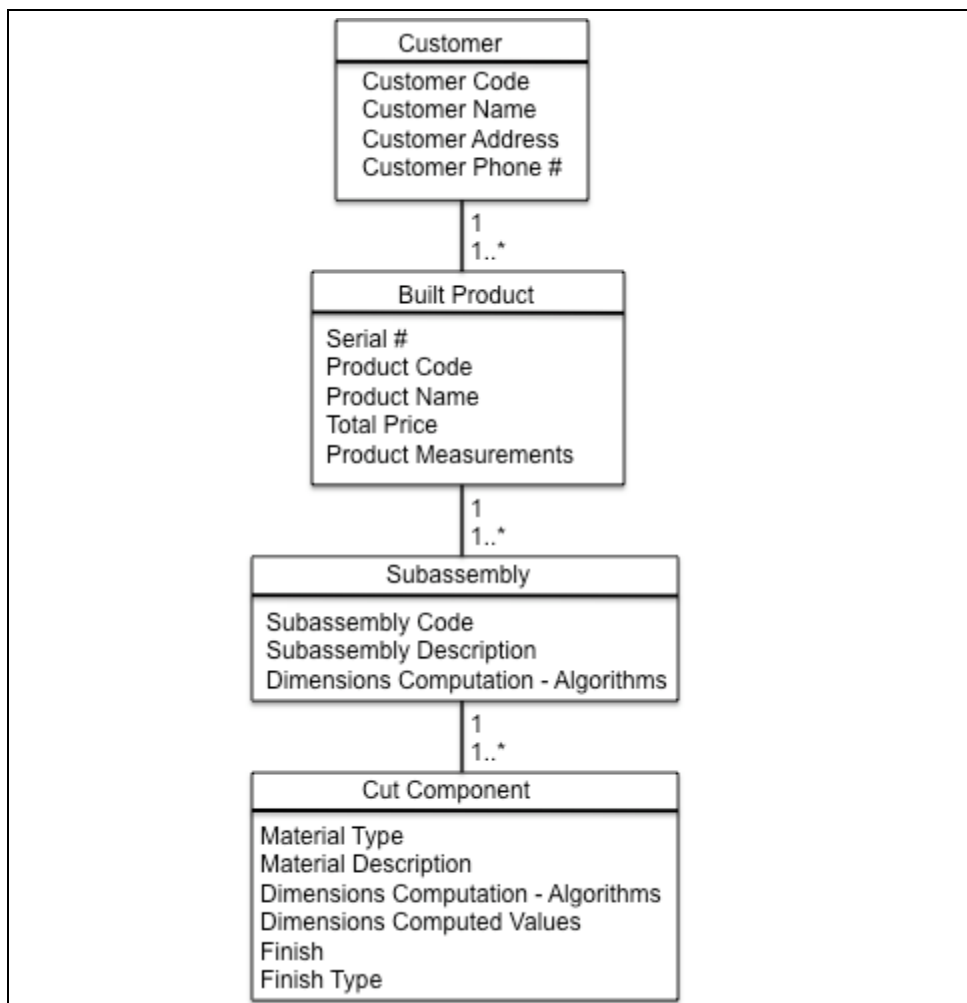


Figure 2. Ontologically Clear Model with Defects

⁴ Note that we employed simplified UML class diagrams because we did not include the names of methods or operations in our object classes. We omitted methods/operations because dynamic phenomena were not the focus of our research.

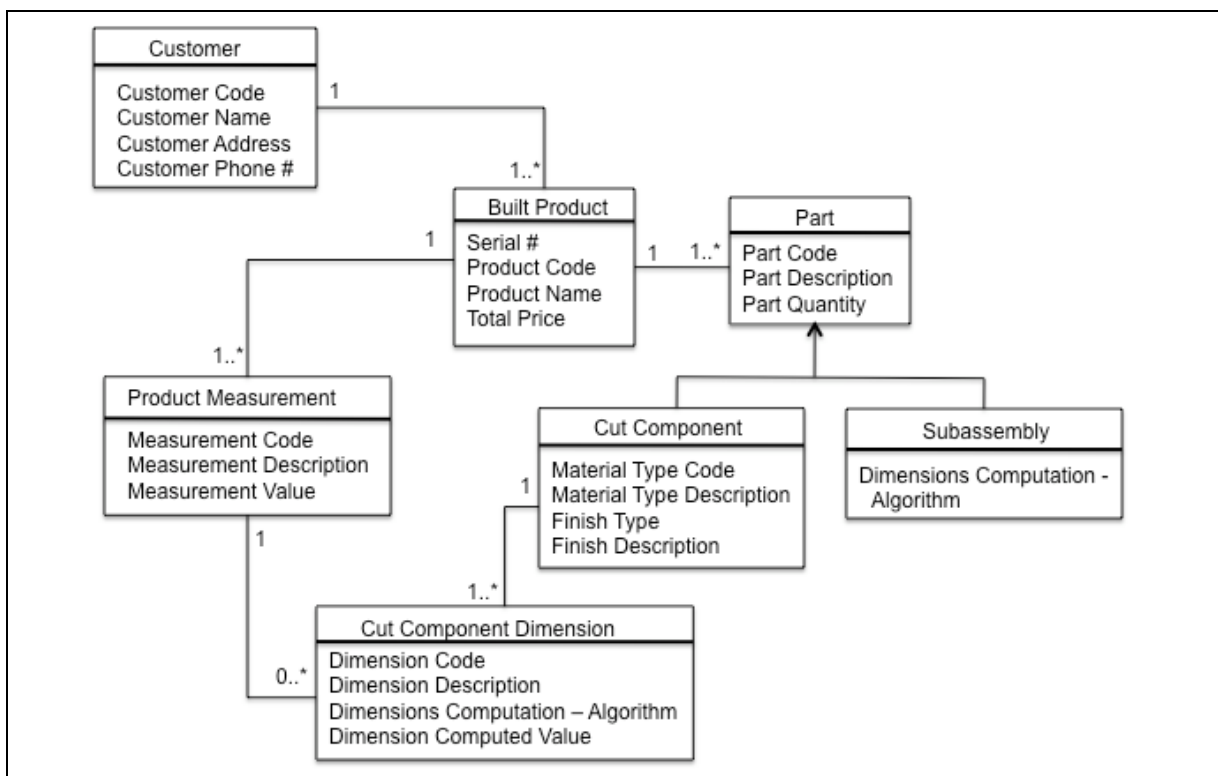


Figure 3. Ontologically Unclear Model with Defects

The materials instantiated the two levels of the “ontological clarity” factor in our experiment. The ontology we used to inform the design of our ontologically clear and ontologically unclear models was Bunge’s (1977) ontology (as modified and extended by Wand & Weber, 1990). We chose Bunge’s ontology because a growing body of research suggests it provides a useful basis for predicting and explaining the experiences of stakeholders when they engage with conceptual models (e.g., Bodart, Sim, Patel, & Weber, 2001; Bowen, O’Farrell, & Rohde, 2006; Burton-Jones & Meso, 2006; Evermann & Wand, 2005; Green & Rosemann, 2000; Opdahl & Henderson-Sellers, 2002; Recker, Rosemann, Green, & Indulska, 2011; Shanks, Tansley, Nuredini, Tobin, & Weber, 2008; Soffer & Hadar, 2007).

To prepare the second set of materials, we first employed an information systems professional who was highly experienced and widely acknowledged as an expert consultant in data modeling. We asked him to prepare case materials for a domain that was moderately complex but nonetheless could be understood by expert data modelers who had a wide range of backgrounds and experience. Based on one of his consulting assignments, he prepared a case description, conceptual models, and a data dictionary for a medium-size organization engaged in made-to-order manufacturing of aluminum-based products for residential and commercial buildings (e.g., shower screens, security doors, and leafless guttering). For our primary experiment, we chose a subset of the materials he prepared – specifically, the case description (Appendix A), conceptual models, and data dictionary pertaining to recording the measurements of shower screen components to be built for customers. We felt this subset was sufficiently rich to allow us to test our propositions but at the same time not overly complex. Moreover, we felt that anyone who had expertise in data modeling could understand the subset and, as such, denotational semantics would dominate connotational semantics. Because the materials were also based on a real case, they also had a high level of external validity.

The second set of materials was refined iteratively. We progressively removed those aspects of the consultant’s model that did not conform to Bunge’s (1977) ontology and replaced them with instances of modeling constructs that did conform to his ontology. For instance, all object classes in the UML models had to represent classes of things (rather than properties in general), and we removed optional attributes and associations from his model. The consultant was interested in the

ways we modified his model, and he engaged with us in a useful debate about the merits of the different models we prepared. For example, he was surprised to see that we did not model an attribute in general such as “product measurement” as a separate class in the ontologically clear model. From the perspective of database design, he recognized that “product measurement” was a repeating group. As a result, he felt it should be represented as a separate class because it would need to be instantiated as a separate table in a normalized database. We explained carefully that our ontologically clear model was founded on Bunge’s (1977) ontology and not database design considerations. He was concerned, also, that ultimately the differences between our models were not so great that we would be comparing “apples with oranges” in our experiment. Care was taken in the refinement process, therefore, as evidenced by the time taken in preparing the materials (nine months’ elapsed time) and the activities undertaken (five two-hour joint workshops and approximately 30 hours of separate work by the consultant).

Eventually, we settled on two models that were not so different from each other that they might be deemed to represent fundamentally different domain phenomena. Figures 4 and 5 show these two “benchmark” models. Note that the two benchmark models contain only the attributes used in the application domain that was our focus. The object classes may have other attributes to support other application domains.

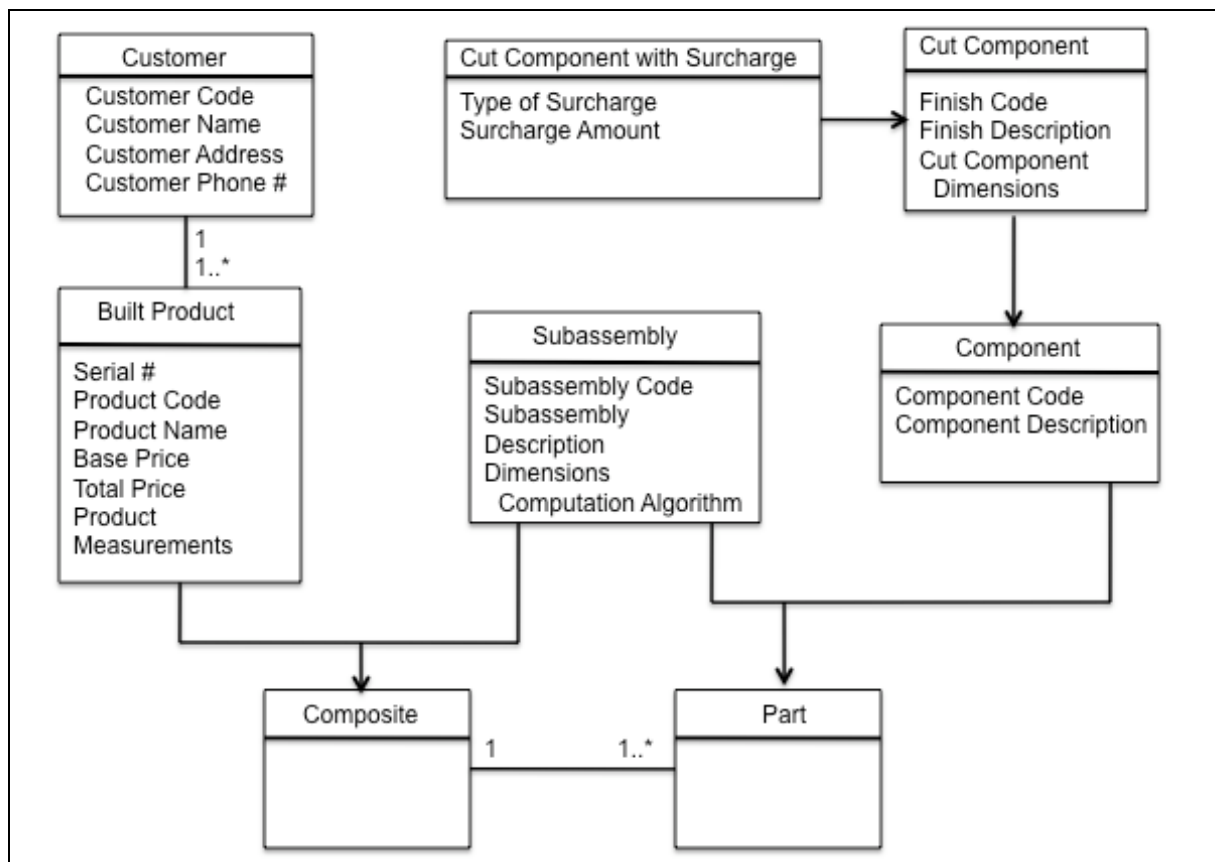


Figure 4. Ontologically Clear Benchmark Model

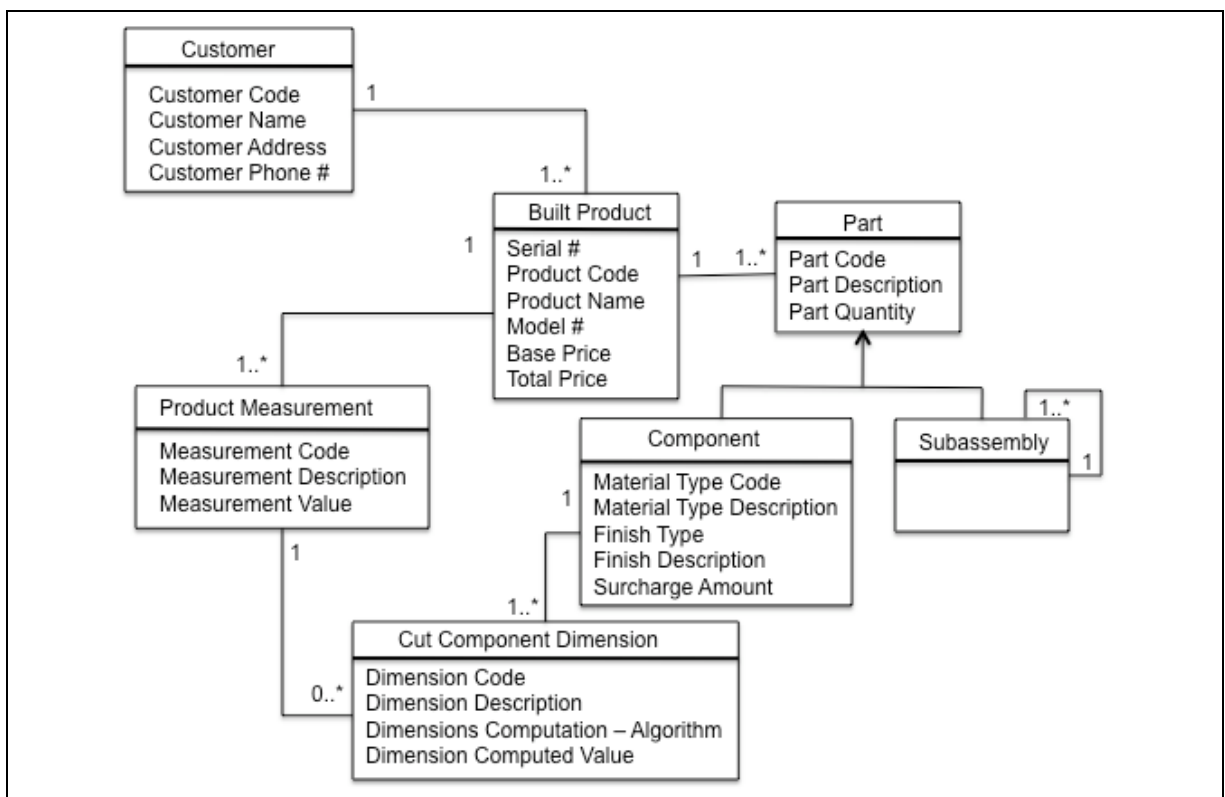


Figure 5. Ontologically Unclear Benchmark Model

In the case of the ontologically unclear conceptual model shown in Figure 5, Table 1 (next page) shows those of its characteristics that undermine its ontological clarity. The first two characteristics are instances of construct overload because a single grammatical construct (an object class) has been used to represent two ontological constructs (a thing and an attribute of a thing). The third characteristic is an instance of construct excess because a grammatical construct (an optional association) does not map to any ontological construct (the ontological benchmark proscribes optional attributes or associations).

As a final step, we seeded both the ontologically clear and ontologically unclear conceptual models with errors and omissions (Table 2, next page) so that they were no longer faithful representations of the case-study domain. We were also careful to ensure that participants in the experiment could be expected to detect the errors we seeded (Appendix B). Our objective was to create experimental materials that allowed us to investigate whether an ontologically clear model facilitated detection of the errors and omissions in Table 2 and an ontologically unclear model inhibited detection of them.

In choosing the defects that we seeded in the model, we tried to include instances of a range of defect types that might be encountered in practice. Some were motivated by discussions we had with the data modeling expert who developed the case study that formed the basis for our experimental materials. Some, we believed, would have been easy to identify; others we believed would have been difficult to identify. Some would have had serious consequences for subsequent system design work; others would have had only minor consequences for subsequent system design work. Given the exploratory nature of our research work, our goal was to try to tease out what effects, if any, ontological clarity and the level of cognitive engagement evoked by a quality evaluation method had on participants' ability to detect defects in a conceptual model.

Table 1. Characteristics of Conceptual Model Shown in Figure 5 that Undermine Ontological Clarity

Characteristic	Reason ontological clarity undermined	References
Product measurement represented as an object class (class of things).	Product measurement is an attribute in general of a class of things and not a class of things itself.	Bunge (1977, pp. 57-62); Weber (1996); Corral, Schuff, and St. Louis (2006)
Cut component dimension represented as an object class (class of things).	Cut component dimension is an attribute in general of a class of things and not a class of things itself.	Bunge (1977, pp. 57-62); Weber (1996); Corral et al., (2006)
Product measurement has an optional association with cut component dimension.	Things do not have optional mutual attributes (associations).	Bunge (1977, pp. 60-61); Wand, Storey, and Weber (1999); Bodart et al. (2001); Gemino and Wand (2005); Gemino (1999)

Table 2. Defects Seeded into Conceptual Models

Defect no.	Explanation	Type of defect	Ont. clear model	Ont. unclear model
1	Subassembly object class is not shown as a subclass of part object class.	Omission	√	
2	Component object class is not shown as a subclass of part object class.	Omission	√	
3	Cut-component object class is not shown as a subclass of component object class.	Omission	√	√
4	Component-with-surcharge object class is not shown as a subclass of cut-component object class.	Omission	√	√
5	Fact that a subassembly comprises one or more other subassemblies and/or one or more parts is not shown.	Omission	√	√
6	Instances of built-product object class are shown as always having one or more subassemblies.	Error	√	
7	Instances of built-product object class are shown as always having one or more subassemblies and/or cut components (note, a built product could comprise components only, none of which are cut components).	Error		√
8	Instances of subassembly object class are shown as always having one or more cut components (note, a subassembly could comprise components only, none of which are cut components).	Error	√	
9	Missing attribute "base price" in built-product object class.	Omission	√	√
10	Missing attribute "model no." in built-product object class.	Omission	√	√
11	Missing attribute "surcharge amount" (might have been shown in cut-component object class even though not all cut components have a surcharge).	Omission	√	√
12	Additional attribute "dimension computation algorithm" in subassembly object class.	Error	√	√

As we seeded the model with defects, it became clear that some defects could be included in both models whereas others could be included in only one model. For instance, a missing attribute could be seeded as a defect in both models, whereas a cardinality error pertaining to an object class that represented a repeating group of attribute values could be seeded only in the ontologically unclear model. We decided, therefore, to seed defects that were common to both models as well as a few defects that were specific to a particular model. We were interested in the outcomes with the common defects as well as those defects that were unique to a particular model.

The third set of materials was a table with five rows where the rows showed representative instances of phenomena in the domain (e.g., things, properties, events) that the conceptual model was supposed to represent. Table 3 shows an excerpt from this table. In essence, each row of the table can be conceived as a use case. It was the means by which we instantiated the two levels of “cognitive engagement of quality evaluation method” factor in our experiment.

We predicted that participants in our experiment would match the contents of the table against the conceptual model they received to determine whether each element of the table could be mapped to a construct in the conceptual model. Similarly, they could map constructs in the conceptual model to elements in the table. They could also check the correspondence between the table, the conceptual model they had been given, and the written case description, which we predicted would lead participants to engage more with the application-domain semantics represented by the conceptual model that they had been given (their cognitive engagement). For example, the case description (Appendix A) indicates that products have a base price, and Table 3 (use-case) shows a column for base price. If participants tried to map the base price to the conceptual models (Figures 2 and 3), they would see quickly that this attribute is missing from the model that they had been given. In these kinds of ways, we predicted that participants would be able to use the table to come to a deeper understanding of the domain represented by the conceptual model that they were given to evaluate. As a result, we believed that they would be better able to determine how clearly the conceptual model represented the domain. Indeed, during a pilot test of the experimental materials (see below), we found our participants had employed the use-case table in these types of ways.

Table 3. Excerpt from Use-Case Table

Customer	Serial no	Product code	Product name	Base price	Total price	Product measurements
Steve	A000002	SS3P	Shower screen	400	460	2010, 1010, 1005
Lyn	A000004	MWD2P	Mirrored wardrobe door	500	530	2000, 1500, 1500
Parts of the product			Materials of the component		Surcharge type	
Sliding door panel-1, sliding door panel-2, fixed panel, aluminum rail frame, head rail, bottom rail, l side jamb, r side jamb.			Glass sheet, glass sheet, glass sheet, aluminum strip, aluminum strip, aluminum strip, aluminum strip, aluminum strip.		Flat, flat, flat, var, var, var, var, var.	
Sliding door panel-1, sliding door panel-2, aluminum rail frame, head rail, bottom rail, l side jamb, r side jamb.			Mirror, mirror, aluminum strip, aluminum strip, aluminum strip, aluminum strip, aluminum strip.		Flat, flat, var, var, var, var, var.	

We faced a dilemma, however, with choosing the way to instantiate the level of cognitive engagement evoked by the quality evaluation method in our experiment. On the one hand, we would have liked to have operationalized the construct by having someone who played the role of an analyst engage in face-to-face and systematic questioning of our participants about the conceptual model they were evaluating. We suspected that this approach would have been the most forceful way to operationalize the construct in our experiment. On the other hand, we also had to constrain the amount of time participants had to expend to complete the experiment. Indeed, the time constraints we faced were

especially salient for us in light of (a) our using busy, practicing professionals as participants in our experiment, and (b) the pilot-test results we had obtained that indicated the experiment might take well over an hour to complete.

Eventually, for four reasons we opted to employ the use-case approach to instantiate our cognitive engagement factor. First, in the context of quality evaluation methods used in practice, we believe that use-cases lead to one of the highest, if not the highest, levels of cognitive engagement. Second, among UML practitioners, Dobing and Parsons (2006) found that use cases are employed extensively to verify user requirements. Third, use cases provided a way to constrain the time required to perform the experimental task. Fourth, during discussions we had with data-modeling practitioners, we found that they employ use-case data as a way to evaluate the quality of a conceptual model. They have adopted this approach because they have often been given limited opportunities to engage with stakeholders in face-to-face interactions (during which they could question stakeholders about the validity of a conceptual model's components).

Nonetheless, our choice of voluntary employment of use cases by our participants as a way to operationalize the cognitive engagement construct was somewhat speculative. Unfortunately, because the construct is novel, we have little prior research that provides guidance on the extent to which different quality evaluation methods are likely to lead stakeholders to engage with the evaluation task. Moreover, we had deliberately traded off what we suspected would be a more forceful operationalization of the construct against the need to constrain the time required to complete the experiment. As a result, we were unsure about how well our treatment would take effect.

The fourth set of materials comprised a summary sheet on which participants, at the conclusion of the experiment, were asked to record all defects they had identified in the conceptual model they were given.

3.3. Participants

Participants in the experiment were 48 expert data modelers in Indian organizations (28 participants) and Sri Lankan organizations (20 participants). We defined an expert as someone who had constructed at least five substantive data models (models that had been used as the basis for an information system that had been implemented) and who had worked in at least two industry sectors. We obtained the cooperation of these experts by either approaching the presidents of local chapters of the Data Management Association or contacting colleagues whom we knew. We asked them to approach expert data modelers whom they knew and to request these experts to contact us if they were willing to assist us in our research. When a data modeler contacted us to indicate that they were willing to assist us, we confirmed with them that they had the necessary experience before enlisting them as participants in our research.

3.4. Procedures

We collected data about the performance of participants in our experiment using a process-tracing technique (Ericsson & Simon, 1984). Specifically, we used the concurrent verbal protocol approach (which is often also called the talk-aloud or think-aloud protocol). Participants were asked to verbalize their thoughts while they performed the experimental task that they were given. By using concurrent verbal protocols, we could trace each step in participants' cognitive processes rather than having to (a) induce their cognitive processes based on task outcomes, or (b) query participants retrospectively about the cognitive processes that they thought they used while they performed the experimental task. We believed the protocols collected from our participants would provide us with rich data about the impact of the ontological clarity of a conceptual model and the quality evaluation approach used on performance outcomes. On average, the experiment took about 90 minutes to complete.

Prior to beginning the primary experiment, the materials and procedures to be used in the experiment were pilot tested with eight information systems professionals. Six were expert data modeling practitioners; two were academic colleagues who had expertise in, and had taught, data modeling. No concerns were identified, although some minor refinements were made to the experimental materials and experimental procedures. The pilot test also allowed the experimenter to develop facility in running the experiment.

In the primary study, participants were first assigned randomly to one of the four treatments in our experiment. Also, to ensure that the experiments were conducted consistently across both treatments, only one of us (henceforth, the experimenter) carried out all experiments.

Participants were run individually through the experiment. Several days before they undertook the experiment, the experimenter sent them an instruction sheet that explained the objectives and nature of the experiment, the tasks that they would perform, and the approximate time it would take them to perform the experimental tasks.

When participants arrived to undertake the experiment, the experimenter greeted them, reiterated the objectives and nature of the experiment, provided an overview of the task that they would undertake, explained their rights as participants in the experiment (e.g., they could withdraw from the experiment at any time without penalty), and asked them to complete the consent form (required under our university's research ethics protocols). The experimenter then explained that they would be given a narrative description of an application domain and a conceptual model that purportedly represented the semantics of this domain. Their primary task was to identify as many errors and omissions as possible in the conceptual model's representation of the domain's semantics. The experimenter then answered any questions that participants had about the experiment.

When participants indicated that they were ready to begin, the experimenter explained the nature of concurrent verbal-protocol procedures. He then indicated that they would be given a warm-up task to assist them to become comfortable with the procedures required. The experimenter subsequently answered any questions that participants had about concurrent verbal-protocol procedures and again underscored the need to talk aloud as they undertook the experimental tasks.

Next, participants were given the warm-up task. The experimenter prompted them to speak aloud if a period of silence ensued. If participants had questions about the task as they performed the experiment, the experimenter responded but took care not to lead them down a particular evaluation path. At the end of the warm-up task, the experimenter addressed any concerns and answered any questions raised by participants.

When participants indicated that they were ready to begin the primary task, they were given the materials for the primary task. All were given the same case description of the manufacturing company. Depending on the treatment to which they had been assigned, however, they were given (a) either the ontologically clear or ontologically unclear conceptual model, and (b) either the use-case table or no use-case table. As with the warm-up task, the experimenter prompted participants to speak aloud if a period of silence ensued and responded to questions asked by the participant as they undertook the primary task. When participants indicated that they had finished the primary task, they were asked to complete the answer sheet that asked them to summarize errors or omissions in the conceptual model they had examined.

After participants had completed the answer sheet, the experimenter debriefed them. In particular, he gave participants an opportunity to raise questions or concerns about the experiment and to provide their views on the research we were undertaking.

4. Results

In the subsections below, we present our results using two types of data analyses. The first, which we call "summative analyses", reports the results of statistical analyses we undertook on the effects of our treatments on the number of conceptual model defects identified by participants. The second, which we call "protocol analyses", reports the results of our analyses of the protocol data we collected.

4.1. Summative Analyses

Prior to undertaking statistical analyses of our summative data, we found that we needed to undertake some data-coding work. We first describe the nature of this work. We then present the results of our analyses.

4.1.1. Summative Data Coding

Our summative data coding involved two steps. First, we had typed transcripts of the recorded protocols prepared. When the typed transcripts were returned, we checked them carefully against the recordings to ensure that they were accurate and complete. Second, using the list of defects that we had seeded in the conceptual models (Table 2), each of us independently compiled a list of defects noted by eight randomly chosen participants equally distributed across the four treatments.

We found that we had to use three sources of evidence to identify the defects that each participant had noted. The first was the summary sheet of defects that participants prepared at the end of the experiment. The second was each participant's protocol. On a few occasions, their protocols contained comments about defects in the conceptual model that they were given, but they did not list these defects on the summary sheet. The third was the sheet of paper showing the conceptual model (either Figure 2 or Figure 3) that they were given. On a few occasions, they marked/described defects directly on the model itself, but they did not list these defects on the summary sheet. From these three sources, we compiled a list of defects that each of the eight participants had noted.

After completing coding for each of the eight participants, we then compared our lists and reconciled any differences that existed. We found only a few differences in our coding, which we could reconcile easily. As a result, only one of us coded the data for the remaining 40 participants.

4.1.2. Summative Results

Table 4 provides descriptive statistics for the proportion of total defects (errors and omissions) in a model identified by participants who received the model. Note that the table is organized according to the two dimensions that are our experimental-design factors – namely, ontological clarity and cognitive engagement of the quality evaluation method. Our focus is the proportion of total defects identified in the specific model that a participant received because the ontologically clear and ontologically unclear models have an unequal number of defects (11 versus 8 defects).

Table 4. Proportion–Mean and (Standard Deviation)–of Defects Identified by Experimental Treatment

Level of cognitive engagement of quality evaluation method \ Level of ontological clarity	Clear	Unclear	Total
	High	0.300 (0.117) N = 12	0.205 (0.143) N = 12
Low	0.315 (0.141) N = 12	0.205 (0.123) N = 12	0.260 (0.141) N = 24
Total	0.308 (0.127) N = 24	0.205 (0.131) N = 24	0.256 (0.138) N = 48

Table 5 shows a similar table for the total number of defects, but the count pertains only to the seven defects common to both models. In other words, the count shown in Table 5 omits defects numbered 1, 2, 6, 7, and 8 in Table 2 because they exist in only one of the two models.

Table 5. Total–Mean and (Standard Deviation)–of Common Defects Identified by Experimental Treatment

Level of cognitive engagement of quality evaluation method \ Level of ontological clarity	Clear	Unclear	Total
High	3.080 (1.240) N = 12	1.670 (1.155) N = 12	2.370 (1.377) N = 24
Low	3.170 (1.467) N = 12	1.670 (0.985) N = 12	2.420 (1.442) N = 24
Total	3.130 (1.329) N = 24	1.670 (1.049) N = 24	2.400 (1.395) N = 48

We used analysis of variance (ANOVA) to conduct two sets of analyses of the main effects and interaction effect of our two factors on the defects identified by our participants. For both analyses, we undertook tests to examine whether the assumptions underlying our ANOVA models were satisfied (e.g., we used Levene's test to evaluate the null hypothesis of equality of error variances of the dependent variable across treatments). While some moderate violations of normality in the dependent variable existed for some treatments, no serious violations existed. Moreover, ANOVA is robust to violations of normality when treatments sizes are equal.

The dependent variable in our first analysis was the proportion of the total number of defects that participants identified in the model they were given. We obtained the following results:

- The main effect for ontological clarity was significant: $F = 7.283$, $df = 1/44$, $p = .010$.
- The main effect for cognitive engagement of the quality evaluation method was not significant: $F = 0.039$, $df = 1/44$, $p = .844$.
- The interaction effect between ontological clarity and cognitive engagement of the quality evaluation method was not significant: $F = 0.039$, $df = 1/44$, $p = .844$.
- The adjusted R^2 for the model was .085.

The dependent variable in our second analysis was the total number of the defects (those that were common to both models) that participants identified in the model they were given. We obtained the following results:

- The main effect for ontological clarity was significant: $F = 17.035$, $df = 1/44$, $p < .001$.
- The main effect for cognitive engagement of the quality evaluation method was not significant: $F = 0.021$, $df = 1/44$, $p = .907$.
- The interaction effect between ontological clarity and cognitive engagement of the quality evaluation method was not significant: $F = 0.021$, $df = 1/44$, $p = .907$.
- The adjusted R^2 for the model was .230.

Because ontological clarity was a significant factor in our ANOVA analyses, we did follow-up chi-square tests on each defect common to both models. We were interested in whether the level of

ontological clarity (clear versus unclear) was associated with the frequency with which participants detected a particular defect in the model.

Table 6 shows the number of participants in the ontologically clear and unclear groups who identified each of the defects in the conceptual models that they were given. The chi-square tests were significant for four of the seven defects that were common to both models: defect 5 ($\chi^2 = 4.752$, $p = .029$), defect 9 ($\chi^2 = 7.378$, $p = .007$), defect 10 ($\chi^2 = 6.701$, $p = .010$), and defect 12 ($\chi^2 = 7.378$, $p = .007$). For these four defects, participants who received the ontologically clear model were more likely to detect the defect.

In short, our summative results provide support for Proposition 1 but not Propositions 2 and 3. Participants who received the ontologically clear conceptual model were better able to detect defects in the model. The cognitive engagement factor, however, had no statistically significant effect on a participant's ability to detect a defect in the model that they received. Moreover, no statistically significant interaction effect existed for ontological clarity and cognitive engagement.

Table 6. Number of Participants Who Detected Seeded Defect

Defect no.	Explanation	Ont. clear model N = 24	Ont. unclear model N = 24
1	Subassembly object class is not shown as a subclass of part object class.	4	–
2	Component object class is not shown as a subclass of part object class.	0	–
3	Cut-component object class is not shown as a subclass of component object class.	1	0
4	Component-with-surcharge object class is not shown as a subclass of cut-component object class.	3	2
5*	Fact that a subassembly comprises one or more other subassemblies and/or one or more parts is not shown.	20	13
6	Instances of built-product object class are shown as always having one or more subassemblies.	4	–
7	Instances of built-product object class are shown as always having one or more subassemblies and/or cut components (note, a built product could comprise components only, none of which are cut components).	–	0
8	Instances of subassembly object class are shown as always having one or more cut components (note, a subassembly could comprise components only, none of which are cut components).	3	–
9*	Missing attribute "base price" in built-product object class.	13	4
10*	Missing attribute "model no." in built-product object class.	8	1
11	Missing attribute "surcharge amount" (might have been shown in cut-component object class even though not all cut components have a surcharge).	20	18
12*	Additional attribute "dimension computation algorithm" in subassembly object class.	13	4

Note: * means statistically significant difference; – means defect not seeded in model.

4.2. Protocol Analyses

The protocol analyses that we undertook were motivated primarily by the results obtained with our summative analyses. In particular, as we studied the transcripts to identify defects in the conceptual models that each participant had identified, we noted a pattern. Specifically, some participants engaged more fully with the conceptual model that they were given as they compared it against the narrative that all participants received. In this light, even though our second factor (level of cognitive engagement evoked by the quality evaluation method) was not statistically significant in our analysis of the summative data, we concluded that the protocol transcripts manifested important differences in the ways that participants undertook the experimental task. We sought to explore these perceived differences further.

As with our summative analyses, we first undertook some data-coding work. We then analyzed plausible relationships among the data that we had coded.

4.2.1. Protocol Data Coding

Following Shanks, Moody, Nuredini, Tobin, and Weber (2010), we coded the cognitive behaviors manifested in our transcripts using four categories of episodes.

1. Understanding case description (type A episode): Used when a participant's utterances indicated that they were reading and seeking to understand the domain semantics evident in the case description.
2. Understanding conceptual model (type B episode): Used when a participant's utterances indicated that they were studying the conceptual model that they have been given in an effort to understand its underlying semantics.
3. Evaluating conceptual model against case description (type C episode): Used when a participant's utterances indicated that they were evaluating the conceptual model that they have been given to determine whether it correctly and completely represented the domain semantics evident in the case description.
4. Null category (type D episode): Used when a participant's utterances indicated that they were no longer "on task" (e.g., they digressed to make a tangential statement about how they believed data modeling in general should be done).

Based on our reading of a participant's complete protocol, we also classified the participant as having low cognitive engagement overall, medium cognitive engagement overall, or high cognitive engagement overall. In this regard, those participants who were distracted and frequently went "off task" were classified as having low cognitive engagement overall. Those who failed to discuss the meaning of the object classes encountered in either the narrative or the conceptual model were also classified as having low cognitive engagement overall.

To be classified as high cognitive engagement overall, participants in their transcripts had to clearly demonstrate all the following behaviors: (1) a systematic approach to the quality evaluation task, (2) a clear attempt to understand the semantics of the case, and (3) a clear attempt to understand the model's that they were given. Behaviors (2) and (3) were evident in their statements about the meaning of objects, attributes, associations, and multiplicities embedded in the narrative and conceptual model that they were given. The remaining participants were classified as having medium cognitive engagement overall. In other words, this group comprised participants who had not been classified as either high-engagement or low-engagement participants.

When coding the protocols, the three of us first worked independently to code five participants' protocols. We then met, compared our codes, and reconciled differences among us. We found that a high level of agreement existed in the codes that we had assigned to each participant's protocol and that any differences that existed could be reconciled easily. In light of our experience in coding five protocols, however, we also concluded it was best for us to code the protocols rather than employ independent coders. Interpreting the protocols often required a deep knowledge of the motivation for and nature of the experiment. We doubted that we could communicate this knowledge to coders

without compromising the independence we sought from them. Accordingly, we then worked in pairs to code the remaining 43 protocols. To increase the likelihood of coding consistency, one of us (the experimenter) was always a member of the coding pair. The other two of us worked with the experimenter to code a subset of the 43 protocols – one of us coded 22 protocols, and one of us coded the remaining 21 protocols.

4.2.2. Protocol Results

Table 7 provides descriptive statistics for the proportion of total defects (errors and omissions) in a conceptual model identified by participants who received the model. As in Table 4, Table 7 has two dimensions: ontological clarity (operationalized via the two levels of the experimental treatment); and cognitive engagement (operationalized via the three levels assigned during our protocol coding).

Table 7. Proportion – Mean and (Standard Deviation) – of Defects Identified (Organized by Ontological Clarity and Cognitive Engagement)

Level of cognitive engagement \ Level of ontological clarity	Clear	Unclear	Total
High	0.376 (0.088) N = 11	0.288 (0.101) N = 6	0.345 (0.100) N = 17
Medium	0.216 (0.163) N = 5	0.157 (0.155) N = 7	0.182 (0.154) N = 12
Low	0.270 (0.108) N = 8	0.190 (0.117) N = 11	0.224 (0.117) N = 19
Total	0.308 (0.127) N = 24	0.205 (0.131) N = 24	0.256 (0.138) N = 48

Table 8 shows a similar table for the total number of defects, but the count pertains only to the seven defects common to both conceptual models.

Table 8. Total – Mean and (Standard Deviation) – of Common Defects (Organized by Ontological Clarity and Cognitive Engagement)

Level of cognitive engagement \ Level of ontological clarity	Clear	Unclear	Total
High	3.820 (0.874) N = 11	2.330 (0.816) N = 6	3.290 (1.105) N = 17
Medium	2.200 (1.643) N = 5	1.290 (1.254) N = 7	1.670 (1.435) N = 12
Low	2.750 (1.282) N = 8	1.550 (0.934) N = 11	2.050 (1.224) N = 19
Total	3.130 (1.329) N = 24	1.670 (1.049) N = 24	2.400 (1.395) N = 48

In light of our focus on the level of cognitive engagement in our analysis of participants' protocols, we identified five plausible relationships that we sought to investigate further.

1. Number of defects identified and effort spent evaluating conceptual model against narrative: To the extent that participants spent more effort evaluating the conceptual model against the narrative, we predicted that they would identify more defects in the model that they were given. As a measure of this effort, we calculated the proportion of words in a participant's transcript coded under a Type C episode.
2. Number of defects identified and number of transitions between episodes where participant was engaged in understanding the narrative and episodes where participant was engaged in evaluating the conceptual model that they had been given against the narrative: To the extent that participants transitioned more frequently between these two types of episode, we predicted that they would identify more defects in the model that they were given. In this light, we counted the number of transitions between (a) a Type A episode and a Type C episode, and (b) a Type C episode and a Type A episode.
3. Number of defects identified and number of transitions between episodes where participant was engaged in understanding the conceptual model and episodes where participant was engaged in evaluating the conceptual model that they were given against the narrative: As with (2) above, to the extent that participants transitioned more frequently between these two types of episode, we predicted that they would identify more defects in the model that they were given. In this light, we counted the number of transitions between (a) a Type B episode and a Type C episode, and (b) a Type C episode and a Type B episode.
4. Number of defects identified and number of transitions between episodes where participant was engaged in understanding the narrative and episodes where participant was engaged in understanding the conceptual model that they had been given: As in (2) and (3) above, to the extent participants transitioned more frequently between these two types of episode, we predicted that they would identify more defects in the model that they were given. In this light, we counted the number of transitions between (a) a Type A episode and a Type B episode, and (b) a Type B episode and a Type A episode.
5. Number of defects identified and overall cognitive engagement level: We predicted that a participant's overall cognitive engagement level (low, medium, or high) would be associated with the number of defects that they identified in the conceptual model that they were given.

Initially, our view was that we would use analysis of covariance (ANCOVA) to conduct two analyses. Both would have four covariates: (a) proportion of words in a participant's transcript coded as Type C (evaluation of conceptual model against narrative), (b) number of transitions between episodes where participant was engaged in understanding the narrative and episodes where participant was engaged in evaluating the model that they were given against the narrative (ACCA), (c) number of transitions between episodes where participant was engaged in understanding the conceptual model and episodes where participant was engaged in evaluating the model that they were given against the narrative (BCCB), and (d) number of transitions between episodes where participant was engaged in understanding the narrative and episodes where participant was engaged in understanding the conceptual model that they were given (ABBA). Both analyses also had two main effects (ontological clarity and cognitive engagement) and an interaction effect (ontological clarity by cognitive engagement).

From our analyses of the summative data, we already knew that the main effect for ontological clarity was statistically significant. We sought to determine, however, whether the main effect for cognitive engagement (coding based on our examination of the protocols) and the interaction effect of

ontological clarity and cognitive engagement were statistically significant. Moreover, by using ANCOVA for the protocol data, we sought to control for any effects on the dependent variables associated with the four covariates before the main and interaction effects were evaluated.

We first correlated the four covariates with our two dependent variables – namely, the proportion of the total number of defects that participants identified in the model that they were given, and the total number of the defects (those that were common to both conceptual models) that participants identified in the model that they were given. Only the proportion of Type C code in a transcript and the count of ACCA transitions in a transcript were significant predictors of both dependent variables ($p < .05$). The count of BCCB transitions and the count of ABBA transitions were not significantly correlated with either dependent variable. Thus, there was no point to include these latter two variables in an ANCOVA model.

When we checked the assumptions underlying our ANCOVA models, we also found that the proportion of Type C code in a transcript was not independent of the treatments (which violates an assumption underlying ANCOVA). Thus, we could not include this covariate in our statistical analyses (in other words, the count of ACCA transitions in a transcript was the only covariate that we could use in our ANCOVA analyses). As with our summative analyses, we also found some moderate violations of normality in the dependent variable existed for some treatments, but no serious violations existed.

The dependent variable in our first analysis was the proportion of the total number of defects that participants identified in the model that they were given. We obtained the following results.

- Consistent with the summative analysis, the main effect for ontological clarity was significant (albeit slightly above the .05 level): $F = 4.001$, $df = 1/41$, $p = .052$.
- The main effect for cognitive engagement was significant: $F = 3.755$, $df = 2/41$, $p = .032$.
- The interaction effect between ontological clarity and cognitive engagement was not significant: F approx. zero, $df = 2/41$, p approx. one.
- The count of ACCA transitions engagement was not significant: $F = 1.654$, $df = 1/41$, $p = .206$.
- The adjusted R^2 for the model was .259.

When the covariate was removed from the analysis, the main effects for ontological clarity and cognitive engagement still remained significant ($F = 4.432$, $df = 1/42$, $p = .041$ and $F = 5.589$, $df = 2/42$, $p = .007$ respectively), and the interaction effect between ontological clarity and cognitive engagement still remained as not significant ($F = 0.052$, $df = 2/42$, $p = .950$). The adjusted R^2 for the model was .248.

For the cognitive engagement factor, a post hoc comparison using Tukey's HSD multiple comparison test shows high-engagement participants on average detected a greater proportion of defects than both low-engagement and medium-engagement participants. The difference between the low-engagement participants and medium-engagement participants, however, was not statistically significant.

The dependent variable in our second analysis was the total number of the defects (those that were common to both conceptual models) that participants identified in the model they were given. We obtained the following results.

- Consistent with the summative analysis, the main effect for ontological clarity was significant: $F = 12.272$, $df = 1/41$, $p = .001$.
- The main effect for cognitive engagement was significant: $F = 3.865$, $df = 2/41$, $p = .029$.

- The interaction effect between ontological clarity and cognitive engagement was not significant: $F = 0.100$, $df = 2/41$, $p = .905$.
- The count of ACCA transitions engagement was not significant: $F = 0.855$, $df = 1/41$, $p = .361$.
- The adjusted R^2 for the model was .369.

When the covariate was removed from the analysis, the main effects for ontological clarity and cognitive engagement remained significant ($F = 12.992$, $df = 1/42$, $p = .001$ and $F = 5.404$, $df = 2/42$, $p = .008$ respectively), and the interaction effect between ontological clarity and cognitive engagement was not significant ($F = 0.223$, $df = 2/42$, $p = .801$). The adjusted R^2 for the model was .371.

For the cognitive engagement factor, a post hoc comparison using Tukey's HSD Multiple Comparison Test shows high-engagement participants on average detected a greater number of defects than both low-engagement and medium-engagement participants. The difference between the low-engagement participants and medium-engagement participants, however, was not statistically significant.

In summary, our protocol results provide support for the notion that participants who had higher levels of cognitive engagement with the conceptual model quality evaluation task were better able to detect defects in the conceptual model that they were given. The level of cognitive engagement that they manifested, however, was not associated with the quality evaluation treatment that they received. Rather, it was associated with broader strategies that they used to approach the quality evaluation task – for instance, whether they systematically evaluated the conceptual model that they were given to determine whether the objects, attributes, associations, and multiplicities represented in the model were congruent with the case description.

5. Discussion

A criticism that might be levied at our research is that our results reflect that the ontologically clear model in Figure 2 is simpler than the ontologically unclear model in Figure 3. Specifically, the former has fewer object classes and fewer associations than the latter. Thus, cognitively it is easier to understand (Batra, 2007).

If Bunge's (1977) ontology is used to provide normative guidance during modeling activities, analysts will often prepare conceptual models with fewer object classes and associations. The reason is that object classes always represent classes (or subclasses) of things in ontologically clear models. In ontologically unclear models, however, object classes represent both classes (or subclasses) of things and attributes of things in a class (especially "repeating-group" attributes (Simsion & Witt, 2005)). Thus, ontologically unclear conceptual models will often have more object classes than ontologically clear models. As a consequence, they will also have more associations because object classes in the ontologically unclear model must be linked to ensure the association between a thing and its attributes can be determined via the model. It is important to note, therefore, that the existence of putatively simpler models under ontologically clear treatments is not an experimental confounding; rather, it is the essence of the treatment. Nonetheless, as Burton-Jones et al. (2009) point out, the effects of perceived simplicity in conceptual models still need to be evaluated empirically.

At times, an ontologically clear model will not have fewer object classes if some classes of things have additional attributes or associations that mean subclasses of these things must be represented. In an ontologically clear model, these subclasses must be represented explicitly via an additional object class construct. In an ontologically unclear model, however, a modeler might choose to represent them implicitly by attaching optional attributes/associations to the object class that represents their super-class. Nonetheless, even when an ontologically clear model has more objects classes and associations than an ontologically unclear model, we predict that the one-one mapping between ontological constructs and grammatical constructs still facilitates its quality evaluation.

To the extent that individuals evaluating an ontologically clear model realize (either consciously or subconsciously) that object classes only ever represent classes (or subclasses) of things, we predict that they will check to ensure that all attributes and associations that “attach” to things in the class are “clustered” around (or with) the object. In an ontologically unclear model, however, the attributes and associations that things in a class possess may be dispersed across multiple object classes. Thus, we suspect that individuals evaluating an ontologically unclear model will sometimes fail to detect missing attributes and associations in their models because the link between things and the attributes and associations that the things possess is not always straightforward.

Moreover, in an ontologically clear model, all attributes of a class are mandatory. As a result, we predict that participants are more likely to notice any attributes that are missing. In an ontologically unclear model, however, some classes have optional attributes. If an attribute is missing from one class, therefore, participants may simply assume incorrectly that it has been attached to another class as an optional attribute.

Our summative results for defects 5, 9, and 10 in our experiment support this prediction – specifically, participants who received the ontologically clear model were more able to detect that attributes were missing in the conceptual model that they were given. We also have evidence to support this prediction in some of our protocol results. For instance, Participant 31, who received the ontologically clear model, first compared all object classes in the conceptual model that he was given against the case description. He then checked to see that all attributes associated with an object class were present in the model. Initially he believed that “base price” should be an attribute of the object class “cut component”. As he tries to make sense of the domain semantics, he realized his mistake: “See the base price do not change at all. Yeah, so that means that we need to have a product object class (confirms his earlier understanding)... So I have a basic product where the base price is the same for all customers” (identifies error 9).

In a similar vein, individuals evaluating an ontologically clear model should be able to detect duplicated attributes or associations more easily. For instance, in the ontologically clear model in Figure 2, the attribute dimension computation algorithm is duplicated in the subassembly and cut-component object classes. Because the subassembly and cut-component object classes clearly represent classes of things, we predict that the existence of the duplicate attribute is easier to identify. In the ontologically unclear model in Figure 3, however, the duplicated attribute is located with the subassembly object class and a repeating-group attribute that is represented as an object class. This object class in turn is related to a cut-component object class. Cognitively, therefore, the duplicated attribute is more “distant” from its associated object class. Thus, we suspect that the fact it is a duplicate is more difficult to detect.

Our summative results for defect 12 in our experiment support this prediction – specifically, that participants who received the ontologically clear model were more able to detect this defect. We also have evidence to support this prediction in our protocol results. For instance, Participant 5, who received the ontologically clear model, commented: “So when I’m still at the subassembly and when I’m questioning, you know, what is, why do you need to represent this at the conceptual level. Are these two different? I mean, dimension computing algorithms are different at a subassembly level and at a cut component level? I am questioning the redundancy of it”. On the other hand, Participant 4, who received the ontologically unclear diagram, illustrates his confusion in relation to the dimension computation algorithm because he is dealing with object classes that represent repeating groups of attributes: “Subassembly dimension computation algorithm. So you have the part code description and the quantity coming and what algorithm? I guess dimension description algorithm is this one. Shouldn’t that have been in the cut component?”

Four of our defects were subclasses that were missing from the models. Two were common to both the ontologically clear and ontologically unclear model (defects 3 and 4). Two were missing only from the ontologically clear model (defects 1 and 2). Which model participants received did not matter – few detected the defects.

Because ontologically clear models have attributes and associations clustered around the things to which they attach, we predicted that participants who received the ontologically clear model would see the value of subclassing the part and component object classes. All participants' protocols revealed, however, that they paid little or no attention to classes and subclasses in the conceptual models they received. We suspect they simply concluded subclasses could be represented implicitly by designating some attributes and associations as optional. In this regard, our narrative had no phenomena where use of optional attributes or associations would have led to a substantial level of confusion about the domain semantics represented by the model. Moreover, because use of optional attributes and associations in practice is widespread, our participants might not have been aware of the sorts of problems pointed out by Bodart et al. (2001) and Gemino and Wand (2005) that can arise when they are used in conceptual models.

Three of our defects were cardinality errors. Two related to the ontologically clear model only (defects 6 and 8), and one related to the ontologically unclear model only (defect 7). As with the subclassing defects, it did not matter whether participants received the ontologically clear or ontologically unclear model. With few exceptions, all participants failed to detect that the cardinalities were erroneous.

We were surprised that our cognitive engagement treatment had no effect on the number of defects identified in a conceptual model. When we examined the protocols of participants who received the use-case treatment, they revealed that 14 of the 24 participants who received the use-case table had employed it in the ways that we had predicted. For instance, they mapped columns in the table to constructs in the conceptual model that they received, and they evaluated whether the model could accommodate the types of data that they found in a column of the table. As such, we undertook chi-square tests for each of the defects that were common to both models to determine whether the cognitive engagement factor had any effect on the likelihood that the defect would be detected. None of the tests were statistically significant.

When we measured cognitive engagement in another way, however, we found participants who manifested a high level of cognitive engagement were better able to detect defects in the conceptual model that they were given. Specifically, when participants' protocols revealed that they remained "on task" and used a systematic strategy to compare the congruence of object classes, attributes, associations, and cardinalities in the conceptual model that they received with those in the narrative, they were better able to detect defects in the model. The protocol measure of cognitive engagement reflects how participants actually behaved when they evaluated the model that they were given rather than how we predicted that they would behave under our experimental treatment.

6. Implications for Research and Practice

Our research has implications for both research and practice. In the following subsections, we briefly canvass these implications.

6.1. Implications for Research

The results we obtained for the impact of cognitive engagement on the effectiveness of conceptual model quality evaluation work were equivocal. On the one hand, our experimental treatment produced no effect. On the other hand, our protocol analyses suggested that participants who were more cognitively engaged with the conceptual model that they were given detected more defects in the model. Moreover, in follow-up work, we found that no relationship existed between the extent to which participants engaged cognitively with the model that they were given (as manifested in the protocols) and whether they received the use-case treatment in the experiment. In other words, factors other than the experimental treatment that they received motivated some participants to have greater levels of cognitive engagement with the model they were given. Future research might seek to determine why some individuals who undertake a quality assessment of a conceptual model engage more with the model that they have to evaluate. In particular, a construct such as "cognitive engagement of stakeholder with conceptual model quality evaluation task" might be used as a mediating construct between various characteristics of the stakeholder (e.g., domain expertise and conceptual modeling expertise), various characteristics of the conceptual modeling method (e.g., level of structure and level of task challenge), and so on.

In the design of our experiment, we decided not to force participants who received the use case to employ it when evaluating the conceptual model that they received. In this exploratory work, we were more interested to see how expert data modelers voluntarily went about evaluating a conceptual model. Given the cognitive-engagement factor did not show an effect in our experiment, future research might, therefore, instruct participants to use the quality assessment method that they have been given to determine whether the method has any effect on quality assessment outcomes.

An important issue to investigate is the extent to which cognitive engagement with a validation method will remain volitional among stakeholders or whether training them in structured validation methods will result in their having greater levels of cognitive engagement with a validation method. To the extent cognitive engagement with a validation method remains volitional, experimental treatments that manipulate validation methods will be ineffective. Instead, research that investigates the factors that lead to stakeholders having greater levels of engagement with a conceptual model is likely to be more fruitful. If greater levels of cognitive engagement can be evoked through non-volitional use of validation methods, however, the characteristics of those validation methods that lead to higher levels of engagement might then be studied.

Our research could be extended in still other ways – for instance:

- Different quality evaluation methods (and combinations of methods) could be evaluated.
- The performance of different stakeholders (e.g., end users) and teams of stakeholders in the quality evaluation process could be investigated (e.g., their ability to identify different kinds of defects in conceptual models).
- Different ontologies could be used as the basis for preparing ontologically clear and unclear conceptual models to test their impact on the effectiveness of conceptual model quality evaluation work.
- Conceptual models of varying levels of complexity could be investigated to determine whether the interaction effect that we postulate in our theory takes effect at some level of complexity.
- The impact of ontologically clear and ontologically unclear conceptual models on stakeholders' ability to detect different kinds of defects from those used in our experiment could be investigated.
- Defects seeded into conceptual models could be weighted according to the severity of their consequences. In our experiment, we have assigned equal weighting to each defect, but some are likely to be more serious than others.

6.2. Implications for Practice

Table 6 shows participants' performance at detecting defects in the conceptual model that they were given was mixed. In terms of the seven defects common to both models, participants who received the ontologically clear model performed moderately well with respect to five of the defects, whereas participants who received the ontologically unclear model performed moderately well with respect to only two of the defects. For the defects that were unique to the models they received, however, neither group performed well.

These results mirror those obtained in early work that investigated the performance of expert programmers in identifying defects in programs (e.g., Myers, 1978). Substantial variability in performance among the programmers was evident. Furthermore, overall performance among the programmers was "rather dismal" (Myers, 1978, p. 763). For practice, therefore, our results suggest a need for organizations to be circumspect about how effective conceptual modeling quality assurance

work is likely to be and to find better ways to undertake conceptual modeling quality assurance work. In this regard, the construction of ontologically clear conceptual models appears to hold promise.

7. Conclusion

In summary, both our summative results and our protocol results support our prediction that ontologically clear conceptual models will be easier to evaluate for quality than ontologically unclear models. In this regard, even though the merits of ontological clarity have been shown in other contexts, our results are the first to show, to the best of our knowledge, the benefits of ontological clarity when the quality of conceptual models must be evaluated.

Our summative results do not support our prediction, however, that quality evaluation methods that lead their users to cognitively engage more with the domain semantics represented by a conceptual model will be more effective in detecting defects in the model. Nonetheless, our protocol results provide some support for our arguments that greater cognitive engagement with the semantics of a conceptual model will lead to higher levels of quality evaluation effectiveness.

As such, future research might investigate what factors lead stakeholders to greater levels of cognitive engagement with conceptual model evaluation tasks and whether greater levels of engagement leverage the benefits of ontological clarity. Future research might also examine whether quality evaluation methods can be structured to facilitate stakeholders cognitively engaging with the semantics of conceptual models and, if so, whether appropriate incentives can be created for them to use the methods.

Acknowledgements

An earlier version of this paper was presented at the 19th Annual Workshop on Information Technologies and Systems, Phoenix, Arizona, USA, 14-15 December 2009. We are indebted to participants at the Workshop for their helpful comments. We are indebted, also, to three reviewers and the senior editor for helpful comments throughout the review process. The research was supported by funds from Australian Research Council Discovery Grant DP0773563.

References

- Bakker, A. B., Schaufeli, W. B., Leiter, M. P., & Taris, T. W. (2008). Work engagement: An emerging concept in occupational health psychology. *Work & Stress, 22*, 187-200.
- Batra, D. (2007). Cognitive complexity in data modeling: Causes and recommendations. *Requirements Engineering, 12*, 231-234.
- Bodart, F, Sim, M., Patel, A., & Weber, R. (2001). Should optional properties be used in conceptual modelling? A theory and three empirical tests. *Information Systems Research, 12*(4), 384-405.
- Bowen, P. L., O'Farrell, R. A., & Rohde, F. H. (2006). Analysis of competing data structures: Does ontological clarity produce better end user query performance? *Journal of the Association for Information Systems, 7*(8), 514-544.
- Bunge, M. (1977). *Treatise on basic philosophy: Volume 3: Ontology I: The Furniture of the World*. Dordrecht, Holland: Reidel.
- 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., Wand, Y., & Weber, R. (2009). Guidelines for empirical evaluations of conceptual modeling grammars. *Journal of the Association for Information Systems, 10*(6), 495-532.
- Connolly, T. M., & Begg, C. E. (2010). *Database systems: A practical approach to design, implementation, and management* (5th ed.). Boston: Pearson Addison-Wesley.
- Corno, L., & Mandinach, E.B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist, 18*(2), 88-108.
- Corral, K., Schuff, D., & St. Louis, R. D. (2006). The impact of alternative diagrams on the accuracy of recall: A comparison of star-schema diagrams and entity-relationship diagrams. *Decision Support Systems, 42*, 450-468.
- Davies, I., Green, P., Rosemann, M, Indulska, M., & Gallo, S. (2006). How do practitioners use conceptual modeling in practice. *Data & Knowledge Engineering, 58*, 358-380.
- Dobing, B., & Parsons, J. (2006). How UML is used. *Communications of the ACM, 49*(5), 109-113.
- Dunn, C. L, Gerard, G. J., & Grabski, S. V. (2005). Critical evaluation of conceptual data models. *International Journal of Accounting Information Systems, 6*(2), 83-106.
- Elmasri, R., & Navathe, S. B. (2011). *Fundamentals of database systems* (6th ed.). Boston: Pearson Addison Wesley.
- Erickson, K. A., & Simon, H. A. (1984). *Protocol analysis*. Cambridge, MA: MIT Press.
- Evermann, J., & Wand, Y. (2005). Ontology based object-oriented domain modelling: Fundamental concepts. *Requirements Engineering, 10*, 146-160.
- Gemino, A. (1999). *Empirical methods for comparing system analysis modelling techniques* (Unpublished PhD thesis). Sauer School of Business Administration, University of British Columbia, Canada.
- Gemino, A., & Wand Y. (2005). Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. *Data & Knowledge Engineering, 55*(3), 301-326.
- Genero, M., Poels, G., & Piattini, M. (2008). Defining and validating metrics for assessing the understandability of entity-relationships diagrams. *Data & Knowledge Engineering, 64*(3), 534-556.
- Green, P., & Rosemann, M. (2000). Integrated process modelling: An ontological evaluation. *Information Systems, 25*(2), 73-87.
- Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly, 30*(3), 611-642.
- International Standards Organisation. (2005). ISO Standard 9000: 2005: Quality Management Systems: Fundamentals and Vocabulary. *International Standards Organisation*.
- Krogstie, J., Lindland, O. I., & Sindre, O. I. (1995). *Towards a deeper understanding of quality in requirements engineering*. Paper presented at the Proceedings of the 7th International Conference on Advanced Information Systems Engineering (CAISE), Jyvaskyla, Finland.
- Lindland, O. I., Sindre, G., Sølvsberg, A. (1994). Understanding quality in conceptual modeling. *IEEE Software, 11*(2), 42-49.
- Maes, A., & Poels, G. (2007). Evaluating quality of conceptual modeling scripts based on user perceptions. *Data & Knowledge Engineering, 63*, 701-724.
- Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153-184.

- Milton, S. K., & Kazmierczak, E. (2004). An ontology of data modelling languages: A study using a common-sense realistic ontology. *Journal of Database Management*, 15(2), 19-38.
- Moody, D. L. (2005). Theoretical and practical issues in evaluating the quality of conceptual models: Current state and future directions. *Data & Knowledge Engineering*, 55, 243-276.
- Moody, D. L., & Shanks, G. (2003). Improving the quality of data models: Empirical validation of a quality management framework. *Information Systems*, 28, 619-650.
- Myers, G. J. (1978). A controlled experiment in program testing and code walkthroughs/ inspections. *Communications of the ACM*, 21(9), 760-768.
- Opdahl, A., & Henderson-Sellers, B. (2002). Ontological evaluation of the UML using the Bunge-Wand-Weber model. *Software and Systems Modeling*, 1, 43-67.
- Radosevich, D. J., Radosevich, D. M., Riddle, M. R., & Hughes, P. A. (2008). Goal orientation as a predictor of cognitive engagement, performance, and satisfaction. *Journal of Academy of Business and Economics*, 8(3), 46-55.
- Recker, J., Rosemann, M., Indulska, M., & Green, P. (2009). Business process modeling – a comparative analysis. *Journal of the Association for Information Systems*, 10(4), 333-363.
- Recker, J., Rosemann, M., Green, P., & Indulska, M. (2011). Do ontological deficiencies in modeling grammars matter? *MIS Quarterly*, 35(1), 57-79.
- Rumbaugh, J., Jacobson, I., & Booch, G. (2005). *The unified modeling language reference manual* (2nd ed.). Boston : Addison-Wesley.
- Shanks, G., Tansley, E., & Weber, R. (2003). Using ontology to help validate conceptual models. *Communications of the ACM*, 46(10), 85-89.
- Shanks, G., Tansley, E., Nuredini, J., Tobin, D., & Weber, R. (2008). Representing part-whole relations in conceptual modelling: An empirical evaluation. *MIS Quarterly*, 32(3), 553-573.
- Shanks, G., Moody, D., Nuredini, J., Tobin, D., & Weber, R. (2010). Representing classes of things and properties in general in conceptual modelling: An empirical evaluation. *Journal of Database Management*, 21(2), 1-25.
- Siau, K., & Rossi, M. (2007). Evaluation techniques for systems analysis and design modelling methods – a review and comparative analysis. *Information Systems Journal*, 21, 249-268.
- Simsion G. C., & Witt G. C. (2005). *Data modelling essentials: Analysis, design and innovation* (2nd ed.). Arizona: Coriolis.
- 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.
- Wand, Y., & Weber, R. (1990). An ontological model of an information system. *IEEE Transactions on Software Engineering*, 16(11), 1282-1292.
- Wand, Y., & Weber, R. (1993). On The ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal*, 3(4), 217-237.
- Wand, Y., Storey, V., & Weber, R. (1999). An ontological analysis of the relationship construct in conceptual modeling. *ACM Transactions on Database Systems*, 24, 494-528.
- Weber, R. (1996). Are attributes entities? A study of database designers' memory structures. *Information Systems Research*, 7(2), 137-162.
- Webster, J., & Hackley, P. (1997). Teaching effectiveness in technology-mediated distance learning," *Academy of Management Journal*, 40(6), 1282-1309.

Appendices

Appendix A. Case Materials Provided to Participants (Case Description)

Triple-A (Acme Aluminum Accessories)

Company Overview

Triple-A (Acme Aluminum Accessories) is a company based in Victoria, Australia. It provides aluminum-based products for residential and commercial buildings. These products include shower screens, security doors, wardrobe doors, leafless guttering, fly screens, and security shutters.

All Triple-A's products are made-to-order. The primary reason is that Triple-A's products are often used in situations that require tailored solutions. For example, a shower screen might be fitted in a place that is 5mm or 10mm wider at the top than at the base – walls that appear straight to the owner's eye often have variations that are significant when it comes to a fit-out.

Purpose of the Conceptual Data Model

A conceptual data model is required to support development of a system for recording the measurements of components of all products actually built (not model types) at AAA to particular customer requirements (measurements and finishes) for factory floor cutting. Please note that conceptual model should not consider any primary or foreign keys or any other logical model aspects.

Product Overview

Triple-A has over 40 models of products. Each model has its own rules on measurements, allowable materials, and component size calculations. As an example, one model of shower screen has a three-panel front (one fixed panel, two sliding panels). The required on-site measurements for this model are height (H), width at top (WT), and width at bottom (WB) of the opening. Another shower screen model has a return over the bath hob. In addition to the height, width-at-top and width-at-bottom measurements described above, the height of the bath and width of the bath must also be measured on site.

Each product will typically comprise several constituent parts. Some of these parts may themselves be composite parts; that is, they are made up of smaller parts. Such composite parts are known as subassemblies. For example, Alex's model "123" shower screen may be made up of a frame, one fixed panel, and two sliding panels. Each panel is a subassembly; that is, while it is a part in the overall product, it is also a composite part that contains smaller parts such as lengths of aluminum strip and a sheet of glass.

The decomposition may be nested several levels deep. For example, a more complex product may be composed of subassemblies that are composed of subassemblies, and so on. However, at the end of the decomposition process, there are parts that, for all practical purposes, cannot be decomposed further. These atomic parts are known as components.

Some components do not need any special processing on the factory floor. For example, some handles are not cut to size, but are simply fitted. And most pieces of aluminum strip and glass sheet are cut to measure on the factory floor.

Each component of a subassembly requiring cutting has several cutting-floor dimensions. For example, the sheet of glass for the door of a particular customer's shower screen may have to be cut slightly off-square to reflect the fact that the walls of the customer's building are not perfectly vertical. The resultant component dimensions might be calculated as:

- Length of glass for the door = 1,706 mm
- Width of glass for the door at the base = 894 mm
- Width of glass for the door at the top = 898 mm.

Similarly, the aluminum strip components will require their dimensions to be calculated.

Each component is made of one material that may be offered in a number of finishes. For example, glass is offered in finishes such as Standard Clear, Safety Clear, and Standard Amber. Similarly, aluminum strip is offered in finishes such as Standard Matt, Standard Brushed, and Bronze Anodized.

Customers incur surcharges (extra costs) for some types of finishes of the materials used to make product components. Depending on the material used, these surcharges are either fixed or variable. For example, \$10 may be the fixed surcharge added to a product's base price if bronze anodized aluminum is the material used to make its aluminum components. Similarly, \$20 per square metre may be the variable surcharge added to a product's base price if amber glass is nominated.

Diagrammatic Product Sample

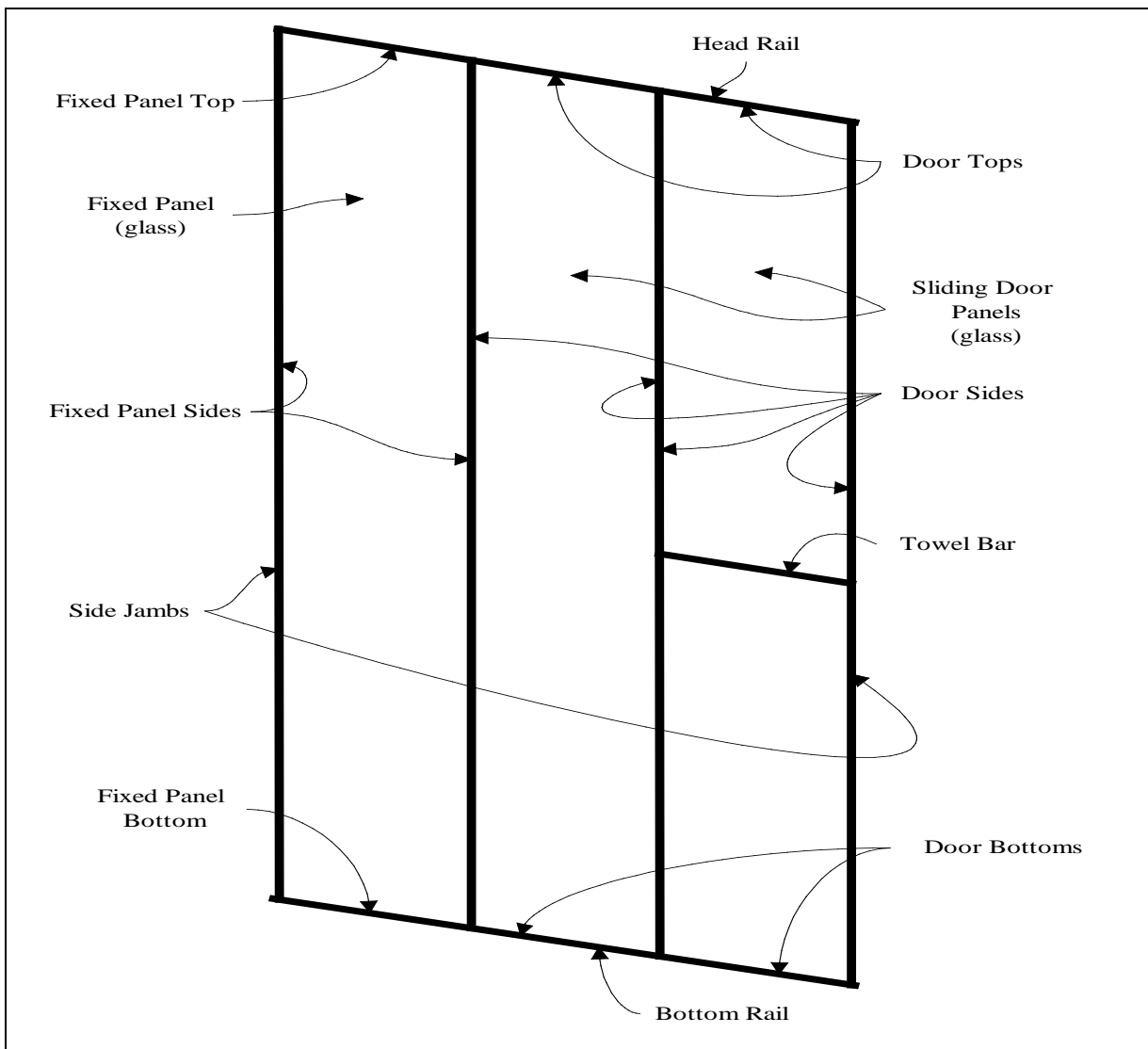


Figure A-1. A Pictorial View of a Simple Shower Screen and its Components

Dimension Computation

For the model portrayed above, assume that three on-site measurements are required – namely, width at the base (“WB”), width at the top (“WT”), and height (“H”). For a particular customer, assume also that the measurements are 860mm for WB, 865mm for WT, and 1850mm for H.

The aluminum strip length calculations are described in Table A-1 below. Formula variables represent either the measurement variables or the abbreviations of the components themselves. For example, in the formula “BR(L) = WB – 7”, BR is the variable for the base rail component (with a dimension of “length”), and WB is the variable for the width-at-base measurement. For the first two entries in the table below, the computation rules could be expressed in English as follows:

- Cut the head rail to equal the measured width of the top of the shower screen.
- Cut the bottom rail to 7 mm less than the measured width of the base of the shower.

Table A-1. Aluminum Strip Calculations

Component name	Abbrev.	Dimension	Formula	Result	Quantity
Head Rail	HR	L (Length)	HRL = WT	865	1
Bottom Rail	BR	L (Length)	BRL = WB – 7	853	1
Side Jamb	SJ	L (Length)	SJL = H – 35	1815	2
Door Top & Bottom	DT&B	L (Length)	DT&BL = (WT + 30) / 3	298	4
Door Side	DS	L (Length)	DSL = H – 60	1790	4
Fixed Panel Top & Bottom	FPT&B	L (Length)	FPT&BL = DT&BL	298	2
Fixed Panel Side	FPS	L (Length)	FPSL = H	1850	2
Towel Bar	TB	L (Length)	TBL = DT&BL – 60	238	1

The glass panel length and width calculations are described in Table A-2 below. The dimensions are again expressed in millimeters.

Table A-2. Glass Panel Height & Width Calculation

Component name	Abbrev.	Dimension	Formula	Result	Quantity
Sliding Door Panel	SDP	H (Height)	SDPH = DSL – 20	1770	2
		W (Width)	SDPW = DT&BL - 20	278	
Fixed Panel	FP	H (Height)	FPH = H – 20	1830	1
		W (Width)	FPW = DT&BL - 20	278	

Appendix B. Defects Seeded in the Conceptual Models with Explanations How These Violate the Case Materials

Table B-1. Defects Seeded into Conceptual Models with Explanation as to How Participants Could Find the Defect in the Case Materials

Defect no.	Description	Type of defect	Explanation of the detection mechanism, expected to be used by participants, based on quotations from case description (see above)
1	Subassembly object class is not shown as a subclass of part object class.	Omission Ont. clear model	In the case description, we find: "Some of these parts may themselves be composite parts; that is, they are made up of smaller parts. Such composite parts are known as subassemblies". Thus, the reader can conclude that some parts are in fact subassemblies.
2	Component object class is not shown as a subclass of part object class.	Omission Ont. clear model	In the case description, we find: "The decomposition may be nested several levels deep. For example, a more complex product may be composed of subassemblies that are composed of subassemblies, and so on. However, at the end of the decomposition process, there are parts that, for all practical purposes, cannot be decomposed further. These atomic parts are known as components". Thus, the reader can conclude that some parts are in fact components.
3	Cut-component object class is not shown as a subclass of Component object class.	Omission Both models	In the case description, we find: "Some components do not need any special processing on the factory floor. For example, some handles are not cut to size, but are simply fitted. And most pieces of aluminum strip and glass sheet are cut to measure on the factory floor." Thus, the reader can conclude that some components require cutting and are therefore may be a cut-components (or similar) class.
4	Component-with-surcharge object class is not shown as a subclass of Cut-component object class.	Omission Both models	In the case description, we find: "Each component is made of one material that may be offered in a number of finishes... Customers incur surcharges (extra costs) for some types of finishes of the materials used to make product components". Thus, the reader can conclude that some components attract this surcharge (but not all) – hence, a subclass would be warranted.
5	Fact that a subassembly is composed of one or more other subassemblies and/or one or more parts is not shown.	Omission Both models	In the case description, we find: "Each product will typically be composed of several constituent parts. Some of these parts may themselves be composite part; that is, they are made up of smaller parts. Such composite parts are known as subassemblies. For example, Alex's model "123" shower screen may be made up of a frame, one fixed panel, and two sliding panels. Each panel is a subassembly; that is, while it is a part within the overall product, it is also a composite part that contains smaller parts such as lengths of aluminum strip and a sheet of glass. The decomposition may be nested several levels deep. For example, a more complex product may be composed of subassemblies that are composed of subassemblies, and so on. However, at the end of the decomposition process, there are parts that, for all practical purposes, cannot be decomposed further. These atomic parts are known as components". Thus, the reader can conclude that each part can be composed of subassemblies and/or one or more parts and that the nesting can be deep.

Table B-1. Defects Seeded into Conceptual Models with Explanation as to How Participants Could Find the Defect in the Case Materials (cont.)

6	Instances of built-product object class are shown as always having one or more subassemblies.	Error Ont. clear model	In the case description we find: "Each product will typically be composed of several constituent parts. Some of these parts may themselves be composite parts; that is, they are made up of smaller parts. Such composite parts are known as subassemblies." The presence of subassemblies is therefore not required because of the use of 'may' in the text.
7	Instances of built-product object class are shown as always having one or more subassemblies and/or cut components (note, a built product could be composed of components only, none of which are cut components).	Error Ont. unclear model	In the case description, we find: "Each product will typically be composed of several constituent parts. Some of these parts may themselves be composite parts; that is, they are made up of smaller parts. Such composite parts are known as subassemblies". These statements are followed a little later in the description by "The decomposition may be nested several levels deep. For example, a more complex product may be composed of subassemblies that are composed of subassemblies, and so on. However, at the end of the decomposition process, there are parts that, for all practical purposes, cannot be decomposed further. These atomic parts are known as components. Some components do not need any special processing on the factory floor". Thus, the reader can conclude that products may be comprised of only components – and that this could be composed of components that are not cut components.
8	Instances of subassembly object class are shown as always having one or more cut components (note, a subassembly could be composed of components only, none of which are cut components).	Error Ont. clear model	Using the same passage as above (error 7), the reader can conclude that it is possible to compose a product that does not contain any cut components.
9	Missing attribute "base price" in built-product object class.	Omission both models	Discussion of the price clearly shows that surcharges are added to the base price: "Similarly, \$20 per square metre may be the variable surcharge added to a product's base price if amber glass is nominated."
10	Missing attribute "model no." in built-product object class.	Omission both models	Implied in the description of subassemblies for a product: "For example, Alex's model "123" shower screen may be made up of...". This discussion would imply that we can identify each product.
11	Missing attribute "surcharge amount" (might have been shown in cut-component object class even though not all cut components have a surcharge).	Omission both models	Surcharges are described clearly in the materials where both base price and surcharge price are discussed: "For example, \$10 may be the fixed surcharge added to a product's base price if bronze anodized aluminum is the material used to make its aluminum components". This discussion makes it clear that knowing the base price and the surcharge is required.
12	Additional attribute "dimension computation algorithm" in subassembly object class.	Error both models	Each assembly has its dimensions calculated based on the measurement for the product. The discussion of dimension computation is set at this level: "For the model portrayed above, assume that three on-site measurements are required...". Later in the case, precise calculation tables are given based on these measurements for cutting each of the cut components (both aluminum strips and glass). This implies that the product dictates the dimension computations not the subassembly.

About the Authors

Simon MILTON received his PhD from the University of Tasmania's Department of Information Systems. His dissertation reported the first comprehensive analysis of data modeling languages using a common-sense realist ontology. Dr Milton continues his interests in the ontological foundations of data modeling languages and the implications of top-level ontological commitments in information systems modeling. His recent work extends to the value and use of ontologies to business and biomedicine. He holds a senior lectureship in the Department of Computing and Information Systems at The University of Melbourne, Australia.

Jayantha RAJAPAKSE received the B.Sc. degree from the University of Peradeniya, Sri Lanka in 1981, Research Masters degree in Computer Science from the University of Queensland, Australia in 1997, and the Ph.D. degree in Information Systems from the University of Melbourne, Australia in 2006. Currently he is a Senior Lecturer at the Faculty of Information Technology, Monash University. He has been involved with teaching and research in information systems and computer science for over 15 years at the University of Queensland, the University of Melbourne, Queensland University of Technology, Manchester Metropolitan University (Sri Lanka), and Monash University.

Ron WEBER is Dean, Faculty of Information Technology, Monash University. He obtained his PhD from the University of Minnesota. His research interests focus primarily on conceptual modeling. Ron is a past president of the Association for Information Systems, a past co-chair of the International Conference on Information Systems, and a past editor-in-chief of the *MIS Quarterly*. He is a Fellow of the Association for Information Systems, the Australian Computer Society, the Institute of Chartered Accountants in Australia, CPA Australia, and the Academy of the Social Sciences of Australia.