On the Usage of Labels and Icons in Business Process Modeling

Jan Mendling, *Humboldt-Universität zu Berlin, Germany*
Jan Recker, *Queensland University of Technology Brisbane, Australia*
Hajo A. Reijers, *Technical University Eindhoven, The Netherlands*

ABSTRACT

The value of business process models is dependent not only on the choice of graphical elements in the model, but also on their annotation with additional textual and graphical information. This research discusses the use of text and icons for labeling the graphical constructs in a process model. We use two established verb classification schemes to examine the choice of activity labels in process modeling practice. Based on our findings, we synthesize a set of twenty-five activity label categories. We propose a systematic approach for graphically representing these label categories through the use of graphical icons, such that the resulting process models are easier and more readily understandable by end users. Our findings contribute to an ongoing stream of research investigating the practice of process modeling and thereby contribute to the body of knowledge about conceptual modeling quality overall.

Keywords: Process Modeling, Labeling, Word Classification, SAP reference model, Icons

INTRODUCTION

Process modeling has emerged as a primary reason to engage in conceptual modeling (Davies, Green, Rosemann, Indulska, & Gallo, 2006) and is sought to provide benefits of process documentation, organizational transparency, and others (Indulska, Green, Recker, & Rosemann, 2009). Similar to other forms of conceptual modeling, process models are first and foremost required to be intuitive and easily understandable, especially in IS project phases concerned with requirements documentation and communication (Dehnert & van der Aalst, 2004). But even though process modeling has been around for some thirty years, surprisingly little is known about the practice of process modeling, and how process modeling can be of value to an organization (Indulska, Recker, Rosemann, & Green, 2009). Research has investigated, for instance, the graphical constructs and their meaning in process models (Rosemann, Recker, Indulska, & Green, 2006), or the expressiveness and validity of workflow aspects in process models (van der Aalst, ter Hofstede, Kiepuszewski, & Barros, 2003). Also work on quality frameworks for conceptual models in general is also available, such as the Guidelines of Modeling (GoM) (Schütte & Rotthowe, 1998) or the SEQUAL framework (Krogstie, Sindre, & Jørgensen, 2006). These and other frameworks provide sets of guidelines on how to conduct process modeling. However, the specific factors that contribute to building a "good" process model, for example one that results in human understanding, has received little attention up until now (Mendling, Reijers, & Cardoso, 2007).

Recent research has started to examine process model comprehension. For instance, the impact of structural properties of the graphical model elements on model understanding is clearly identified (Mendling, Neumann, & van der Aalst, 2007). However, it has also been shown that the choice of the graphical language used for process modeling has only insignificant effects on process model understanding (Recker & Dreiling, 2007). This situation raises the question of which aspects – other than the choice of graphical constructs and their structural layout – influence the way a process model is understood by end users.

In our work we continue along this line of work towards more understandable process models. In particular, we assert that, to date, little attention has been devoted to a very essential task in process modeling: the labeling of the graphical constructs, in particular of the constructs standing for "activities" (or "tasks", or "operations" – in other words, work to be performed) in a process model. This is surprising given that, clearly, the true meaning of any construct in a process model is only revealed when model users read – and intuitively understand – the labels assigned to a construct. Current practice indicates that the labeling of activity constructs is a rather arbitrary task in modeling initiatives and one that is sometimes done without a great deal of thought (Storey, 2005). This can undermine the understanding of the resulting models in cases where the meaning of the labels is unclear, not readily understandable or simply counter-intuitive to the reader. In prior work (Mendling, Reijers, & Recker, 2009), for instance, we found that the choice of the right convention for labeling activity constructs (e.g., a "verb-object"-convention versus an "action-noun"-convention) has a significant impact on the perceived ambiguity and perceived usefulness of the labels.

This situation indicates a demand for more sophisticated methodical support in the act of labeling activity constructs in process models. We identify two challenges in particular. First, we argue in line with other studies (e.g., Born, Dörr, & Weber, 2007; Greco, Guzzo, Pontieri, & Sacca, 2004; Mendling et al., 2009) that more support is needed to select adequate terms in the labeling of constructs. Second, research in cognitive science suggests that incorporating graphical icons in textual messages improves reader understanding (Mayer, 1989; Paivio, 1991). This work suggests that – in addition to improving the choice of textual labels, a second, complementary contribution can be made by examining the use of additional graphical icons to assist model users in understanding the (textual) labels in process models. And indeed, several modeling tools already provide mechanisms to assign an icon to an activity construct such that its meaning can be grasped faster and more intuitively. Yet, none of the tools that we are aware of deals with icons in a systematic way.

Accordingly, we address two research objectives in this paper. First, we examine the use of terms in activity labels in process models, so as to arrive at a set of meaningful terms that can be used to guide process modelers in their labeling efforts. Second, we provide a systematic approach to extend textual labels in process models with a set of graphical icons. We proceed as follows. In the next section we discuss the background to our work. We review work in cognitive science that forms the theoretical basis for our elaborations on process model understanding. We also review existing approaches to labeling as implemented in process modeling tools and the way they support the assignment of icons to activities. Next, we perform a linguistic analysis of textual labels of process models found in practice, on the basis of two verb classification schemes. Then, we suggest an approach to complement textual labels with adequate graphical icons. We conclude our paper by discussing contributions, limitations, and future research directions.

BACKGROUND

Theories of Understanding

Process models are used to convey information about organizational procedures and workflows to users, with the objective of assisting them in tasks such as workflow design (van der Aalst et al., 2003), process performance measurement (Kueng, 2000), organizational re-design (Danesh & Kock, 2005), and others.

The Dual Coding Theory (Paivio, 1991) suggests that individuals have two separate channels (visual and auditory) that they use when processing information that is, for instance, embodied in a process model. The two channels complement each other such that simultaneously receiving information through each channel improves understanding compared to receiving information through one channel only. In other words, individuals understand informational material better when it is provided through both auditory (i.e., words) and visual (i.e., images) channels. Indeed, people tend to read by speaking out the words of the text in their mind, which even suppresses visual activation (Brooks, 1967). Therefore, the textual activity labels in process models are cognitively processed via the auditory channel.

Based on this observation, the Cognitive Theory of Multimedia Learning (CTML, Mayer, 1989) suggests that learning material intended to be received, understood and retained by its recipients should be presented using both words and pictures. This argument applies directly to the task of process modeling, where both visual (graphical constructs) and auditory (labels and text annotations) material are available to convey information about a business process in a process model. This argument suggests that, to warrant process model understanding, the choice of graphical constructs (e.g., events, activities, business rules, decisions, messages, and so forth) and the choice of appropriate labels for these constructs (e.g., "perform assessment", "message received", and so forth) are both of importance.

Due to the overall limited number of graphical constructs used in a process model (there are typically few if not only one graphical construct for representing "tasks" or "activities"), most of the critical domain information is typically contained in the textual labels of the constructs – in other words, in auditory channels. Indeed, zur Muehlen and Recker (2008) showed that in process modeling practice, users often use a very limited set of graphical constructs in their process modeling, and frequently use textual annotations to convey the information they need to have articulated in the model.

Based on these observations, we thus argue that process model understanding can be increased if additional graphical images (such as icons) are added to the labeling of process model constructs, if performed in a meaningful and systematic way. Similar work has, for instance, been carried out in the data modeling domain (Masri, Parker, & Gemino, 2008). Work in the object-oriented modeling domain (Moody & van Hillegersberg, 2008) and in the process modeling domain (e.g., Laue & Mendling, 2010; Recker, Rosemann, Indulska, & Green, 2009), however, has to date mostly focused on visual and structural aspects without taking into consideration textual labeling considerations. Our contention is now that graphical icons, if meaningfully and systematically complementary to the choice of textual labels in process modeling, could potentially significantly improve process model understanding. To that end, we investigate the current state of process modeling practice, to assert to what extent current use of icons is linked to the (textual) specification of a process modeling construct, if at all.

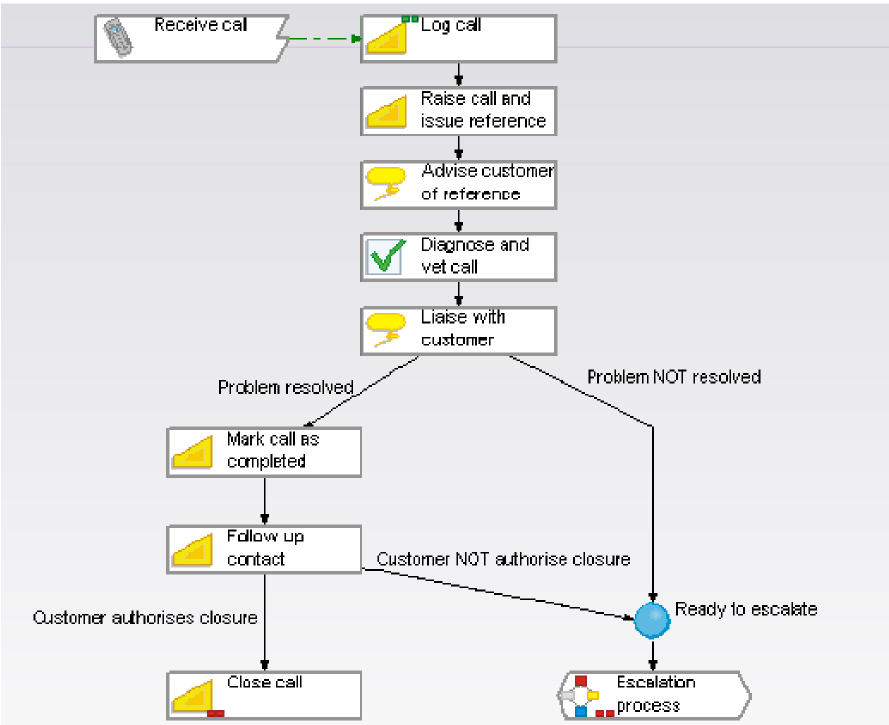
Labels and Icons in Commercial Process Modeling

The labeling of graphical process model constructs, such as activities, is often more art than science. In practice, a number of guidelines exist that typically suggest a verb-object convention (e.g., "approve order", "verify invoice") for labeling activities (Malone, Crowston, & Herman, 2003; Miles, 1961). Similarly, research established a perceived superiority of the verb-object convention over other modeling conventions (Mendling et al., 2009).

Intuitively, one may presume that the more information contained in the labels, the clearer the meaning is to the reader, which would argue for long and elaborate textual labels. Recent research, however, uncovered that shorter activity labels improve model understanding (Mendling & Strembeck, 2008). This observation, in combination with the prevalent verb-object convention, would suggest that the verb term that signifies the action to be performed as part of the process activity should be specified in both a precise and concise manner, to warrant intuitive and easy understanding. Following CTML (Mayer, 1989) we can further expect that the understanding of these terms could be supported even more so through the use of additional graphical information, for instance, through the use of an icon that matches the semantics of the activity.

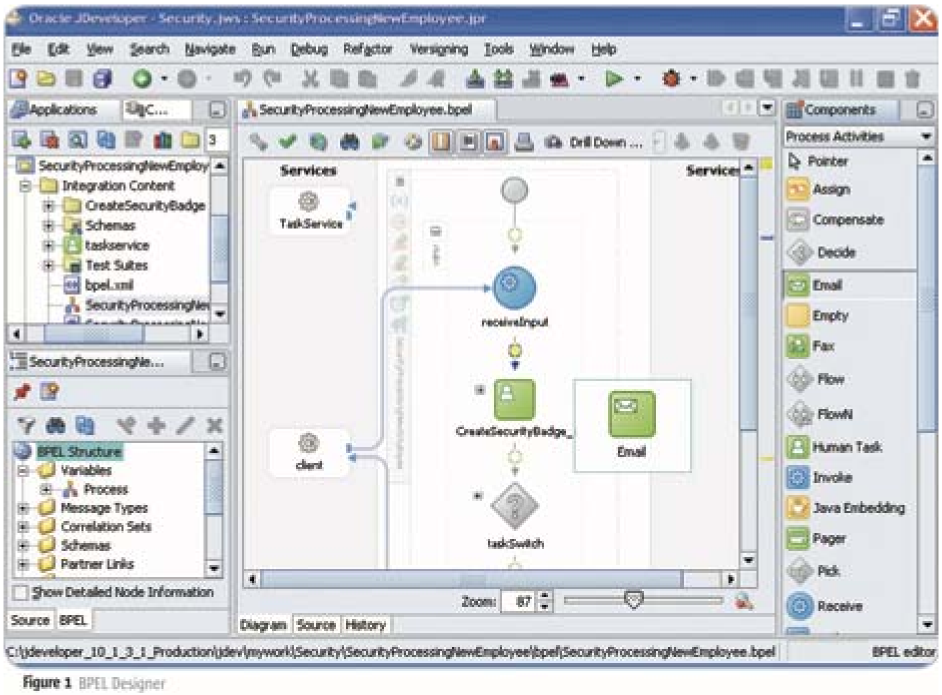
And indeed, some modeling tools already allow for the embedding of additional graphics. Protos (http://www.pallas-athena.com), for instance, allows for selecting specific types of activities (e.g., Basic, Logistics, Authorize, Communication and Check) and represents these different types by means of different images. So, when one sets the activity in a Protos model to be a Communication activity, the image becomes a 'talk balloon'. An example is shown in Figure 1. Similarly, ARIS in its Version 7.02 (http://www.ids-scheer.com) allows users to right-click on activities in BPMN diagrams to select one of several pre-defined graphical markers to distinguish basic tasks from automated transactions, sub-processes or ad-hoc processes.

Figure 1. Protos Screenshot



Similar to these examples, Intalio's BPMS Version 5 (http://bpms.intalio.com) graphically distinguishes manual from automated tasks while - as shown in Figure 2 - Oracle's BPEL Process Manager (http://www.oracle.com) offers graphical icons to distinguish, for instance, “invoke” from ”receive” activities.

Figure 2. Oracle BPEL Process Manager Screenshot



However, while the graphical or iconic annotations in current process modeling tools do provide some graphical differentiation, they do not actually visualize its intended domain-specific and contextual meaning. Our argument, however, would be that a suitable differentiation scheme – and graphical representation – of different types of relevant domain semantics of process model constructs would be of benefit to process model readers. A first step towards meeting this objective is accordingly to identify a suitable differentiation scheme to distinguish different relevant semantics of process modeling constructs.

THEORETICAL ANALYSIS OF ACTIVITY LABELS IN PROCESS MODELING PRACTICE

A major challenge during process modeling is to identify which construct to use in the creation of a process model—that is, to identify, which aspects of a domain should be represented as an event, an activity, a message, an actor and so forth. Useful guidelines have emerged for choosing appropriate sets of graphical construct for representation (e.g., Recker et al., 2009; Rosemann et al., 2006; van der Aalst et al., 2003; zur Muehlen & Recker, 2008) and for choosing appropriate model structures (e.g., Laue & Mendling, 2010; Reijers & Mendling, 2008). The terms used to label these constructs, however, are usually selected by the model designer, sometimes without a great deal of thought given to finding the one that best reflects the semantics of the business domain (Mendling et al., 2009; Storey, 2005). This can lead to problems, especially when interpreting or comparing process models.

We argue that having some mechanism to compare the semantics of different activity labels – in particular the *verb phrases* used in activity labels – would be useful to process model application areas, for instance when comparing different process models during process integration, process benchmarking, process re-design or process documentation. It would also be valuable for comparing different designs of a process models (or template) to facilitate process model re-use.

Given that a process model, in essence, represents a conceptualization, or simplified view, of a real-world business domain (Recker et al., 2009), a taxonomy could be developed that would facilitate the sharing of activity label terms amongst different models, or model designers. This taxonomy could then be used, for instance, for process model comparison (Ehrig, Koschmider, & Oberweis, 2007) or verification (Wynn, Verbeek, Van der Aalst, ter Hofstede, & Edmond, 2009). Even if fully automated techniques for doing such tasks are unlikely, a taxonomy can still assist users in such tasks.

Our line of work follows the example given by Storey (2005) in the data modeling domain. She developed an ontology for the semantic classification of relationship-type constructs in data models based on dictionaries, business taxonomies and previous research. Similarly, we will use established taxonomies of knowledge in the business process domain to differentiate different categories of terms used to annotate activity constructs in process models with textual information. We focus on activity constructs, and the labels thereof, because of the centrality of the 'activity' concept in process modeling, over and above other notions such as 'event', 'message' or 'orchestration' (Mendling et al., 2009).

Domain semantics define the real-world meaning, or essence thereof, of the terms used in any conceptual model, that is, of words and phrases used to label constructs (Storey, 2005). The tricky part is that some of these semantics are well-known and unambiguous while others may vary with context, i.e., they can be subject to multiple interpretations. Accordingly, it would appear logical to develop a verb classification scheme based on the business process management context, in which process models are used.

We turn to two verb classification schemes: the MIT Process Handbook (Malone et al., 2003) and the Verb Classes proposed by Levin (1993). Both verb classification schemes are well-known libraries for sharing and managing knowledge, and the MIT Process Handbook focuses specifically on business processes and organizations.

We apply both schemes in the classification of the verbs used in the activity labels of the SAP Reference Model (Keller & Teufel, 1998). The SAP reference model contains overall 604 process models capturing various information about the SAP R/3 functionality to support the business processes in an organization. The SAP reference model denotes a frequently used tool in the implementation of SAP systems (Daneva, 2004), and much literature covers its development and use (Curran, Keller, & Ladd, 1997; Keller & Meinhardt, 1994; Keller & Teufel, 1998). With the SAP solution being the market leading tool in the Enterprise Systems market we feel that the examination of SAP process models gives us a good understanding of the use of process models in real-life business contexts. We extracted 19,839 activity labels from the 604 process models for our analysis. 4,553 of these labels are unique and they refer to 309 different verbs (including corresponding gerunds or nouns). Table 1 lists the 30 most frequently used verbs of the SAP Reference Model. Clearly, some verbs are semantically overlapping like “to determine” and “to check”. The following sections discuss how verb classifications are suited to resolve these overlaps.

Table 1: The 30 most frequently used verbs in the SAP Reference Model

Verbs 1-10 (occurrences in the SAP reference model)	**Verbs 11-20 (occurrences in the SAP reference model)**	**Verbs 21-30 (occurrences in the SAP reference model)**
to process (2003)	to post (330)	to update (203)
to enter (1922)	to release (328)	to analyse (191)
to determine (1755)	to maintain (316)	to settle (186)
to check (971)	to calculate (271)	to allocate (180)
to create (665)	to assign (261)	to transmit (171)
to plan (614)	to define (258)	to copy (164)
to transfer (510)	to edit (258)	to print (162)
to select (349)	to perform (228)	to generate (141)
to confirm (345)	to specify (226)	to change (136)
to carry out (337)	to evaluate (203)	to display (131)

Using the MIT Process Handbook

The MIT Process Handbook Project started in 1991 with the aim to establish an online library for sharing knowledge about business processes that is freely available to the general public under a form of "open source” licensing (Malone et al., 2003). The Process Handbook includes entries for over 5000 business activities, together with an extensive set of software tools for viewing and modifying the knowledge base.

The business processes in the MIT Process Handbook library are organized hierarchically to facilitate an easy navigation. The hierarchy builds on an inheritance relationship between verbs that refer to the represented business activity. A list of eight generic verbs including “create”, “modify”, “preserve”, “destroy”, “combine”, ”separate”, ”decide”, and “manage” have been identified using the lexical database WordNet (Miller, 1995), an online library containing over 21,000 verb word forms divided into 15 semantic files. We used WordNet to build a list of all synonyms of the eight generic verbs. Using them we linked the verbs of the SAP Reference Model to the verb classes of the MIT Process Handbook. Since some lexical verbs have common synonyms, verbs are potentially related to more than one class. Table 2 shows instantiations of the eight generic verb classes of the MIT Process Handbook, and provides examples found in activity labels in the SAP reference model.

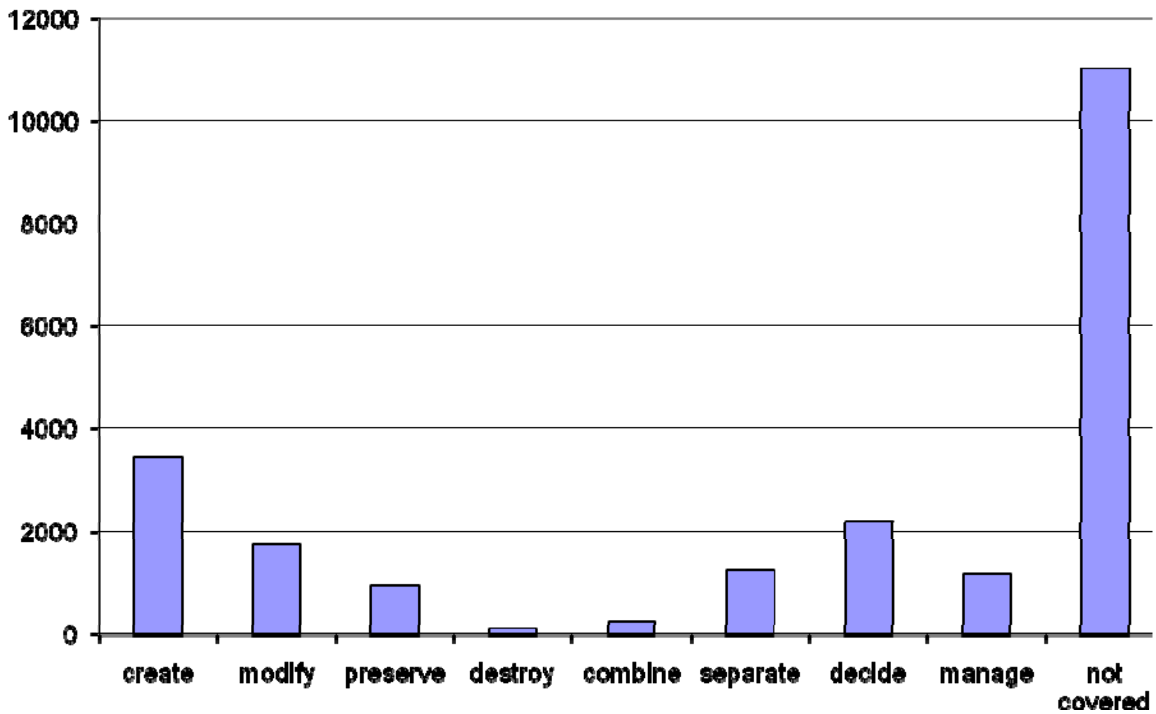
Table 2: Interpretation of generic MIT Process Handbook terms, and examples from the SAP data set.

Verb Class	**Synonyms**	**Examples from the SAP reference model**
to create	to build	Automatic load building
to modify	to change	Change loan contract
to preserve	to store	Shipping unit storage
to destroy	to eliminate	Elimination of IC Profit and Loss in Current Assets
to combine	to include	Include third person in call
to separate	to break	Break determination
to decide	to determine	Determine variance tolerance
to manage	to negotiate	Work Contract Negotiation

Figure 3 shows the final results from our classification effort. The eight generic verbs of the MIT Process Handbook cover 21,046 verb occurrences (some classified in multiple categories) while 11,029 could not be automatically classified using the synonyms of WordNet (note that verbs can be assigned to more than one class). Yet, the eight generic verbs of the MIT Process Handbook occur at least 122 and up to 3,455 times in our data sample. Hence, we can conclude that the eight generic verbs provide a substantial part of the spectrum of activity labels found in the SAP reference model data set.

In performing the classification, we encountered two types of problems. First, there are several terms that are too technical to be covered by the synonyms of WordNet. Examples of such terms include “to dun” (15 occurrences) and “to accrue” (16 occurrences). Second, there are some verbs that are not covered at all, although they are arguably synonyms of one or many of the generic verbs provided. A prominent example is the verb “to process”, which in fact is the most frequently used term in the SAP set of models (2003 occurrences). Even though it shares a number of characteristics with, for instance, “to modify”, this relationship is not documented in WordNet.

Figure 3. Generic verbs of the MIT Process Handbook and Occurrences in the SAP Reference Model



Using Levin's Word Classes

The systematic work on verb classes by Levin is an important contribution to understanding the use of languages. It defines 49 semantic classes of verbs and categorizes more than 3,000 English verbs (Levin, 1993). In contrast to the MIT Process Handbook that builds upon WordNet, Levin's verb classes are derived from a linguistic analysis of 'English in use'. Each verb class is divided into sub-classes that list some prominent example verbs. In Levin's work, verbs that display the same or similar set of diathesis alternations are assumed to share a similar semantic meaning and can thus be organized into a semantically coherent verb class. For instance, the class "Break verbs" (class 45.1) refers to actions that bring about a change in the material integrity of some entity.

We used an online version of this classification hierarchy for our analysis (see http://www-personal.umich.edu/~jlawler/levin.html). Again, multiple assignments between a verb and a class were possible. Selected examples are given in Table 3.

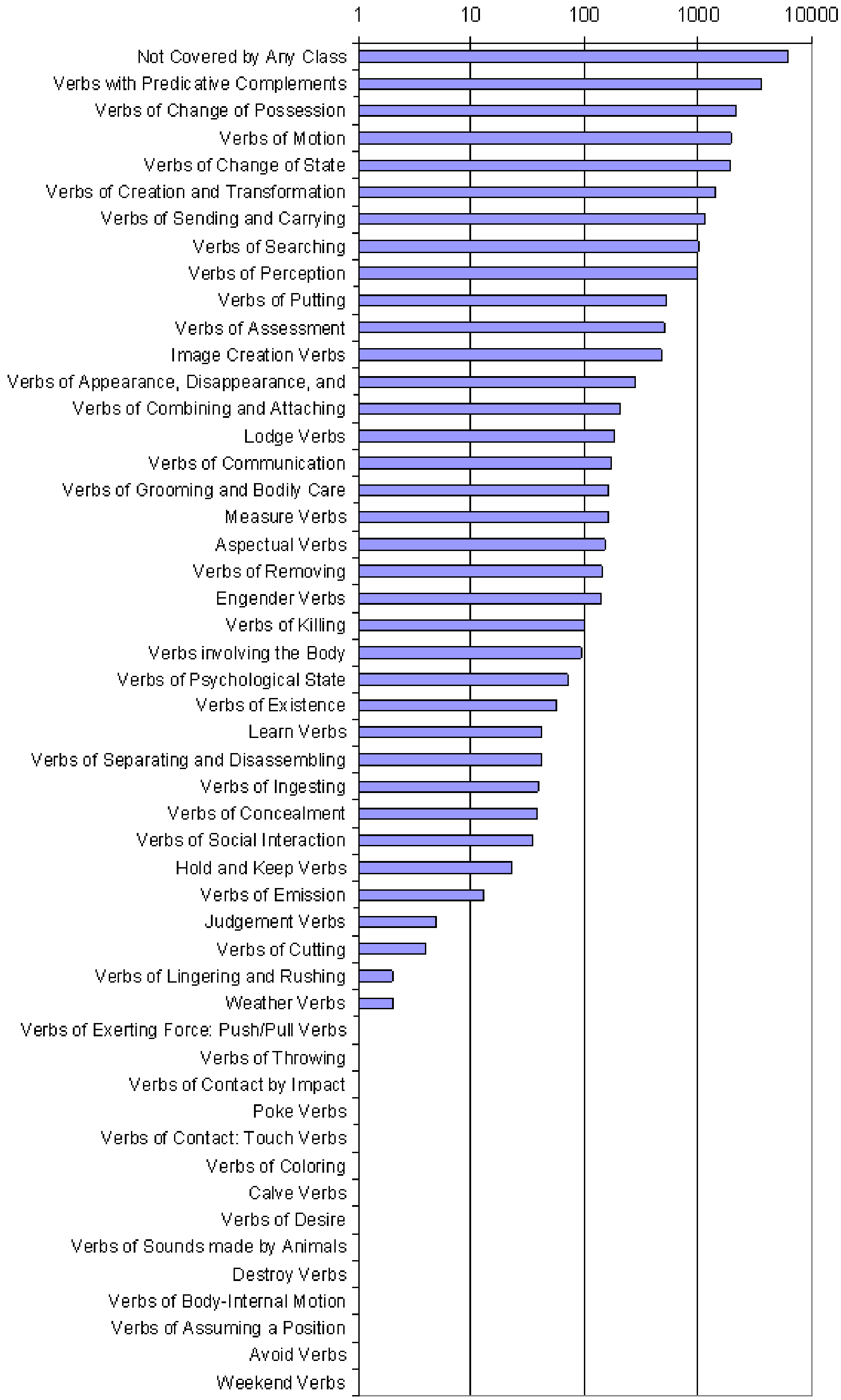
Table 3: Interpretation of selected generic terms from Levin’s word classes, and examples from the data set.

Verb Class	**Synonyms**	**Examples from the SAP reference model**
Verbs with predicative complement	to accept	Accept registration order of business event attendee
Verbs of perception	to check	Check and Assign Dummy Commitment Items
Verbs of appearance and disappearance	to define	Define license type
Verbs of communication	to announce	Announce Official Project Start
aspectual verbs	to begin	Begin warranty check
Verbs involving the body	to close	Closing of Insurance Contract
Verbs of psychological state	to approve	Approve profit plan

Figure 4 shows the final classification results using Levin's verb classes. Altogether 17,868 occurrences could be categorized in one or more classes, 6,232 were not covered. The most frequently occurring verb class relates to verbs with predicative complement. These are basically speech acts (Auramäki, Lehtinen, & Lyytinen, 1988) such as “to accept”, “to acknowledge”, or “to select”. This could be seen as an indication for a focus of business process on inter-personal communication and decision-making scenarios.

We faced classification problems similar to our usage of the MIT Process Handbook: some verbs were too specific, others not covered. Interestingly, again the verb “to process” was not included in any class. We also faced homonym problems. For instance, the verb “to change” appears in different classes, among others in the class “verbs of grooming and bodily care”. Although this aspect is not touched in the SAP Reference Model, our automatic classification counts the verb also for this class, as shown in Figure 4.

Figure 4. Verb Classes by Levin and Occurrences in the SAP Reference Model



Synthesis

We extracted 19,839 activity construct labels from the SAP reference models, from which some terms were classified in multiple verb classes. We can assess the coverage of the two classification schemes as (1 – notCovered) / 19,839. Accordingly, the coverage of the MIT Handbook is (1 – 11,029) / 19,839 = 0.44 while the Levin classes cover (1 – 6,232) / 19,839 = 0.68. This would indicate a preference for Levin's work.

In both categorization schemes, a fraction of activity labels could not be covered. In most cases, unspecific terms such as “to dun” or “to process” were such candidates. These and similar verbs are typically prone to different types of ambiguity, such as action-object ambiguity, verb-inference ambiguity, or zero-derivation ambiguity (Mendling et al., 2009). These types of ambiguity stem from the ambiguity problem of the English language in general (Small, Cottrell, & Tanenhaus, 1988). We further note current conventions for process modeling (e.g., Davis, 2001; Rosemann, 2003) typically recommend to avoid unspecific verbs such as “to process.” The fact that our findings indicate the frequent usage of such verbs indicates that such conventions are being violated in practice, which suggests a need for a more rigorous enforcement of modeling conventions. The work by Becker et al. (2009) could be potential avenue enforcing labeling conventions through tool support.

To achieve a better coverage ratio, we assume that a most suitable categorization of activity verb terms can be obtained when combining both classification schemes considered. More specifically, based on our initial analysis we assume that in particular those verb classes with more than 100 noted occurrences can be viewed suitable candidates for consideration in a taxonomy. Table 4 shows the 25 resulting generic verbs we consider based on this assumption, and gives examples for actual SAP reference model labels. Given the overlap between the two considered schemes (e.g., there are several Levin classes that are subcategories of the MIT generic verb “to modify”), we decided to consider the more specific terms (e.g., verbs such as “to create” and “to transform” instead of “to modify”). Note that we skipped the Levin term “change state” since this is essentially the definition of exactly what an activity of a process sets out to do. Overall, the mapping of activity labels from the SAP reference model to the combined classification scheme shown in Table 4 resulted in coverage of 95 per cent, which we deem acceptable and confirmatory for the validity of our analysis.

Based on this analysis, we were able to specify a set of 25 generic classes of verbs that can typically be found in activity labels in process models. This thesaurus can be of use, not only in the act of comparing process models (Ehrig et al., 2007) or verifying process models (Wynn et al., 2009), but also in the act of creating process models. For instance, the thesaurus can be applied to the design of organizational modeling conventions (Rosemann, 2003) to structure, guide and govern process modeling initiatives by providing a set of guidelines for the consistent, standardized creation of process models. Process model designers can be assisted in their effort to create meaningful and understandable models, for instance, through the use of a tool-based modeling wizard that offers the different verb classes to select a most appropriate activity type, and then displays a set of potential synonyms for further specification. For the act of reading or interpreting process models, we could envisage the developed set of generic classes useful for assisting human understanding of process models, by displaying appropriate graphical information that assists differentiating task types in a process models. The next section reports on such an attempt.

Table 4: A set of 25 generic verbs for describing activities in business process models, with examples from the SAP reference model in brackets

to appear (Open Commitment Document Selection)	*to complete* (Complete order technically)	*to engender* (Define availability of material component (lead time/lead-time offset))	*to move* (Goods Movements)	*to remove* (Remove material from stock)
to assess (Confirm materials used)	*to create* (Create planned order)	*to lodge* (Allocate invoice)	*to perceive* (Identify certificate recipient)	*to search* (Determine validity period of BOM)
to care (Child Care)	*to decide* (Decide if budget values should be accumulated)	*to manage* (Plan total activity input quantity)	*to preserve* (Maintain units of measure / dimensions)	*to send* (Send purchasing document)
to combine (Combine deliveries into a wave)	*to destroy* (Delete Without Archiving)	*to measure* (Product Cost Analysis)	*to promise* (Guarantee Net Amount)	*to separate* (Split costs to activity types)
to communicate (Request work list)	*to display* (Display reference object data)	*to modify* (Update quality level (dynamic modification data)	*to put* (Set the Deletion Flag on Individual Objects or for the Entire Project)	*to transform* (Change employee status)

ON THE USE OF ICONS

Our previous discussion revealed that indeed we are able to reduce the set of activity labels used in process models to a restricted set of semantically different task or activity terms.

On the basis of these findings we argue that a suitable strategy for making process models more understandable is to develop iconic representations for the different identified verb classes. This would allow model users to intuitively identify – by mere visual inspection – the most common classes of activities contained in any process. For instance, a model reader could instantly identify in any given model how often (s)he is required to communicate with other stakeholders and how often a process object needs to be modified etc. Detailed information about the exact type of activity (e.g., what form of communication, what type of process object, what type of modification and so forth) can then be obtained from the textual label of the construct.

We argue that the use of iconic representations is conducive to improving process model understanding even more so, because graphical icons essentially have become part of our daily lives (think of the hourglass in Windows, the telephone symbol in Skype or the use of emoticons in text messages).

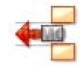
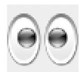
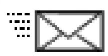
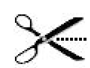
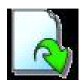
Accordingly, our endeavor was also to investigate the development of suitable iconic representations for the identified verb classes. Unfortunately, similarly to activity labeling in process modeling, icon development has been more of an art than science (Chen, 2003). Yet, some guidelines based on research in graphical user interface design exist to support our undertaking (Caplin, 2001; Horton, 1994):

1. Semantics-oriented: Icon selection should emphasize the easiness of interpretation by the users (icons should be natural to users), resemblance (to the things or tasks it refers to) and differentiation (all icons should be easily differentiated from each other and should not be subject to mis-interpretation).
2. User-oriented: Icon selection should be based on user preferences and extensive user testing.
3. Composition principle: Icon composition rules should be natural and easy to understand and learn. The Multiple-level icon composition principle, for instance, suggests rules for composing high-level icons from low-level icons based on similar concepts used in data/system modeling and the English grammar (Chen, 1983). The grouping principle, on the other hand, provides some rules to design icons in groups based on the type and instance concepts found in data/system modeling and icon-based natural languages (Chen, 1997).
4. Interpretation rules: Icon composition rules should be transferable to different models and audiences.

We have attempted to adhere to these principles in our effort to provide icons for the most frequently used generic verb classes. For instance, we scanned Google Images for icons that conformed to the principles of resemblance, differentiation, and easiness of interpretation, following the use of iconic images in popular applications such as Facebook, Twitter or Skype. Our icons also conform to the interpretation rule principle, in that the icons can be used for different models, and different model audiences. At this stage of the research, however, we were not fully able to adhere to the principles of user orientation and composition. While we based the design of our icons on some user preferences (those of the research team), we cannot claim that user preference examination or extensive user testing has been conducted yet. Specifically, we note that our icons may require further investigation in regard to interpretation ambiguity. We note this as an interesting future research avenue that could follow the research presented in this paper. Similarly, it would be interesting to distill detailed composition principles for icon design that could follow principles of process (Recker et al., 2009) or systems decomposition (Burton-Jones & Meso, 2006).

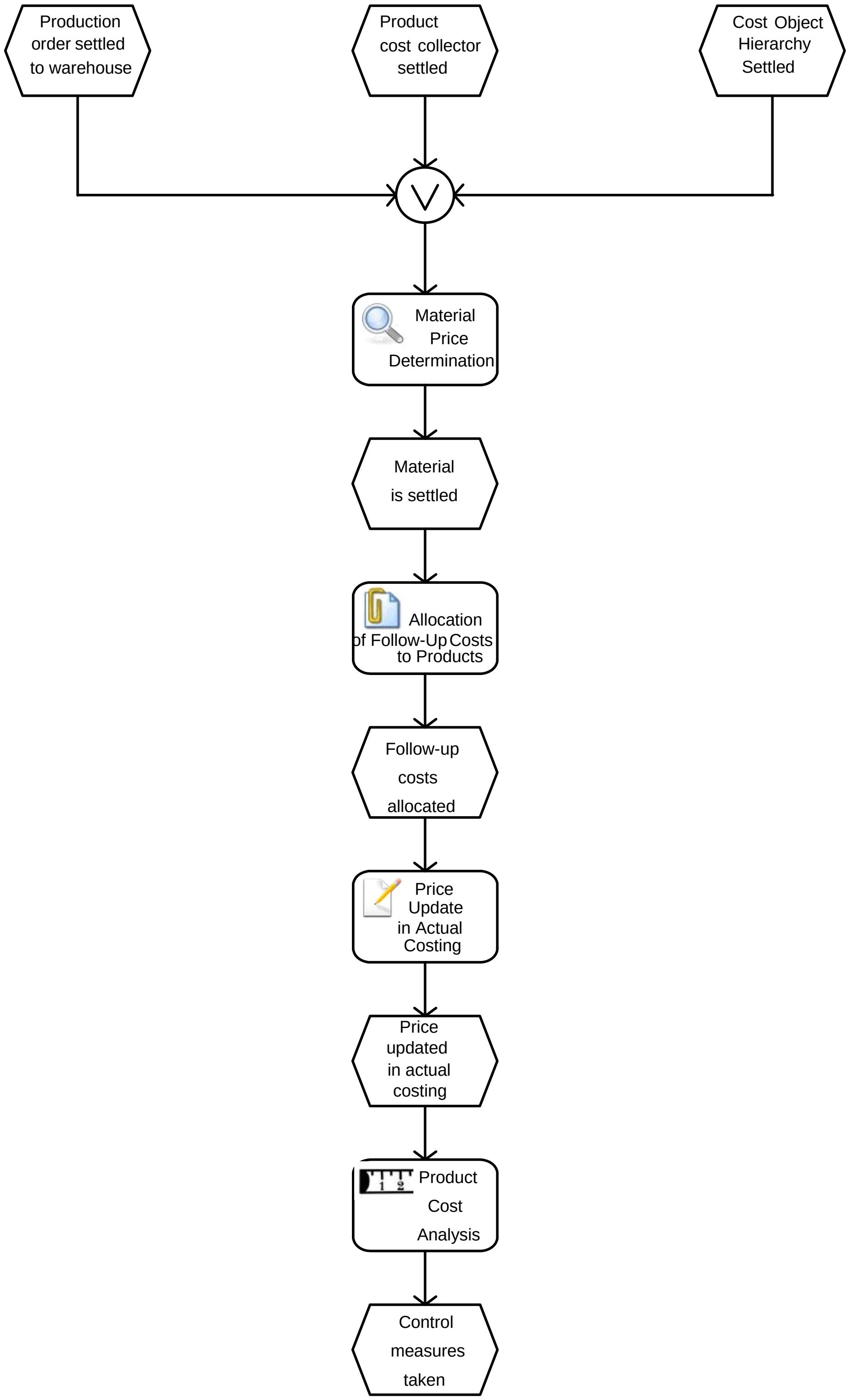
Table 5 shows the results from our design effort. We should note that our attempt to provide suitable iconic representations can still be subject to revision and enhancement. We are aware that the selected iconic representations may not necessarily be conducive to all type of modeling scenarios, or model end user audiences. This, however, is a question of an empirical nature and thus requires further research in the form of empirical testing. This is a noted future research direction.

Table 5: Iconic representations for the 25 identified generic activity label verb classes

 To appear	 To complete	 To engender	 To move	 To remove
 To assess	 To create	 To lodge	 To perceive	 To search
 To care	 To decide	 To manage	 To preserve	 To send
 To combine	 To destroy	 To measure	 To promise	 To separate
 To communicate	 To display	 To modify	 To put	 To transform

As an application example, we consider the Period-End Closing: Material Ledger process from the SAP module Revenue and Cost Controlling. In this process, four tasks are specified, which are described with the verbs “to determine”, “to allocate”, “to update“, and “to analyze”. As per the classification scheme we used, these verbs are instantiations of the four generic verb classes “to search”, “to lodge”, “to modify” and “to measure”. Accordingly, Figure 5 shows the process model annotated with icons for these generic verb classes to illustrate our approach.

Figure 5. Example Process Model with Additional Icons



CONCLUSIONS

In this paper we discussed an essential yet under-researched aspect of process modeling practice: the labeling of graphical constructs in a process model. This complements the existing streams of research investigating other dimensions of process modeling (e.g., the data, resource, or control-flow perspectives). Our line of research is based on the assumption that process model understanding can be improved if a more systematic way of labeling constructs can be found. First, through linguistic analysis we used two verb classification schemes to identify from a sample of over 3,000 process models the twenty-five most frequently used verb classes, and their synonyms. Second, based on Dual Coding Theory and CTML we argued that understanding can be improved if labels in process models can be complemented by iconic graphical representations. Based on the results from our linguistic analysis we provided a first set of iconic representations to match the identified verb classes.

We do not consider our research complete. For instance, even though we examined a considerably large number of process models, we only considered process models of the SAP reference model. This may limit the extent to which our results can be generalized. In particular, we realize that the task labels contained in the SAP reference may not be representative of all potential process modeling domains. However, the way we described the conduct of our analyses will allow researchers to replicate our study in other contexts, e.g., using different model sets, or different process modeling notations such as BPMN or UML Activity Diagrams. Also, we investigated the choice of verb terms in activity labels. Yet, in the prevalent 'verb-object' labeling style, also the word items used for object terms are of importance to understanding the process model produced. To that end, some experts recommend a preparatory step called 'technical term modeling' before modeling the actual process (Rosemann, 2003). Technical term models capture the key entities and object involved in a business process, and delineate their hierarchy and semantic relationships. Often, entity-relationship diagrams or class diagrams are employed as a relevant modeling language by practitioners. Some niche tools like Semtalk already support an integration of ontologies and process models for this purpose (Fillies, Wood-Albrecht, & Weichardt, 2003), but major tools such as ARIS or Telelogic System Architect do not.

Our forthcoming research is as follows. We will examine empirically the suitability of verb classification schemes for classifying activity tasks in process models. Similar to the experiment described by Storey (2005), we will have students classify activity tasks in a number of process models as per the verb classification schemes to establish the viability of these schemes. This will involve other reference models such as the Supply Chain Operations Reference Model (SCOR). After identifying the most common verb classes used in process modeling, we will further develop the set of iconic representations for these verb classes. Usability studies on different icon alternatives, potentially involving users with different cultural backgrounds, are further important tasks. There is work in the visual programming (Green & Petre, 1996) and software development discipline (Holzinger, 2005) that could guide such research, and we would like to invite fellow scholars to contribute to this task. In another stream of research we will then investigate empirically whether the inclusion of words (labels) and images (icons) in process models does in fact warrant improved model understandability. CTML suggests three outcomes of understanding – retention, recall and transfer – that can be used as measures in a related empirical study. In conducting such a study we can refer to the works of Gemino and Wand

(2005) and Recker and Dreiling (2007) that both used exactly these measures for examining understanding generated through data and process modeling, respectively. This study is currently underway, and we invite interested readers to take part, by contacting the authors.

In a different stream of research, we are investigating to what extent tools and techniques from natural language processing can be applied for achieving a better label quality. For instance, the frequent occurrence of *to process* in the SAP Reference Model points to a quality issue as it is a very unspecific verb. We are currently testing capabilities of WordNet for defining quality metrics for specificity and potential ambiguity. One potential way of avoiding ambiguity in labels is by selecting a labeling style that is the least susceptible to ambiguity problems. Mendling et al. (2009) identified the verb-object convention to be such a labeling style. Combining the style of labeling with the work in this paper on the different semantic categories of activity labels could potentially further decrease ambiguity and specificity problems in process modeling. We hope that the work presented in this paper inspires fellow researchers to distill a set of normative guidelines for the act of process modeling from our work and that of our colleagues.

REFERENCES

Auramäki, E., Lehtinen, E., & Lyytinen, K. (1988). A Speech-act-based Office Modeling Approach. *ACM Transactions on Information Systems, 6*(2), 126-152.

Becker, J., Delfmann, P., Herwig, S., Lis, L., & Stein, A. (2009). *Towards Increased Comparability of Conceptual Models - Enforcing Naming Conventions through Domain Thesauri and Linguistic Grammars.* Paper presented at the 17th European Conference on Information Systems, Verona, Italy.

Born, M., Dörr, F., & Weber, I. (2007). User-Friendly Semantic Annotation in Business Process Modeling In M. Weske, M.-S. Hacid & C. Godart (Eds.), *Web Information Systems Engineering – WISE 2007 Workshops* (Vol. 4832, pp. 260-271). Nancy, France: Springer.

Brooks, L. R. (1967). The Suppression of Visualization by Reading. *The Quarterly Journal of Experimental Psychology, 19*(4), 289-299.

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.

Caplin, S. (2001). *Icon Design: Graphic Icons in Computer Interface Design*. London, England: Watson-Guptill.

Chen, P. P.-S. (1983). English Sentence Structures and Entity-Relationship Diagrams. *Information Sciences, 29*(2-3), 127-149.

Chen, P. P.-S. (1997). English, Chinese and ER diagrams. *Data & Knowledge Engineering, 23*(1), 5-16.

Chen, P. P.-S. (2003). *Toward a Methodology of Graphical Icon Design.* Paper presented at the 5th International Symposium on Multimedia Software Engineering, Taichun, Taiwan.

Curran, T., Keller, G., & Ladd, A. (1997). *SAP R/3 Business Blueprint: Understanding the Business Process Reference Model*. Upper Saddle River, New Jersey: Prentice Hall.

Danesh, A., & Kock, N. (2005). An Experimental Study of Two Process Representation Approaches and their Impact on Perceived Modeling Quality and Redesign Success. *Business Process Management Journal, 11*(6), 724-735.

Daneva, M. (2004). ERP Requirements Engineering Practice: Lessons Learned. *IEEE Software, 21*(2), 26-33.

Davies, I., Green, P., Rosemann, M., Indulska, M., & Gallo, S. (2006). How do Practitioners Use Conceptual Modeling in Practice? *Data & Knowledge Engineering, 58*(3), 358-380.

Davis, R. B. (2001). *Business Process Modelling with ARIS: A Practical Guide*. London, England: Springer.

Dehnert, J., & van der Aalst, W. M. P. (2004). Bridging The Gap Between Business Models And Workflow Specifications. *International Journal of Cooperative Information Systems, 13*(3), 289-332.

Ehrig, M., Koschmider, A., & Oberweis, A. (2007). Measuring Similarity between Semantic Business Process Models. In J. F. Roddick & A. Hinze (Eds.), *Conceptual Modelling 2007* (Vol. 67, pp. 71-80). Ballarat, Australia: Australian Computer Science Communications.

Fillies, C., Wood-Albrecht, G., & Weichardt, F. (2003). Pragmatic Applications of the Semantic Web Using SemTalk. *Computer Networks, 42*(5), 599-615.

Gemino, A., & Wand, Y. (2005). Complexity and Clarity in Conceptual Modeling: Comparison of Mandatory and Optional Properties. *Data & Knowledge Engineering, 55*(3), 301-326.

Greco, G., Guzzo, A., Pontieri, L., & Sacca, D. (2004). An Ontology-Driven Process Modelling Framework. In F. Galindo, M. Takizawa & R. Traunmüller (Eds.), *Database and Expert Systems Applications - DEXA 2004* (Vol. 3180, pp. 13-23). Zaragoza, Spain: Springer.

Green, T. R. G., & Petre, M. (1996). Usability Analysis of Visual Programming Environments: A 'Cognitive Dimensions' Framework. *Journal of Visual Languages and Computing, 7*(2), 131-174.

Holzinger, A. (2005). Usability Engineering Methods for Software Developers. *Communications of the ACM, 48*(1), 71-74.

Horton, W. (1994). *The Icon Book: Visual Symbols for Computer Systems and Documentation*. New York, New York: John Wiley & Sons.

Indulska, M., Green, P., Recker, J., & Rosemann, M. (2009). Business Process Modeling: Perceived Benefits. In S. Castano, U. Dayal & A. H. F. Laender (Eds.), *Conceptual Modeling - ER 2009*. Gramado, Brazil: Springer.

Indulska, M., Recker, J., Rosemann, M., & Green, P. (2009). Process Modeling: Current Issues and Future Challenges. In P. van Eck, J. Gordijn & R. Wieringa (Eds.), *Advanced Information Systems Engineering - CAiSE 2009* (Vol. 5565, pp. 501-514). Amsterdam, The Netherlands: Springer.

Keller, G., & Meinhardt, S. (1994). *SAP R/3 Analyzer: Business Process Reengineering based on the R/3 Reference Model*. Philadelphia, Pennsylvania: SAP North-America.

Keller, G., & Teufel, T. (1998). *SAP R/3 Process Oriented Implementation: Iterative Process Prototyping*. Boston, Massachusetts: Addison-Wesley.

Krogstie, J., Sindre, G., & Jørgensen, H. D. (2006). Process Models Representing Knowledge for Action: a Revised Quality Framework. *European Journal of Information Systems, 15*(1), 91-102.

Kueng, P. (2000). Process Performance Measurement System: A Tool to Support Process-based Organizations. *Total Quality Management, 11*(1), 67-85.

Laue, R., & Mendling, J. (2010). Structuredness and its Significance for Correctness of Process Models. *Information Systems and e-Business Management, 9*(In Press).

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, Illinois: University of Chicago Press.

Malone, T. W., Crowston, K., & Herman, G. A. (2003). *Organizing Business Knowledge: The MIT Process Handbook*. Cambridge, Massachusetts: The MIT Press.

Masri, K., Parker, D. C., & Gemino, A. (2008). Using Iconic Graphics in Entity-Relationship Diagrams: The Impact on Understanding. *Journal of Database Management, 19*(3), 22-41.

Mayer, R. E. (1989). Models for Understanding. *Review of Educational Research, 59*(1), 43-64.

Mendling, J., Neumann, G., & van der Aalst, W. M. P. (2007). Understanding the Occurrence of Errors in Process Models based on Metrics. In R. Meersman & Z. Tari (Eds.), *On the Move to Meaningful Internet Systems 2007* (Vol. 4803, pp. 113-130). Vilamoura, Portugal: Springer.

Mendling, J., Reijers, H. A., & Cardoso, J. (2007). What Makes Process Models Understandable? In G. Alonso, P. Dadam & M. Rosemann (Eds.), *Business Process Management - BPM 2007* (Vol. 4714, pp. 48-63). Brisbane, Australia: Springer.

Mendling, J., Reijers, H. A., & Recker, J. (2009). Activity Labeling in Process Modeling: Empirical Insights and Recommendations. *Information Systems, 35*(In Press).

Mendling, J., & Strembeck, M. (2008). Influence Factors of Understanding Business Process Models. In W. Abramowicz & D. Fensel (Eds.), *Business Information Systems - BIS 2008* (Vol. 7, pp. 142-153). Innsbruck, Austria: Springer.

Miles, L. D. (1961). *Techniques of Value Analysis and Engineering*. New York, New York: McGraw-Hill.

Miller, G. A. (1995). WordNet - A Lexical Database for English. *Communications of the ACM, 38*(11), 39-41.

Moody, D. L., & van Hillegersberg, J. (2008). Evaluating the Visual Syntax of UML: An Analysis of the Cognitive Effectiveness of the UML Family of Diagrams. In D. Gasevic, R. Lämmel & E. van Wyk (Eds.), *Software Language Engineering - SLE 2008 Revised Selected Papers* (Vol. 5452, pp. 163-134). Toulouse, France: Springer.

Paivio, A. (1991). Dual Coding Theory: Retrospect and Current Status. *Canadian Journal of Psychology, 45*(3), 255-287.

Recker, J., & Dreiling, A. (2007). *Does It Matter Which Process Modelling Language We Teach or Use? An Experimental Study on Understanding Process Modelling Languages without Formal Education.* Paper presented at the 18th Australasian Conference on Information Systems, Toowoomba, Australia.

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.

Reijers, H. A., & Mendling, J. (2008). Modularity in Process Models: Review and Effects. In M. Dumas, M. Reichert & M.-C. Shan (Eds.), *Business Process Management - BPM 2008* (Vol. 5240, pp. 20-35). Milan, Italy: Springer.

Rosemann, M. (2003). Preparation of Process Modeling. In J. Becker, M. Kugeler & M. Rosemann (Eds.), *Process Management: A Guide for the Design of Business Processes* (pp. 41-78). Berlin, Germany: Springer.

Rosemann, M., Recker, J., Indulska, M., & Green, P. (2006). A Study of the Evolution of the Representational Capabilities of Process Modeling Grammars. In E. Dubois & K. Pohl (Eds.), *Advanced Information Systems Engineering - CAiSE 2006* (Vol. 4001, pp. 447-461). Luxembourg, Grand-Duchy of Luxembourg: Springer.

Schütte, R., & Rotthowe, T. (1998). *The Guidelines of Modeling - An Approach to Enhance the Quality in Information Models.* Paper presented at the Conceptual Modeling - ER 1998, Singapore.

Small, S., Cottrell, G., & Tanenhaus, M. (Eds.). (1988). *Lexical Ambiguity Resolution: Perspective from Psycholinguistics, Neuropsychology and Artificial Intelligence*. San Francisco, California: Morgan Kaufman.

Storey, V. C. (2005). Comparing Relationships in Conceptual Modeling: Mapping to Semantic Classifications. *IEEE Transactions on Knowledge and Data Engineering, 17*(11), 1478-1489.

van der Aalst, W. M. P., ter Hofstede, A. H. M., Kiepuszewski, B., & Barros, A. P. (2003). Workflow Patterns. *Distributed and Parallel Databases, 14*(1), 5-51.

Wynn, M. T., Verbeek, H. M. V., Van der Aalst, W. M. P., ter Hofstede, A. H. M., & Edmond, D. (2009). Business Process Verification – Finally a Reality! . *Business Process Management Journal, 15*(1), 74-92.

zur Muehlen, M., & Recker, J. (2008). How Much Language is Enough? Theoretical and Practical Use of the Business Process Modeling Notation. In M. Léonard & Z. Bellahsène (Eds.), *Advanced Information Systems Engineering - CAiSE 2008* (pp. 465-479). Montpellier, France: Springer.