
© Copyright 2010 Elsevier BV
The Measurement of Perceived Ontological Deficiencies of Conceptual Modeling Grammars

Jan Recker and Michael Rosemann
Information Systems Discipline
Queensland University of Technology
Brisbane, Australia
Email: j.recker@qut.edu.au; m.rosemann@qut.edu.au

Abstract

Over recent years, many scholars have studied the conceptual modeling of information systems based on a theory of ontological expressiveness. This theory offers four constructs that inform properties of modeling grammars in the form of ontological deficiencies, and their implications for development and use of conceptual modeling in IS practice. In this paper we report on the development of a valid and reliable instrument for measuring the perceptions that individuals have of the ontological deficiencies of conceptual modeling grammars. We describe a multi-stage approach for instrument development that incorporates feedback from expert and user panels. We also report on a field test of the instrument with 590 modeling practitioners. We further study how different levels of modeling experience influence user perceptions of ontological deficiencies of modeling grammars. We provide implications for practice and future research.

Keywords: Theory of ontological expressiveness, instrument development, perception measurement, conceptual modeling

1 Corresponding author: Jan Recker, Queensland University of Technology, 126 Margaret Street, Brisbane QLD 4000, Australia, Tel: +61 (07) 3138 9479, Fax: +61 (07) 3138 9390, Email: j.recker@qut.edu.au
The Measurement of Perceived Ontological Deficiencies of Conceptual Modeling Grammars

1. Introduction

In IS practice, analysts use conceptual models to represent features of a business domain intended to be supported by an existing or newly built information system [1]. These conceptual models (scripts) are specified using a conceptual modeling grammar, which consists of a set of constructs with a graphical representation, and rules to combine those constructs. The entity-relationship grammar, for example, includes the constructs ‘entity’ (represented as a rectangle) and ‘relationship’ (represented as a diamond). A rule of that grammar specifies that two entities can be associated only via a relationship. Conceptual modeling is conducted using a method (i.e., procedures by which the grammar can be used), and applies within an organizational context (i.e., the setting in which the modeling occurs) [1].

The practice of conceptual modeling has been found to be a very conducive way of articulating knowledge and perceptions about features of a real world domain that are, or should be, supported by an Information System (IS) [2]. In line with its popularity in practice, conceptual modeling has also emerged as a popular and relevant area of research in IS [1]. However, by far most of the published research in this domain is conceptual in nature, with the share of empirical papers in this space reportedly being less than 20 per cent [3].

One reason for the lack of empirical research in this area can be seen in the lack of mature theoretical foundations [4], upon which empirical inquiries could be based. To
address this lack of theory, Wand and Weber [5-7] developed a theory of ontological expressiveness, to guide researchers working in the space of conceptual modeling. Over recent years, this theory has become a popular framework for the study of conceptual modeling and related phenomena; both on a conceptual and empirical level [8]. Wand and Weber’s theory of ontological expressiveness, however, is not without its criticisms. Most notably, academics have criticized a lack of empirical validation of the principles that stem from the use of the theory [e.g., 9].

One of the related problems is the lack of validated measurement instruments that could be used to tap into the fundamental constructs of Wand and Weber’s theory of ontological expressiveness, and that could be used to study empirically conceptual modeling practice. The aim of this paper, therefore, is to report on research we undertook to develop a measurement instrument designed to capture the perceptions that an individual may have of the ontological deficiencies of conceptual modeling grammars.

User perceptions of ontological deficiencies of modeling grammars can be viewed as secondary attributes of the grammar in question [e.g., 10]. A primary attribute would be an actual ontological deficiency of the grammar. These deficiencies may be examined using the method of ontological analysis [6], and a large body of literature exists on such work (for an overview see, for instance, [11]). Secondary attributes of process modeling grammars (i.e., user perceptions of ontological deficiencies), however, are also an important aspect of study, because individuals do not necessarily perceive ontological deficiencies to be existent in the modeling grammars they work with. These perceptions, consequently, impact the way how they use these grammars in modeling practice. Consider the following example from a study on the usage of a process-oriented conceptual modeling grammar [12]. The grammar in use has
deficiencies in assisting users in the decomposition of process-oriented information systems [13]. Some of the grammar users indeed noted this deficiency:

“I think if the standard allows for a large amount of decomposition, my understanding is that it doesn’t at the moment, [...], we definitely need something to link.” (interview transcription data from [12])

Other users in the same organizations working with the same grammar, however, did not perceive this deficiency to be existent:

“I wouldn’t really say it [the lack of support for decomposition] is a problem right now because people are happy they can read it, they can understand it.” (interview transcription data from [12])

To be able to examine how actual ontological deficiencies (a primary attribute) in conceptual modeling grammars are perceived by the users working with these grammars (a secondary attribute), our first research objective is to develop a multi-item instrument to measure the user perceptions of the ontological deficiencies of conceptual modeling grammars. We then show how this instrument can be applied in a study of the use of conceptual modeling grammars in practice. Thus, our second research objective is to examine empirically how levels of modeling experience influence user perceptions of the ontological deficiencies of a conceptual modeling grammar. Accordingly, our two research questions are:

1. How do users perceive ontological deficiencies of a conceptual modeling grammar?

2. How do user perceptions of ontological deficiencies of a conceptual modeling grammar change when modeling experience is acquired?
We proceed as follows. Section 2 introduces, firstly, the conceptual modeling domain; secondly, describes Wand and Weber’s theory of ontological expressiveness; and thirdly, reviews related empirical work based on this theory. We report in Section 3 on the procedure we employed for measurement instrument development, and report on the results from a confirmatory field study with modeling practitioners. In Section 4 we present an application of the measurement instrument in a field study of conceptual modelers with different levels of modeling experience. We conclude in Section 5 with a discussion of the conclusions from our research, and provide suggestions for future research.

2. Background and Theory

2.1. Conceptual Modeling

Conceptual modeling is an essential cornerstone of many IS analysis and design methodologies. This is because conceptual modeling helps articulating knowledge about relevant business domain features and identifying errors at early stages of IS development [14]. Not surprisingly, a variety of research streams have addressed conceptual modeling. Halper et al. [15], for instance, suggest extensions to data modeling grammars to express ownership semantics. Decker et al. [16] specify extended semantics for a conceptual modeling grammars to specify service interactions. Similarly, Zhang et al. [17] report on the development of a new conceptual modeling grammar to articulate agile integration among participating work systems.

Undoubtedly, conceptual modeling is an important information systems practice; however, it is one that is yet to be fully understood. While some analytical work has been carried out to examine, and improve, the quality of conceptual modeling in
theory [18, 19], empirical studies of conceptual modeling practice [e.g., 20, 21, 22] still report on faulty or incomplete models, varying modeling performance, lacking utilization and other mishaps. Other authors have speculated a lack of understandability when using modeling grammars in isolation [23] or combination [24]. One reason for this situation is suggested to stem from the lack of widely acknowledged quality frameworks for conceptual modeling [3]. In particular, it is argued that little knowledge exists about the grammars used for conceptual modeling, and the characteristics that differentiate “good” from “bad” modeling grammars [e.g., 11].

2.2. Ontological Expressiveness of Modeling Grammars

Searching for a theoretical basis to aid IS researchers in establishing insights into the characteristics of conceptual modeling grammars, we turn to a theory of ontological expressiveness [5-7] to facilitate an understanding of the properties of conceptual modeling grammars. The theory of ontological expressiveness was developed from the adaptation of an ontology proposed by Bunge [25]. Generally, ontology studies the nature of the world and attempts to organize and describe what exists in reality, in terms of the properties of, the structure of, and the interactions between real-world things [25]. With information systems essentially being human-created representations of real-world systems, Wand and Weber [5-7] suggest that ontology may help in devising conceptual structures on which modelers can base their representations of these systems.

From Bunge’s ontological theory [25], Wand and Weber adapted an ontological model of representation, which specifies a set of rigorously defined ontological constructs to describe all types of real-world phenomena that a modeling grammar
user may desire to have represented in a conceptual model of an information systems domain. More details about Wand and Weber’s ontological model of representation is available elsewhere [5-7].

Based on this model, Wand and Weber formulated a theory of ontological expressiveness. Their theory purports to account for variations in the ability of conceptual modelers to use conceptual modeling grammars to develop diagrams of real-world phenomena that are ontologically complete and clear [26].

Wand and Weber’s theory of ontological expressiveness is founded on the nature of the mapping between representations and real-world phenomena. Following theories of representation in human vision [27], Wand and Weber argue that for a grammar to be ontologically expressive, such mappings should be isomorphic. Thereby, the theory identifies four theoretical constructs to describe four properties of conceptual modeling grammars (more precisely, four ontological deficiencies) that potentially undermine its ontological expressiveness as shown in Figure 1.

![Figure 1: Main constructs and relations in Wand and Weber’s Theory of Ontological Expressiveness](image)

To examine construct deficit, construct redundancy, construct overload, and construct excess in a conceptual modeling grammar, it is required to perform an ontological analysis of a conceptual modeling grammar [6]. In this process, the constructs of a
conceptual modeling grammar are mapped to the constructs specified in Wand and Weber’s ontological model of an information system. Through this mapping process, construct deficit, construct redundancy, construct overload, and construct excess in a modeling grammar are defined as follows (see Table 2).

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontological Expressiveness</td>
<td>The ability of a modeling grammar to describe real-world phenomena completely and clearly, as defined by an underlying ontological model of representation.</td>
</tr>
<tr>
<td>Ontological Completeness</td>
<td>Ontological completeness is achieved when the user of a conceptual modeling grammar is able to articulate with the grammar all the types of real-world phenomena he/she seeks to have articulated in a model.</td>
</tr>
<tr>
<td>Construct deficit</td>
<td>An ontological construct exists that has no mapping from any conceptual modeling grammar construct (a 1:0 mapping), resulting in the grammar lacking constructs to describe a particular real-world phenomena.</td>
</tr>
<tr>
<td>Ontological Clarity</td>
<td>Ontological clarity is achieved when the user of a conceptual modeling grammar is able to articulate in the model all those real-world phenomena in a manner that allows for unambiguous interpretation.</td>
</tr>
<tr>
<td>Construct redundancy</td>
<td>Two or more conceptual modeling grammar constructs map to a single ontological construct (a 1:m mapping), resulting in the grammar offering multiple constructs to describe the same type of real-world phenomena.</td>
</tr>
<tr>
<td>Construct overload</td>
<td>A single conceptual modeling grammar construct maps to two or more ontological constructs (a m:1 mapping), resulting in the grammar offering constructs that share the same real-world meaning.</td>
</tr>
<tr>
<td>Construct excess</td>
<td>A conceptual modeling grammar construct does not map onto any ontological construct (a 0:1 mapping), resulting in the grammar offering constructs that have no real-world meaning.</td>
</tr>
</tbody>
</table>

Wand and Weber’s theory of ontological expressiveness argues that lack of ontological completeness and lack of ontological clarity – expressed in the four ontological deficiencies of construct deficit, construct redundancy, construct overload, and construct excess – in a conceptual modeling grammar undermine a
user’s ability to create models of real-world phenomena that contain representations of all required real-world phenomena and/or that are difficult to understand, confusing or ambiguous.

In this paper, our interest is in the four theoretical constructs *construct deficit*, *construct redundancy*, *construct overload*, and *construct excess*. Specifically, we are concerned with measuring user perceptions of these constructs. We realize that our work presupposes an ontological analysis through which construct deficit, construct redundancy, construct overload, and construct excess in a modeling grammar can be identified. Such work has been carried out for a wide range of conceptual modeling grammars (see [11, 28] for overviews) and the process of an ontological analysis itself has also been widely discussed [e.g., 29, 30], providing ample support for scholars interested in such research.

Foreshadowing our elaborations below, we note that in our own prior work, we performed an ontological analysis of the conceptual modeling grammar we examine in our study, the ‘Business Process Modeling Notation’ (BPMN) [31]. Details of this analysis are reported in [11, 13]. This study found that BPMN exhibits construct deficit related to the articulation of business rules, the history of state changes, and process structure and decomposition, and construct redundancy related to the articulation of real-world objects, transformation and events. Further, the analysis revealed construct overload in the BPMN construct ‘Pool’ and ‘Lane’, and construct excess in overall twelve BPMN grammar constructs. These findings are relevant to our measurement instrument development, which we report in section 3 of this paper.
2.3. Related Work

Wand and Weber’s work based on Bunge’s theory is not the only case of ontology-based research. Today, interest in, and applicability of ontology, is a widely discussed topic in research on conceptual modeling quality [18, 24] and development [17], but also extends to areas beyond conceptual modeling. For instance, ontologies have been developed to share organization and problem specific knowledge among supply chain’s stakeholders [32], or to support evolution and re-use within the Semantic Web [33, 34]. As these selected examples show, the theory of ontology has emerged as a fruitful base from which a wide range of theoretical contributions in the IS discipline stem from.

Concerning Wand and Weber’s theory of ontological expressiveness, a number of studies have been carried out on the basis of their work to investigate the quality of conceptual modeling. This work has in common that it shows how ontological completeness and ontological clarity affects the quality of conceptual modeling procedures and outcomes. Green et al. [24], for instance, showed how multiple grammars can be combined to mitigate construct deficit exhibited by each of the grammars in isolation. Bodart et al. [35] and Gemino and Wand [36] showed how construct excess resulted in users making more understanding errors when reading the model created with the grammar. Similarly, Shanks et al. [26] demonstrated that construct overload undermines users’ ability to understand the information contained in the model produced.

Other authors have undertaken similar empirical tests of the validity of the predictions stemming from Wand and Weber’s theory [e.g., 37, 38]. These studies also found that Wand and Weber’s theory indeed informs researchers about conceptual modeling
activities, outcomes and success, and, moreover, leverages ‘better’ conceptual modeling. Recker [12] reviewed the main works concerning Wand and Weber’s theory of ontological expressiveness and its use in combination with various conceptual modeling grammars such as structured, data-oriented, object-oriented, process–oriented, enterprise systems interoperability, use case, or reference modeling grammars. However, of the seventy-four works reviewed in [12], only 12.2 per cent of works involved empirical studies of some kind, whilst over 87.8 per cent of the studies applied Wand and Weber’s theory to study conceptual modeling analytically.

The limited number of empirical studies may be attributed to a missing operationalisation of the theory constructs [39]. Without the existence of a theoretically sound, rigorously developed and thoroughly tested measurement instrument, however, research efforts are endangered to lead to mixed and inconclusive outcomes. Accordingly, the next section reports on our effort to derive a measurement instrument for the four main constructs of Wand and Weber’s theory of ontological expressiveness (viz., construct deficit, construct redundancy, construct overload, and construct excess).

### 3. Instrument Development

A measurement instrument for the four theoretical constructs of interest (construct deficit, construct redundancy, construct overload, and construct excess), is required to meet the following criteria for validity and reliability [40]:

1. **Content validity**: Does the instrumentation pull in a representative manner from all of the ways that could be used to measure the content of a given theory construct?
2. **Construct validity**: Are the instrument items selected for a given concept, considered together (convergent validity) and compared to other (discriminant validity) latent constructs, a reasonable operationalization of the theory construct?

3. **Reliability**: Are the instrument items selected for a given theory concept, taken together, error-prone operationalizations of that theory construct?

To meet these objectives, we employ a procedural model that extends, and consolidates, suggestions for measurement instrument development firstly described by Davis [41] in his study of technology acceptance, and later extended by Moore and Benbasat [10] as well as Recker and Rosemann [42]. Figure 2 shows the procedural model. It describes in five stages the different tasks to be performed (grey rounded boxes), related inputs and outputs (white rectangles), and the source of decision making, i.e., the relevant literature or the source of empirical data where applicable (dark grey rectangles). In the column to the right, relevant source literature is displayed about previous uses of the procedural model tasks.
Stage One:
Item Creation
- Relevant domain literature
- Define theoretical constructs
- Generate candidate items

Stage Two:
Substrata Identification
- Theory expert panel
- Identify domain substrata
- Assess candidate items
- Candidate items
- Relevance domain substrata

Stage Three:
Item Identification
- Theory expert panel
- Rank candidate items
- Select suitable items
- Top candidate items

Stage Four:
Item Selection and Revision
- Practitioner panel
- Sort items in categories
- Revise items
- Final candidate items

Stage Five:
Instrument Validation
- Conduct field study
- Establish reliability
- Establish validity
- Survey data
- Validated items

Construct definitions
Relevant domain literature
Theory expert panel
Practitioner panel
Survey with practitioners

Literature
[33, 34]
[49]
[14, 35]
[14, 36]
[34, 71]

Figure 2. Instrument Development Process
As shown in Figure 2, the first stage of the procedural model is *item creation*, which is concerned with specifying the theoretical constructs for which measurement items are to be developed, and deriving pools of candidate items for each constructs. This task is carried out through an analysis of the relevant literature. The next stage is *substrata identification*, the purpose of which is to sort the candidate items into meaningful separate domain sub categories to display convergent and discriminant validity. This task is carried out with the help of a panel study with experts of the selected theory. The third stage is *item identification*, the purpose of which is to identify from the pool of candidate items a revised set of items that show good potential for high content validity. This task is also carried out by means of a theory expert panel study. The fourth stage is *item selection and revision*, the purpose of which is to re-specify and further condense the set of candidate items as well as to get an initial indication of reliability and validity. This task is carried out through a practitioner panel study. The last stage is *instrument validation*, which is concerned with obtaining statistical evidence for reliability and validity of the developed measurement items. This task is carried out by means of the survey research method.

In the following, we report in detail on how we carried out the measurement instrument development process, using the case of the BPMN modeling grammar [31]. BPMN is an important industry standard for the design of process-aware information systems, web services and service-oriented architectures alike [43]. BPMN contains overall 38 grammar constructs, grouped into four basic categories of elements, viz., Flow Objects, Connecting Objects, Swimlanes and Artefacts. For further information on BPMN, the reader can refer to its specification [31] as well as various textbooks about BPMN [e.g., 44].
We note that we use the example of BPMN for illustration purposes only. While the selection of BPMN as a target grammar limits the scope of our research effort, and whilst we lack evidence for this, we have no reason not to believe that our procedure can be adopted for studies of other conceptual modeling grammars (e.g., data-oriented or object-oriented grammars) for which an ontological analysis has been, or will be, performed.

3.1 Stage 1: Item Creation

The objective of the item creation step is to establish content validity of the measurement items. To that end, a sound specification of the theoretical constructs to be measured is required. This is because items are required to closely fit the content domains of the construct definitions [40].

In our case, Wand and Weber’s theory of ontological expressiveness offers four constructs (properties of conceptual modeling grammars) that together inform the ontological completeness and clarity of a modeling grammar, these being construct deficit, construct redundancy, construct overload and construct excess. Typically, the extent to which modeling grammar exhibit any of these ontological deficiencies can be established by means of ontological analysis [6]. However, it needs to be considered that the four ontological deficiencies (indicated by the extent to which a grammar exhibits construct deficit, redundancy, overload or excess) do not in all cases imply a practically observable disadvantage or issue [36]. In other words, ontological deficiencies are not necessarily perceived as such. Therefore, we consider the secondary attributes of modeling grammars (i.e., user perceptions of the levels of ontological completeness and clarity), instead of the primary attributes of modeling grammars (i.e., their actual levels of ontological completeness and clarity).
Our decision follows the arguments voiced by Downs Jr. and Mohr [45] who concede that *secondary attributes* of an artifact, i.e., an individual’s perceptions of its *primary attributes*, determine the formation of an attitude towards it, and thereby influence associated usage behavior. They argue that it is most often not the actual attributes of an artifact that will influence an attitude towards it, but rather the perception of the attributes by the user [10]. Consider the case of a product that is annotated with a certain price. A purchase decision will not be made on the basis of the actual (i.e., primary) price attribute but rather on whether an individual perceives the price to be reasonable or too expensive (i.e., the secondary attribute).

In studies of conceptual modeling grammars, therefore, it should be considered whether users of conceptual modeling grammars perceive primary attributes (in this case, the actual levels of construct deficit, construct redundancy, construct overload, and construct excess) as indeed being existent (a secondary attribute). There are two aspects to consider:

- whether or not a modeling grammar has a perceived ontological deficiency. For instance, whether users of a conceptual modeling grammar experience a deficit of construct and/or a number of constructs that are unclear (i.e., overloaded, redundant, or excess), and

- whether or not these grammar deficiencies have a practical or observable impact on users using the grammar. For instance, whether users of a modeling grammar perceive that they are unable to articulate a certain phenomenon in a conceptual model because there is construct deficit in the grammar.

Taking these considerations into account, we adopted the original definitions of the four theoretical constructs in Wand and Weber’s theory [5-7] as follows (note that
during the theory expert panel test reported in section 3.2 and section 3.3 below, the participating scholars were asked for feedback on the definitions. The definitions shown below incorporate the suggestions voiced):

1. **Perceived construct deficit (PCD):** The extent to which a conceptual modeling grammar user perceives the grammar to have a deficit of constructs that (s)he would require to describe all real-world phenomena that (s)he seeks to have represented in a conceptual model.

2. **Perceived construct redundancy (PCR):** The extent to which a conceptual modeling grammar user perceives the grammar to provide more constructs than required to describe a single real-world phenomena that (s)he seeks to have represented in a conceptual model.

3. **Perceived construct overload (PCO):** The extent to which a conceptual modeling grammar user perceives the grammar to provide constructs that can each be used to describe more than one single real-world phenomena in a conceptual model.

4. **Perceived construct excess (PCE):** The extent to which a conceptual modeling grammar user perceives the grammar to provide constructs that do not describe any relevant real-world phenomena in a conceptual model.

Forthcoming from the specification of these four definitions is the need to pursue a set of appropriate candidate items for each of these ontological deficiencies, with the objective being to ensure content validity [10].

The identification of candidate items is very important to the further conduct, and outcomes, of the measurement instrument development. This is because measurement items that are slightly differently worded may yield different answers in an empirical
study [46]. Therefore, offering a large pool of candidate items with slightly different nuances of meaning is instrumental to providing a good foundation for identifying an appropriate final measurement item.

In identifying appropriate candidate items, typically literature reviews are recommended [10, 41, 46]. We examined the available literature concerning Wand and Weber’s theory of ontological expressiveness in order to generate pools of candidate items for each of the four ontological deficiencies. More specifically, we scanned the 74 studies reviewed in [12] to identify candidate items for each theory construct.

Given that content validity generally cannot be empirically assessed, it cannot be guaranteed that enough instrument items are drawn because the content universe of the items itself is indeterminate [40]. Some guidelines, however, exist. For instance, the Spearman-Brown Prophecy formula [47] estimates the number of items needed to achieve reliability levels of at least 0.80 based on the number of items and reliability of comparable existing scales. We used literature on technology acceptance [41] and innovation adoption [10] as reference studies, and extrapolated a target number of ten candidate items, similar to other cases of measurement instrument development reported [10, 41, 42, 48]. With the creation of at least ten candidate items per construct, we felt convinced that we could adequately cover all potential dimensions or interpretations of the theoretical constructs.

As per specification of the candidate items, we considered Ajzen and Fishbein’s [49] suggestions to include in the definition of the items the actual behavior (e.g., using a conceptual modeling grammar to create models), the target at which the behavior is directed (e.g., the Entity-Relationship Modeling Notation as a conceptual modeling grammar in use), the context in which the behavior occurs (e.g., for database
specification tasks) and, where possible, a time frame (e.g., the most recent database design project).

Following this approach, we specified the candidate items in explicit relation to the case of the BPMN grammar [31]. This was done to make the item more tangible and understandable, as well as to refer to a specific context of behavior, viz., conceptual modeling for specifying process-aware information systems.

Further, in our item specification, instead of measuring user perceptions of the ontological deficiencies of the BPMN grammar as a whole (e.g., through items such as “I believe that the BPMN grammar is redundant”), we operationalized the measures for perceived construct deficit, perceived construct redundancy, perceived construct overload, and perceived construct excess by relating them to the specific BPMN grammar construct(s) to which they apply. In doing so, we used the findings from the ontological analysis of BPMN grammar as reported in [11, 13]. For instance, the BPMN grammar construct ‘Lane’ was found to exhibit construct overload (an actual ontological deficiency) [13]. Instead of measuring this deficiency in a general fashion (e.g., through items such as “I believe that the BPMN grammar contains construct overload”), our measurement instrument operates on the level of the actual construct that exhibits the deficiency in question – in this example the Lane construct.

This way of specifying the items for the measurement instrument, in turn, allows us to examine the actual features of a grammar, viz., the nature and type of its constructs. Wand and Weber’s theory of ontological expressiveness allows researchers to speculate about the nature and implications of conceptual modeling grammar constructs. Accordingly, we derive measurement items that operate on the same level as the actual ontological deficiencies of a modeling grammar, as identified through an ontological analysis. Adopting this view also allows us to instantiate the generically
worded measurement items (e.g., “The BPMN modeling grammar does not provide sufficient symbols to represent certain real-world phenomena in process models”) using concrete examples with specific wording (e.g., “The BPMN modeling grammar does not provide sufficient symbols to represent business rules in process models”). This approach, in turn, will likely make the final items more understandable to end users participating in field tests.

In our case, from the ontological analysis of BPMN described in [13] that identified overall twelve types of ontological deficiencies (viz., three instances of construct deficit, three instances of construct redundancy, two instances of construct overload, and four instances of construct excess), we used eight ontological deficiencies identified (two per type of ontological deficiency) to operationalize measurement items. In effect, we covered each type of ontological deficiency without making the final survey instrument unnecessarily long.

In the interest of brevity, we omit the lists of initial candidate item pools generated from the literature review. The lists of candidate item pools are available from the contact author upon request.

3.2 Stage 2: Substrata Identification

Forthcoming from the generation of an initial pool of candidate items is the establishment of construct validity of the candidate items, in particular the display of convergent and discriminant validity. To that end, we employed a procedure called ‘own category test’ [50]. In this test, a panel of judges is asked to sort candidate items into a number of construct categories so that the statements within a category are most similar in meaning to each other and most dissimilar in meaning from those in other categories. The resulting categories, in turn, can be considered a reflection of the domain substrata for each theoretical construct, i.e., they are representative of the
different dimensions of meaning a theoretical construct may carry [41]. The categories are also to be labeled. The labels are then used to assess whether the identified substrata appropriately reflect the item’s intent.

In our case, a panel of four recognized academics with a strong track record in studies using Wand and Weber’s theory was asked to sort the initial candidate items into construct categories. Panel members were selected from the overall population of IS researchers that have published articles on studies that build upon Wand and Weber’s theory of ontological expressiveness [e.g., 11, 26, 38, 51-57]. Each panel member was contacted individually and received instructions including a sample categorization test. This step was done to ensure the mechanics of the test were fully understood by the participating panel members.

The panel members were free to select as many categories as they deemed appropriate for clustering the candidate items. In this case, the panel members used four or fewer categories. The substrata identification results can be obtained from the contact author upon request.

The data was then cluster analyzed. This was done to identify domain substrata of the primary theoretical constructs. In performing the clustering of the categories obtained from the panel members, two coders separately clustered the categories, then met to defend their clusters and created a joint draft, thereby reducing subjectivity in the coding procedure. We calculated inter-coder reliability by means of Cohen’s [58] Kappa. We obtained a Kappa value of .72 when comparing the individual clustering results, indicating sufficient reliability [40]. We found two very clear and strong clusters to emerge across all four theoretical constructs, one tapping into ‘type of grammar deficiency’ and the other tapping into ‘impact of the deficiency on the user of the grammar’. For each of the four theoretical constructs, these two domain
substrata, in turn, reflect the basic premise of Wand and Weber’s theory, namely that (a) a grammar may exhibit a specific type of an ontological deficiency, and (b) this deficiency may reduce the completeness and/or clarity of any model created with the grammar. Consequently, for all subsequent tasks the identification of these two domain substrata serves as a reference point to ensure adequate coverage of the domain substrata for each of the theoretical constructs.

A second step was to assess whether panel members consistently placed the same candidate items in these clusters. Following [10], we demonstrate reliability of the cluster scheme by assessing the percentage of items placed in the target cluster across all panel members (the column placement ratio), which indicates the degree of inter-judge agreement. Also, the items that obtained high placement percentages across the panel show high potential for high construct validity and reliability. The four panel members agreed in at least 75% of their item placements when mapped to the two coded substrate deficiency and impact, which in turn increased confidence in the validity and reliability of the cluster scheme.

3.3 Stage 3: Item Identification
The goal of the item identification stage was to establish differences in content validity between the candidate items in order to be able to drop items that show little potential for high validity. To that end, a panel of theory experts familiar with Wand and Weber’s theory of ontological expressiveness was asked to assess, on a 7-point scale, the correspondence between the candidate items and the definitions of the constructs they are intended to measure. This step followed the procedures firstly documented by Davis [41].

The expert panel consisted of recognized IS academics with varying yet strong track records in studies on basis of Wand and Weber’s theory. By forming a panel of theory
experts, higher potential for correctly assessing the validity of candidate items was achieved. In addition, we sought informal, qualitative feedback on our development procedure. Overall, twenty-three contributors to studies on basis of Wand and Weber’s theory of ontological expressiveness were identified, based on the works reviewed in [12], and contacted. A response rate of 57% (13) was obtained. Each of the thirteen participating panel members received instructions, including a sample test as well as construct definitions and candidate items. Additional space was allowed for extended reasoning and feedback.

Following the approach described in [41], the responses of the panel members were averaged and ranked to obtain an order of candidate items with respect to their content validity. This was done to identify potential candidates for elimination. In eliminating items, however, it had to be considered whether the remaining item pool contains appropriate representativeness of the domain substrata (deficiency and impact) of the theoretical construct [59]. Hence, in analyzing the results attention was also paid to the results of the categorization task (see previous section) in order to identify domain substrata of which the item pool may have excessive, or inadequate, coverage. As an example, one candidate item (“BPMN lacks capacities for representing certain real-world phenomena in process models”) received a relative good ranking but was found not to resemble any of the identified two domain substrata. However, given its high ranking, another candidate item (“BPMN users lack capacities to represent certain real-world phenomena in process models”) was reworded to incorporate some of the content of the other (“BPMN does not provide me with sufficient capacities for representing certain real-world phenomena in process models”).

Overall, this step resulted in an ordered ranking of potential construct validity of the candidate items. This ranking was then used to eliminate items that demonstrated low
potential for validity. Both the ranking and categorization data can be obtained from the contact author upon request.

In making the final decision about whether to drop, retain or merge each of the candidate items, the ranking data together with the categorization data should be considered together with qualitative feedback from the panel. This is because the step of establishing content validity (the objective of this stage) relies on heuristics such as literature reviews, expert panels or judges, and also requires subjective judgment by the research team. Accordingly, not only the quantitative data (ranking, cluster placement ratio) must be considered but also qualitative feedback and judgmental evaluation, to avoid potential discriminations of content validity. The consequences of the decisions made at this stage will then be evaluated during item selection and revision (stage 4) and instrument validation (stage 5). Any potential violations that may be noted during these stages would then require reverting to this earlier stage of the instrument development process.

In our case, some items were reworded based on suggestions, others were reworded to incorporate content from items that exhibited high rankings but failed to cluster appropriately, some items were reworded with the view to improving content validity ranking whilst maintaining the clustering into one of the two domain substrata, and some items were merged in order to improve content and wording. In addition, informal qualitative feedback from the panel members was used to improve the wording of some items.

3.4 Stage 4: Item Selection and Revision
The objective of the fourth stage of the process was to revise the reduced set of candidate items to a final set of high potential candidate items and to further improve
their validity and reliability. An appropriate procedure for this type of task is the index card sorting test suggested by Moore and Benbasat [10].

In this test, a panel of judges is randomly given the items printed on index cards and asked to sort these cards into categories. In different rounds of this test, the categories in which the items are to be sorted into are either given to the panel of judges or not. As suggested by Moore and Benbasat [10], we used four rounds of sorting, each with a different panel, and alternated between given and not-given categories. In rounds one and three, judges independently had to make up categories for the items, which were later compared to the originally intended categories. In rounds two and four, judges were asked to sort items into categories given to them, and to identify items that are ambiguous or indeterminate.

At this stage of the procedure, the panel should be indicative of the target population of the final field study [10], in order to further improve content validity. The objective is to specify measurement items that are most likely to be well understood in the final field test. Accordingly, it is imperative to identify the key characteristics of the intended target population (e.g., novice versus experts, managers versus field workers, professionals versus students etc.), and to select members for the panel so that these key characteristics can be met. We selected panel members that were not familiar with the underlying theory of ontological expressiveness. Overall, sixteen judges, including professional staff, consultants, analysts and post-graduate students, participated over the four rounds, none of them familiar with the study domain but all of them familiar, in varying degrees, with conceptual modeling. This panel, therefore, is an adequate proxy for varying types of modeling practitioners. In each round, the panel size varied between three and five members. Each panel of judges was gathered together in a
face-to-face setting to explain the intent and mechanics of the test. Two trial sorts were conducted prior to the actual sorting to increase familiarity with the procedure.

To assess the reliability of the sorting conducted by the judges, two measurements wereestablished. Table 2 summarizes coding reliability results in terms of placement ratio summaries across all four rounds of sorting and also displays inter-judge agreements measured using Cohen’s Kappa. Round-by-round revisions helped improve reliability so that, in the end, generally recommended Kappa levels of .7 [40] were met.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
<th>Round 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Average Kappa</em></td>
<td>0.34</td>
<td>0.64</td>
<td>0.43</td>
<td>0.80</td>
</tr>
<tr>
<td>Placement ratio summary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct deficit</td>
<td>100.00%</td>
<td>100.00%</td>
<td>87.50%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Construct redundancy</td>
<td>58.33%</td>
<td>87.50%</td>
<td>50.00%</td>
<td>66.67%</td>
</tr>
<tr>
<td>Construct overload</td>
<td>16.67%</td>
<td>62.50%</td>
<td>66.67%</td>
<td>80.00%</td>
</tr>
<tr>
<td>Construct excess</td>
<td>33.33%</td>
<td>75.00%</td>
<td>87.50%</td>
<td>70.00%</td>
</tr>
<tr>
<td><em>Average</em></td>
<td>52.08%</td>
<td>81.25%</td>
<td>72.92%</td>
<td>79.17%</td>
</tr>
</tbody>
</table>

From Table 2 it can be observed how results vary between Rounds 1,3 and 2,4, respectively. This situation was to be expected given that in rounds 1 and 3 judges were not given item categories, which made it harder to categorize the items correctly. There is no definite guidance available as to when to stop this exercise. It is advisable, however, to proceed with the rounds of sorting until inter-judge agreement meets acceptable reliability levels for the rounds in which categories were provided as well as for those rounds where categories were not provided. Landis and Koch [60] suggest different Kappa thresholds, from fair and substantial (0.4 and above) to excellent (0.8 and above), that could be used as a basis for decision making. As can be seen, these
thresholds were met in our coding exercise after round one, and reached excellent levels after round four, which we deemed sufficient.

After each round, each set of items was inspected and, if deemed necessary, reworded. This was done by contacting one member of the panel involved in the rounds to discuss the findings, and to inquire about potential ways of improving the wording of the items. Some items that were repeatedly misplaced (and thus showed only little potential for high validity) were dropped. The Appendix gives an overview of the resulting top three candidate items for each construct after these four stages of instrument development (note that the Appendix displays the items in their final wording after pre- and pilot-tests (see section 3.5). In the Appendix, we show, for each theory construct as used in the final field test (see section 3.5), the three measurement items (No, and Item Definition), worded in accordance to the actual ontological deficiency exhibited by the BPMN grammar.

We selected three items per construct for the pragmatic reason of keeping the overall instrument concise and short and for the technical reason of maintaining the minimum number of items required for appropriate measurement model estimation [61]. For each construct, both the ‘deficiency’ substratum of the construct and the ‘impact’ substratum are covered by at least one item.

3.5 Stage 5: Instrument Validation
The next step was to conduct a confirmatory factor analysis of the measurement instrument developed with a sample of modeling practitioners. The objective was to ensure that the mechanics of compiling the measurement instrument had been adequate and to obtain formal measures for reliability and validity. To that end, we implemented the candidate items listed in the Appendix using the example of the BPMN modeling grammar in a survey instrument, which is the typical way of
validating measurement instruments in IS [40]. The Appendix lists the all items in the measurement instrument as used in the final field test.

Before administering the field study we ran a pre-test and a pilot test. In the pre-test four academics with knowledge of the study were asked to complete a paper-based version of the survey instrument in face-to-face meetings. During survey completion, notes were taken based on comments received. After instrument revision, the measurement instrument was pilot-tested with a sample of 41 post-graduate students with knowledge of the target grammar. After exploratory factor analysis, changes were made to the measurement instrument and to the items that indicated problems in meeting required validity and reliability thresholds.

The population of interest for this study included conceptual modelers who have knowledge of a certain modeling grammar, viz., BPMN. To that end, a web-based survey instrument was crafted and announced via modeling practitioner forums and online groups. Overall, 590 usable results were obtained over a period of four months during 2007. Participants were selected using a judgmental sampling technique. Chi-square tests of key demographic variables showed no significant differences in responses between early and late respondents, indicating the absence of non-response bias.

Reliability and validity for the measurement instrument was assessed via confirmatory factor analysis (CFA) techniques implemented in LISREL Version 8.80. Each measurement item was modeled as a reflective indicator of its hypothesized latent construct. All constructs were allowed to co-vary in the CFA model. Table 3 gives the results from the item validation and Table 3 gives the corresponding factor correlation matrix.
<table>
<thead>
<tr>
<th>Construct</th>
<th>Item</th>
<th>Item loading</th>
<th>t-statistic (for $\lambda$)</th>
<th>Cronbach’s α</th>
<th>$\rho_c$</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCD1 [Business Rules]</td>
<td>PCD1_1</td>
<td>0.916</td>
<td>26.957</td>
<td>0.868</td>
<td>0.847</td>
<td>0.916</td>
</tr>
<tr>
<td></td>
<td>PCD1_2</td>
<td>0.856</td>
<td>27.075</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCD1_3</td>
<td>0.889</td>
<td>26.958</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCD2 [Process structure]</td>
<td>PCD2_1</td>
<td>0.939</td>
<td>37.226</td>
<td>0.925</td>
<td>0.883</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>PCD2_2</td>
<td>0.874</td>
<td>33.499</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCD2_3</td>
<td>0.889</td>
<td>37.224</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR1 [Transformation s]</td>
<td>PCR1_1</td>
<td>0.832</td>
<td>27.435</td>
<td>0.911</td>
<td>0.863</td>
<td>0.928</td>
</tr>
<tr>
<td></td>
<td>PCR1_2</td>
<td>0.891</td>
<td>29.785</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR1_3</td>
<td>0.906</td>
<td>27.434</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR2 [Events]</td>
<td>PCR2_1</td>
<td>0.922</td>
<td>31.012</td>
<td>0.900</td>
<td>0.877</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td>PCR2_2</td>
<td>0.939</td>
<td>33.157</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR2_3</td>
<td>0.880</td>
<td>31.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCO1 [Pool]</td>
<td>PCO1_1</td>
<td>0.811</td>
<td>30.349</td>
<td>0.917</td>
<td>0.855</td>
<td>0.926</td>
</tr>
<tr>
<td></td>
<td>PCO1_2</td>
<td>0.871</td>
<td>31.148</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCO1_3</td>
<td>0.842</td>
<td>30.358</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCO2 [Lane]</td>
<td>PCO2_1</td>
<td>0.878</td>
<td>31.853</td>
<td>0.914</td>
<td>0.864</td>
<td>0.932</td>
</tr>
<tr>
<td></td>
<td>PCO2_2</td>
<td>0.798</td>
<td>33.617</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCO2_3</td>
<td>0.828</td>
<td>31.938</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCE1 [Off-page connector]</td>
<td>PCE1_1</td>
<td>0.927</td>
<td>42.984</td>
<td>0.967</td>
<td>0.924</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>PCE1_2</td>
<td>0.952</td>
<td>43.603</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCE1_3</td>
<td>0.920</td>
<td>43.955</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCE2 [Multiple Instances]</td>
<td>PCE2_1</td>
<td>0.943</td>
<td>27.453</td>
<td>0.945</td>
<td>0.907</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>PCE2_2</td>
<td>0.917</td>
<td>27.145</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCE2_3</td>
<td>0.913</td>
<td>28.813</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>EXP_1</td>
<td>0.713</td>
<td>6.253</td>
<td>N/A</td>
<td>0.812</td>
<td>0.897</td>
</tr>
<tr>
<td></td>
<td>EXP_2</td>
<td>0.820</td>
<td>7.096</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EXP_3</td>
<td>0.994</td>
<td>5.543</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on the data obtained and displayed in Table 3 and Table 4, four tests can be performed [40]. Regarding uni-dimensionality, Cronbach’s $\alpha$ should be greater than or equal to .7 to consider items to be uni-dimensional and to be combinable in an index.
Table 3 shows that all constructs have $\alpha$ of at least .8, thereby meeting the test of uni-dimensionality.

<table>
<thead>
<tr>
<th></th>
<th>EXP</th>
<th>PCD1</th>
<th>PCD2</th>
<th>PCR1</th>
<th>PCR2</th>
<th>PCO1</th>
<th>PCO2</th>
<th>PCE1</th>
<th>PCE2</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCD1</td>
<td>-0.218</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCD2</td>
<td>0.107</td>
<td>-0.146</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR1</td>
<td>-0.163</td>
<td>0.332</td>
<td>-0.078</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR2</td>
<td>-0.214</td>
<td>0.479</td>
<td>-0.059</td>
<td>0.481</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCO1</td>
<td>-0.114</td>
<td>0.347</td>
<td>0.050</td>
<td>0.356</td>
<td>0.268</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCO2</td>
<td>-0.075</td>
<td>0.355</td>
<td>-0.005</td>
<td>0.132</td>
<td>0.152</td>
<td>0.475</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCE1</td>
<td>0.029</td>
<td>0.146</td>
<td>-0.106</td>
<td>-0.051</td>
<td>-0.049</td>
<td>0.192</td>
<td>0.085</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>PCE2</td>
<td>-0.101</td>
<td>0.276</td>
<td>-0.016</td>
<td>0.227</td>
<td>0.310</td>
<td>0.210</td>
<td>0.404</td>
<td>0.046</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Reliability refers to the internal consistency of a measurement instrument. Again, the most widely used test for internal consistency is Cronbach’s $\alpha$, which should be higher than .8. A second test uses the composite reliability measure $\rho_c$, which represents the proportion of measure variance attributable to the underlying trait. Scales with $\rho_c$ greater than .5 are considered to be reliable. Table 3 shows that all constructs obtained $\alpha$ of at least .8 and also meet the required $\rho_c$ cut-off value of .5. These results suggest adequate reliability.

Convergent validity tests if measures that should be related are in fact related. Convergent validity can be tested using three criteria suggested by Fornell and Larcker [62]: (1) all indicator factor loadings ($\lambda$) should be significant and exceed 0.60, (2) construct composite reliabilities $\rho_c$ should exceed 0.80 and (3) average variance extracted (AVE) by each construct should exceed the variance due to measurement error for that construct (i.e., AVE should exceed 0.50). Table 3 shows that all factor loadings $\lambda$ are significant at $p < .001$ (see the reported t-values) and
exceed the recommended threshold of 0.60. In terms of composite reliabilities, Table 3 shows that pc exceeded 0.80 for all constructs. As reported in Table 3, AVE for each construct is higher than 0.90 suggesting that for all constructs AVE well exceeded the variance due to measurement error. Overall, it is concluded that the conditions for convergent validity were met.

Discriminant validity tests if measures that should not be related are in fact unrelated. Fornell and Larcker [62] recommend a test of discriminant validity, where the AVE for each construct should exceed the squared correlation between that and any other construct considered in the factor correlation matrix. In the present study, the largest squared correlations between any pair of constructs within the measurement model was 0.231 (between PCR1 and PCR2, see Table 4), while the smallest obtained AVE value was 0.916 (PCD, see Table 3). These results suggest that the test of discriminant validity is met.

4. An Application of the Instrument

To illustrate the utility of the measurement instrument developed, we consider the question whether the perceptions of ontological deficiencies of modeling grammars change between users with different levels of modeling experience.

Modeling is essentially a problem solving activity where humans create models of a problem domain to aid them with tasks such as information systems analysis, organizational re-design, simulation, requirements specification and others. *Experience* is an important factor in such problem solving activities [63]. Experienced modelers typically possess a repertoire of work-arounds, or patterns, for modeling problems they encountered before, resulting in differences to novices in the way
conceptual modeling is being conducted, and modeling grammars applied for modeling-related tasks [64].

We believe that less experienced modelers often have not yet encountered modeling scenarios in which certain ontological deficiencies would induce problems in the use of the grammar. For example, if a modeler has not yet used a certain potentially ambiguous (i.e., ontologically overloaded) grammar construct (s)he would not know how critically the related grammar deficiency would impact his/her modeling. Likewise, if not required for modeling, construct deficit in the grammar simply will not manifest in a user’s perception. In contrast, more experienced modelers are likely to have already encountered a wide range of problematic situations with the use of a modeling grammar, which may have enticed them to develop, and use, a repertoire of workarounds for such situations. This situation would suggest that weakness inherent in a grammar (that may stem from a lack of ontological completeness and/or clarity) will manifest themselves more strongly in the perceptions of end users working with the grammars, when experience is acquired. We formalize these arguments in the following hypothesis:

\[ \text{H1. Perceptions of ontological deficiencies of a modeling grammar will be higher for users with high levels of modeling experience.} \]

In order to test this hypothesis, we collected data about the levels of prior modeling experience brought to bear by the users of the BPMN modeling grammar. We collected three measures of experience, similar to prior studies [36, 65]:

- Self-reported approximate number of years experience in modeling overall (EXP1),
- Self-reported approximate number of months experience in the modeling grammar under observation, BPMN (EXP2), and
• Self-reported approximate number of models created (EXP3).

We tested validity and reliability of the compound experience (EXP) scale in the same fashion as for the other scales (see above), and obtained adequate results (see Table 3 and Table 4). Factor loadings were significant and exceeded 0.700, and composite reliability of the EXP scale was 0.812 (Cronbach’s Alpha could not be computed due to the continuous nature of the scale). AVE was 0.897, which exceeded the largest squared correlation of any factor pair (see Table 4).

Examination of hypothesis H1 was carried out using structural equation modeling (SEM) implemented in LISREL Version 8.80 [61]. SEM is appropriate for testing theoretically justified models [40], as was the case in this study. Each indicator was modeled in a reflective manner, and the theoretical constructs were linked as hypothesized in hypothesis H1. Results of our examinations of the suggested hypotheses are presented in Figure 3. The structural model showed adequate fit to the data (GFI = 0.921, AGFI = 0.913, NFI = 0.985, NNFI = 0.986, CFI = 0.989, SRMR = 0.0498, RMSEA = 0.070, $\chi^2 = 1132.21$, df = 316).

The results displayed in Figure 3 support hypothesis H1. For all eight instances of ontological deficiencies considered, scores were higher for users with higher levels of modeling experience (as indicated by the positive $\beta$ values in Figure 3). In seven out of eight cases, these paths were significant at $p < 0.001$, except for the path Modeling Experience – Perceived Construct Excess [Off-page Connector], where the $\beta$ value was insignificant at $p > 0.05$. Overall, the significance and positive directionality of the paths suggest that indeed more experienced modelers more strongly perceive ontological deficiencies in modeling grammars. On basis of these results, we can speculate that this situation is because more experienced modelers have, over time,
been exposed to a wide range of modeling scenarios, and are thus more likely to have been exposed to situations in which ontological deficiencies manifest.

![Diagram showing relationships between modeling experience and perceived constructs]

Figure 3. Summary of Model Results

5. Conclusions

5.1. Contributions

The research presented in this paper provides three central contributions:

1. This paper reported on the process of creating an instrument to measure user perceptions of ontological deficiencies of conceptual modeling grammars. The procedure employed was shown to ensure high levels of content validity, construct validity and reliability of the final instrument. This situation should motivate researchers to adopt this procedure in other empirical studies.
2. The measurement instrument can be used in various studies to investigate how individuals perceive ontological deficiencies of conceptual modeling grammars, and what the impact of such perceptions may be. Below we outline a number of research directions that could benefit from the use of such a measurement instrument.

3. We showed empirically how modeling experience influences the way users perceive ontological deficiencies in modeling grammars to be existent. More specifically, we showed that more experienced modelers perceive ontological deficiencies in a modeling grammar more strongly than inexperienced modelers.

5.2. Limitations

Notwithstanding the contributions provided, there are some caveats to be considered. The measurement instrument developed in this research addresses perceptions of users about the ontological deficiencies of conceptual modeling grammars they use for IS analysis and design. While perceptions and beliefs are fundamental to understanding actual behaviors [45], there are some concerns with the approach taken:

1. The instrument taps into secondary attributes of conceptual modeling grammars (user perceptions of ontological deficiencies) but not their primary attributes (actual ontological deficiencies). Analytical studies could be carried out to examine objective attributes of modeling grammars and to contrast these with the user perceptions of these attributes.

2. The instrument builds upon findings from ontological analyses of conceptual modeling grammars – which potentially induces subjective bias in the analysis [66]. Further, the measurement of user perceptions of ontological deficiencies...
is based on the assumption that the ontological deficiencies can be faithfully identified through an ontological analysis, and that the underlying ontological model is appropriate. In our study, we undertook several steps to ensure reliability and validity of the ontological analysis on which we based our work. More details about this analysis are reported in [11, 13]. Last, ontological analyses examine the graphical constructs in a conceptual modeling grammar but, typically, not the rules in the grammar that specify lawful combinations of the constructs. We note, however, that there are some examples that consider combination of constructs [e.g., 67].

3. During the development of the measurement instrument, the specificity of the measurement items may vary, because the pool of candidate items used in stages one and two of the procedure are worded more generally than the final items. This is because during stages three and four of the procedure, items are continuously re-assessed and re-specified based on the feedback from the expert and user panels, to increase validity as well as reliability of the items. In effect, the specificity of the wording of some of the items can be considerably different from the initial specification. The fifth stage of the procedure, however, enables researchers to validate statistically that all prior work on the items in fact improved validity and reliability of the items.

4. In this paper, we reported on the development of a measurement instrument for the case of the BPMN grammar specifically. Hence, all results from the measurement instrument development, as well as the findings from the application of the measurements, should be interpreted before the background of BPMN only. Specifically, we cannot guarantee beyond all levels of concerns that the measurement instrument can be adopted to other modeling
contexts or grammars.

5.3. Implications for Practice

Our work has implications for various stakeholders engaged in conceptual modeling, e.g., end users, modeling consultants, and managers in charge for modeling. Specifically, our examination of perceptions of ontological deficiencies may inform training programs in process modeling as well as the development of appropriate modeling conventions.

Regarding training, the fact that modelers indeed perceive ontological deficiencies in process modeling grammars, and the fact that more experienced modelers perceive these deficiencies even stronger, suggests that training programs in process modeling grammars should place an emphasis not only on the expressiveness and capabilities of a process modeling grammar (e.g., how do I use BPMN to effectively support exception handling) but also inform users about deficiencies and areas of confusion about the grammar (e.g., selecting the right event type, and using the Lane construct appropriately). Informing deficiencies of a grammar can establish awareness in a grammar user to critically reflect on the modeling capabilities offered by a process modeling grammar, and may ultimately lead to a more effective and efficient use of the grammar.

Second, our identification of modeler perceptions of ontological deficiencies of process modeling grammars can inform the development of process modeling conventions that could govern the use of a particular process modeling grammar in an organization. For instance, the fact the that users perceive construct overload in the Lane and Pool construct in BPMN should inspire the development of an organizational convention that restricts the areas of real-world meanings of these constructs to a predefined subset in an organization (e.g., Pools for organizational
entities, and Lanes for functional roles). This way, the ontological deficiency could be mitigated. Similar conventions could address the handling of business rules, the choice of appropriate event constructs, and the handling of excess constructs.

5.4. Implications for Future Research

Aside from practical merits of our work, we identify a number of research streams in conceptual modeling that could proceed on the basis of the research presented in this paper:

1. Studies of the post-adoptive usage behavior of conceptual modeling grammars [64] could study user perceptions of the attributes of a conceptual modeling grammar (e.g., the ontological deficiencies of its constructs), and examine the effects that these perceptions have on usage intentions or usability beliefs.

2. The ongoing stream of research that investigates the quality of conceptual models [e.g., 2] could use the instrument to understand how perceptions about the ontological deficiencies of a modeling grammar ultimately influence the formation of a perception of quality, and ultimate usage, of a model produced with the grammar under observation.

3. Studies that examine the effect of modeling experience [64] or modeling training [68] on modeling processes or outcomes could study in more depth how user perceptions of attributes of conceptual modeling grammars (e.g., their ontological deficiencies) change when expertise (through training or direct application) is acquired.

In the end, we hope that our research and the directions outlined above motivate fellow researchers to contribute to the body of knowledge on conceptual modeling practice.
6. References


<table>
<thead>
<tr>
<th>Theory Construct</th>
<th>No</th>
<th>Item Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCD1</strong> (Business Rules)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCD1_1</td>
<td>BPMN does not provide sufficient symbols to represent business rules in process models.</td>
<td></td>
</tr>
<tr>
<td>PCD1_2</td>
<td>BPMN could be made more complete by adding new symbols for representing business rules in process models.</td>
<td></td>
</tr>
<tr>
<td>PCD1_3</td>
<td>I often cannot use BPMN to adequately represent business rules in process models.</td>
<td></td>
</tr>
<tr>
<td><strong>PCD2</strong> (Process structure)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCD2_1</td>
<td>BPMN does not provide sufficient symbols to represent the process structure and decomposition in process models.</td>
<td></td>
</tr>
<tr>
<td>PCD2_2</td>
<td>BPMN could be made more complete by adding new symbols for representing the process structure and decomposition in process models.</td>
<td></td>
</tr>
<tr>
<td>PCD2_3</td>
<td>I often cannot use BPMN to adequately represent the process structure and decomposition in process models.</td>
<td></td>
</tr>
<tr>
<td><strong>PCR1</strong> (Transformations)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR1_1</td>
<td>I often have to choose between a number of BPMN symbols to represent one kind of a transformation in a process model.</td>
<td></td>
</tr>
<tr>
<td>PCR1_2</td>
<td>BPMN often provides two or more symbols that can be used to represent the same kind of transformation in a process model.</td>
<td></td>
</tr>
<tr>
<td>PCR1_3</td>
<td>In a process model, one kind of a transformation can often be represented by different BPMN symbols.</td>
<td></td>
</tr>
<tr>
<td><strong>PCR2</strong> (Events)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR2_1</td>
<td>I often have to choose between a number of BPMN symbols to represent one kind of an event in a process model.</td>
<td></td>
</tr>
<tr>
<td>PCR2_2</td>
<td>BPMN often provides two or more symbols that can be used to represent the same kind of event in a process model.</td>
<td></td>
</tr>
<tr>
<td>PCR2_3</td>
<td>In a process model, one kind of an event can often be represented by different BPMN symbols.</td>
<td></td>
</tr>
<tr>
<td><strong>PCO1</strong> (Pool)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCO1_1</td>
<td>I often have to provide additional information to clarify the context in which I want to use the Pool construct in a process model.</td>
<td></td>
</tr>
<tr>
<td>PCO1_2</td>
<td>The Pool construct in BPMN can have more than one meaning in a process model.</td>
<td></td>
</tr>
<tr>
<td>PCO1_3</td>
<td>I often use the Pool construct to represent more than one real-world phenomena in a process model.</td>
<td></td>
</tr>
<tr>
<td><strong>PCO2</strong> (Lane)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCO2_1</td>
<td>I often have to provide additional information to clarify the context in which I want to use the Lane construct in a process model.</td>
<td></td>
</tr>
<tr>
<td>PCO2_2</td>
<td>The Lane construct in BPMN can have more than one meaning in a process model.</td>
<td></td>
</tr>
<tr>
<td>PCO2_3</td>
<td>I often use the Lane construct to represent more than one real-world phenomena in a process model.</td>
<td></td>
</tr>
<tr>
<td>PCE1 [Off-page connector]</td>
<td>PCE1_1</td>
<td>The Off-page Connector construct does not have any real-world meaning in a process model.</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------</td>
<td>----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>PCE1_2</td>
<td>I often cannot precisely articulate the meaning of the Off-page Connector construct in a process model.</td>
</tr>
<tr>
<td></td>
<td>PCE1_3</td>
<td>The Off-page Connector construct does not represent any relevant real-world phenomena in a process model.</td>
</tr>
<tr>
<td>PCE2 [Multiple Instances]</td>
<td>PCE2_1</td>
<td>The Multiple Instances construct does not have any real-world meaning in a process model.</td>
</tr>
<tr>
<td></td>
<td>PCE2_2</td>
<td>I often cannot precisely articulate the meaning of the Multiple Instances construct in a process model.</td>
</tr>
<tr>
<td></td>
<td>PCE2_3</td>
<td>The Multiple Instances construct does not represent any relevant real-world phenomena in a process model.</td>
</tr>
</tbody>
</table>