Requirements of Process Modeling Languages – Results from an Empirical Investigation

Susanne Patig
University of Bern, IWI
Engehaldenstr. 8
CH-3012 Bern
susanne.patig@iwi.unibe.ch

Vanessa Casanova-Brito
University of Bern, IWI
Engehaldenstr. 8
CH-3012 Bern
vanessa.casanova-brito@iwi.unibe.ch

ABSTRACT
The majority of large and mid-sized companies are active in Business Process Management (BPM). Documenting business processes is a key task of BPM, but the variety of process modeling languages makes it difficult to determine 'the best' one. Basically, the suitability of a process modeling language depends on the companies' requirements. In this paper we adopt a bird's eye view on the issue: By an empirical investigation of 130 public companies from all over the world and any sector, we gather the common requirements of process modeling languages and use them to assess the most popular ones (i.e., BPMN, UML Activity Diagrams, Event-driven Process Chains). Our results show that these languages are (1) equally expressive and (2) presumably equally understandable concerning the common core notion of 'business process'; thus, they can be used interchangeably. However, the BPMN is the most complex process modeling language.

Categories and Subject Descriptors
D.2.1 [**Requirements**]: Languages; BPMN, UML, EPC

General Terms
Languages, Measurement

Keywords
Business process management, Requirements, process modeling languages, empirical investigation

1. INTRODUCTION
According to the BP Trends Report 2010, the majority of large and mid-sized companies in the world are active in business process management [31]. *Business process management (BPM)* comprises the design, administration, configuration, enactment and analysis of business processes [27]. All of these BPM activities require that the business processes are documented, which can be achieved by (business) *process modeling languages*; for brevity, we omit the term 'business' in the following. But, the large number of these languages confronts the companies with the problem of selecting 'the best one'. It is the aim of our research to contribute to the solution of this problem.

Several evaluations and comparisons of process modeling languages have been proposed; they are summarized in Section 2. The main weakness of the existing work is that the requirements, which are used to evaluate particular process modeling languages, stem from literature or BPM tools, but not from the companies' reality. Especially focusing on BPM tools is dangerous because this market is highly competitive, and some functionalities of BPM tools and features of process modeling languages rather satisfy the tool vendor's marketing department (and its attempts to diversify from other software solutions) than the customers' needs.

The research presented here overcomes this weakness: We have conducted an empirical investigation to gather the requirements concerning process modeling languages (see Section 3) and use them to assess the most popular process modeling languages (see Section 4). Our results (Section 5) are relieving for those trying to find 'the best' process modeling language.

2. RESEARCH BACKGROUND
Evaluations of process modeling languages either account for language pluralism or concentrate on the Business Process Modeling Notation (BPMN) [14] as the current de facto standard [31] of process modeling.

The pluralism-driven research addresses the question of 'the best' process modeling language as measured by expressiveness, understandability or complexity. *Expressiveness* means the capability to represent any meaning intended for some purpose, and it is usually assessed based on a reference [15]. In the field of process modeling languages, two types of *references* can be identified (see also Table 11 in Section 4.2):

1) *Widespread evaluation frameworks*[1] such as the Bunge-Wand-Weber (BWW) representation model (e.g., used in [5]) or different types of workflow patterns (e.g., used in [22]), and

2) *New evaluation frameworks* that were derived from the process modeling literature, e.g., [11], [24], [9].

The outcomes of such evaluations are either lists containing the equivalent, incomplete, overloaded, redundant or excess constructs of the analyzed process modeling languages [5] or evaluation vectors whose symbols indicate to which degree the elements of the reference are supported by the respective process modeling languages (e.g., [22], [9], [11]). Usually, both types of outcome convey the impression that neither process modeling language is 'good enough' in the sense of expressiveness. However,

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

[1] For a comparison between the ontology- and the pattern-based evaluation of process modeling languages see [20].

this impression is misleading as the evaluation references do not reflect the companies' requirements, but theoretical literature [5], [11], [24], [9] or functionality provided by BPM tools [22] - and provided functionality is neither necessarily needed nor used [17].

Research on the 'best' process modeling language in terms of *understandability* has just started out [12], [18]. Though no definite answer can be given yet, first findings indicate that process modeling languages are equally expressive as far as the common notion of 'business process' is concerned [18]. In contrast, if the 'best' process modeling language is the one with the smallest *complexity*, then UML activity diagrams are superior to the BPMN [21]. The complexity was calculated from the numbers of objects, relationships and properties in the meta models of the process modeling languages.

The standard-oriented evaluation research concentrates on the BPMN as the most widespread process modeling language [31] and tries to find out how good it is in real-life conditions. Once more, an important quality criterion is expressiveness: So, a series of interviews with nineteen participants from Australian organizations was conducted to test hypotheses about missing, redundant, superfluous and overloaded BPMN constructs [19]. The hypotheses were derived from the BWW representation model, and most of them had no or limited support in practice. On the one hand, these findings challenge the BWW representation model as a basis to assess process modeling languages; on the other hand they are first hints at the constructs needed in practice to model processes. We will discuss the results in more detail in Section 4.1.

Other research efforts analyze the use of the BPMN in real-life projects to identify unused constructs as well as the most frequently used ones [32]. It turned out that only 20% of the BPMN vocabulary appears in process models (see Section 4.1 for details). However, this analysis was conducted ex post facto and relied on given process models. Thus, it does not provide information on the constructs that would have been needed to adequately represent the companies' processes.

Altogether, the existing research does not help companies in selecting the process modeling language to be used, mainly for the following reasons: First, the academic way of assessing expressiveness does not start from the companies' requirements. Secondly, many of the process modeling languages evaluated in [9], [11], and [24] are not usable in large or mid-sized companies because appropriate tool support is missing. Though this argument does not apply to the BPMN, it is just one process modeling language among so many others. Recommendations, thirdly, should be given language-neutral, starting from the characteristics of the companies' processes and the resulting requirements. To remove these deficiencies, we have conducted an empirical investigation that is described in the next section.

3. EMPIRICAL INVESTIGATION ON THE REQUIREMENTS OF BUSINESS PROCESS MODELING

3.1 Research Goal

Our investigation adopts a bird's eye view on the common requirements of process modeling (i.e., irrespective of the individual companies' specifics) to enable a general assessment whether or not the current process modeling languages are sufficient and which one is possibly superior.

Since recent research findings indicate that there may be no significant difference in the understandability of process modeling languages if a common notion of 'business process' is kept (see Section 2), we concentrate on this common notion and try to define it more precisely. In particular, we want to find out what must be represented by process modeling languages to describe real-life business processes. By *business process* we mean a set of linked activities that collectively realize a business objective or policy goal [29]. Section 3.2 describes how we proceeded to reach this goal.

3.2 Method

Participants: The basic population for our investigation consisted of the companies from the 'Forbes Global 2000' list [3]. This list is an annual ranking of the top 2,000 public companies in the world based on sales, profit, assets and market value. As automation of process execution is the fourth most important reason to model processes [7], we require 'e-readiness'. *E-Readiness* describes the ability of a country and its businesses to use information and communication technology to their benefits [2]. The 'e-readiness' of 70 countries is assessed yearly on a scale between 1 (lowest) and 10 [2]. Since our investigation was conducted in 2009, we used the e-readiness rating of the year 2008, for which the average e-readiness score amounted to 6.4. All companies that are headquartered in countries with below-average e-readiness were discounted. Altogether, we draw a sample of 1,172 companies (by random numbers), which were contacted as described below.

Table 1. Numbers of responses per country

Countries	Responses	
	Per country	Total
United States	17	17
Switzerland	15	15
Germany; Japan	11	22
UK; Canada	9	18
Australia; France; Spain	5	15
Hong Kong; Italy; Portugal; Sweden; Taiwan	4	20
Austria; Belgium	3	6
Netherlands; Singapore; Greece; Denmark; Ireland; Korea; Norway	2	14
New Zealand; Luxemburg; Iceland	1	3
Sum N		130

In total, N = 130 companies responded (see Table 1); so, the response rate was 11%. Some questions were not answered by all companies; the resulting *missing values* are represented by '*na*' in the tables of this paper. Companies whose headquarters are located in Europe account for the majority (58%) of responses, followed by companies from North America (20%), Asia (18%) and Oceania (5%). All sectors are represented. Most responses came from the banking sector (19%), followed by utilities (8%), transportation and insurance (7% in each case) as well as oil & gas, and technology hardware (6% in each case). Some of the companies operate in more than one sector.

Half (53%) the companies' representatives who answered the questionnaire work in IT departments; other affiliations include departments for BPM (22.9%) or for company organization (13.3%), functional areas (9.6%) and product divisions (1.2%). In addition to knowledge about the processes of their own departments (29.4%), the participants stated that they also knew the processes of other departments (23.9%) or even have a company-wide picture (46.7%) of the processes.

Materials: The questionnaire consisted of 42 questions that were grouped in several sections dealing with BPM (current status, tools), processes (characteristics, change and statistics), process modeling (procedure, languages) and socio-demographic information. In this paper we concentrate on process modeling.

All questions of the questionnaire were partially open-ended, i.e., they provided a list with alternatives as well as an alternative 'other' to enter free text for unanticipated answers. The provided alternatives were derived from the process modeling literature (e.g. [1], [25], [6], [27]) and existing standards [28], [29]. Some questions were optional. The data was collected on nominal scales (participants were asked to select *all* alternatives that applied to their case) or on ordinal scale (participants were asked to rate some alternative or to make a ranking). All rating scales had four levels (to avoid neutral answers) and an additional level (e.g., 'Don't know', 'Not applicable', 'Not needed', 'None') to avoid forced ratings [8]. The resulting total number of five rating levels is generally reckoned optimal since the ability to differentiate between ratings decreases with an increasing number of rating levels [8]. In detail, we have used the following *rating scale types*:

- *Scale 1*: 1=Essential, 2=Frequently needed, 3=Occasionally needed, 4=Rarely needed, 5=Not at all needed
- *Scale 2*: 1=Always, 2=Very Often, 3=Sometimes, 4=Rarely, 5=Never
- *Scale 3*: 1=All processes, 2=Most processes, 3=Some processes, 4=A few processes, 5=No processes
- *Scale 4*: 1=Very important, 2=Important, 3=Not so important, 4=Not at all important, 5=Don't know

The questionnaire, including definitions of key terms used in the questions, was written in English and implemented as an online form with the 'LimeSurvey' tool [10]. To contact the companies, we sent letters by surface mail containing the link to the online form and explaining the goals and importance of our investigation. These letters were written in English, German, Spanish and Japanese – depending on the location of the company's headquarter – to make the goals and importance of our research understandable for the addressees.

Procedure: We conducted a pretest of the questionnaire with 10 BPM experts from practice and academia to check the questions for understandability and unambiguity of responses; afterwards, the questionnaire was revised. As little is known about the internal organization of BPM in a company, we sent the letters with the link to the revised questionnaire to the CIOs or CEOs of the companies in the sample and asked them to forward the letters to the persons responsible for BPM. After four to six weeks, we phoned the offices of the CIOs or CEOs to inquire after the status of our information request. After six to eight weeks, we sent reminder letters by surface mail. The survey was conducted from January to December 2009. No incentives were given; the companies were only offered the opportunity to obtain the results of the investigation free of charge.

3.3 Results

The wording of the questions (Q1) to (Q17) discussed here is given in the Appendix. As our questions on *nominal scale* ('choose *all* alternatives that apply …') allow more than one answer, the sum of counts c_i for an alternative (answer) can exceed the number N of responding companies. The answer alternative with the highest count (absolute frequency) represents the *mode* [4]. In the text, we give the count c_i for each alternative and the *percentage of responses* (c_i / N), which relates the count of an alternative to the number N of responses ($N \leq 130$, depending on the number *na* of *missing values*). If questions involved entering *open text*, this was optional, and we refer to the total number of received answers (Σc_o). For the open text questions, we state the *percentage of answers* (c_i / Σc_o). Questions on *ordinal scale* can be recognized by the rating scale type from Section 3.2. According to this rating scale type, a numerical value is associated with each rating level to calculate the mean rating (μ) and the standard deviation (σ) of ratings; the numerical values 1 and 5 correspond to the highest and the lowest rating, respectively.

Naturally, the requirements concerning process modeling languages depend on the characteristics of the processes to be described. For that reason we first gathered information on the nature of the companies' processes (Questions (Q1) to (Q6)).

Concerning their scope **(Q1)**, the processes in the companies of our survey are related to (N=130, na=0) products (78/60.9%), administration (71/55.5%), customer contact (67/52.3%), system integration (66/51.6%), system development (65/50.8%), emergency procedures (33/25.8%) and other things (12/9.4%). To check whether the understanding of what constitutes a business process agrees among the companies, we asked the participants (as an open text question) to freely list at least one title or short description of a typical process **(Q2)**. Having obtained Σc_o =265 short descriptions, we clustered similar answers. The clusters of the most frequently mentioned titles can be summarized as 'Order-to-pay' (18/7%) and 'Purchase-to-Pay' (17/6%). The next smaller clusters are related to project management (9/3%), human resources issues (9/3%) and incident management (8/3%). Other process titles mentioned characterize development processes (7/3%) and change management processes (7/3%). Across all clusters, 7 titles (3%) emphasize that the processes involve approval. Altogether, the entered answers to Question (Q2) basically confirm our results of Question (Q1).

Table 2. Nature of processes according to modes

Question	Rating			N/na
	Most	Some	A few	
Distribution of processes (Q3)	1 Company & >1 Department (83)	1 Department & >1 Person (51); >1 Company (55)	1 Department & 1 Person (60)	130/0
Measure of average process run time (Q4)	In days (54)	In weeks (49)	In month (34); in years (30)	87/43
Execution frequencies of processes (Q5)	Several times a day (59)	Several times a week (58) / a month (56)	Several times a year (50)	102/ 128

In addition to the process scope, we inquired after the distribution of the processes **(Q3)**, the average run time **(Q4)** and execution frequency **(Q5)**, see Table 2, as well as process statistics (Q6). The cells of Table 2 show all alternatives of the questions and their most frequent rating (i.e., the mode). The rating scale (Type

3) corresponds to the columns; the particular counts are given in brackets. Obviously, most processes stay within one company, but span several departments. The processes are executed several times a day and have a run time that is measured in days.

Table 3 gives the statistics for the companies' average processes **(Q6)**. The numbers confirm Table 2 in that the processes involve few other companies (μ = 1.60), but usually several departments (μ = 3.70).

Table 3. Statistics for an average process (N=128, na=2)

(Q6) Please estimate the…	Min	Max	μ	σ	N/na
Number of involved persons from the same department	1	55	7.31	11.59	70/60
Number of involved departments	1	18	3.70	2.56	73/57
Number of other companies involved	0	8	1.60	1.39	58/72
Number of applications involved	1	24	4.27	4.22	69/61
Number of tasks	3	120	19.09	20.82	60/70

The aim of the Questions (Q1) to (Q6) was to get an idea of the processes in the surveyed companies – also to enable the proper assessment of the companies' answers concerning process modeling. We were mainly interested in the concepts needed to describe processes (Questions (Q8) to (Q16)). These concepts may not only depend on the processes' characteristics, but also on the modeling style. Therefore we first asked how the companies proceed in modeling processes **(Q7)**, in particular, what constitutes the starting point for creating process models. According to their responses (N=120; na=10), the companies start to model processes from:

1. The interactions (including data flow) between departments or responsible persons (72/60%)
2. Some goal and the activities to achieve it (57/48%)
3. The products of the company and their transformation from an initial to a final form (44/37%)

Several companies (Σc_o =17) named other procedures: Most often, 'top-down approaches' (4/3%) and 'customer touch points' (4/3%) were mentioned, followed by the reuse of results from existing ERP projects (3/2.5%). Top-down approaches are similar to our alternative 'starting from goals', whereas customer touch points are close to 'interactions'.

Our research goal was to find out which concepts companies need to describe their processes, and we posed this as a question **(Q8)**. Table 4 shows the results (rating scale of Type 1). Tasks are the by far most important concept, followed by relationships, persons or roles, documents, events and resources. Time information and timely distances between tasks are rated as least important. When asked openly to list other concepts needed (Σc_o =19), constructs to describe exceptions and exception handling (3/2%) were mentioned most often, immediately followed by business rules (2/2%) and the data flow between tasks (2/2%). Here, the companies did not recognize that exceptions are a special type of events.

Table 4. Concepts needed to describe processes in the company/department (N= 130, na=0)

Answer	μ	σ	Rank
Tasks to be done in a process	1.34	.63	1
Relationships between tasks	1.60	.84	2
Persons or roles who execute the tasks	1.68	.86	3
Documents or objects relevant for a task	1.91	.84	4
Events that happen and influence a task	2.22	.86	5
Resources needed to fulfill the tasks	2.24	.89	6
Time information related to a task	2.42	.89	7
Timely distances between tasks	2.83	.94	8

The previous question has shown that tasks and their relationships are the most important concepts. Question **(Q9)** investigated the typical relationships between tasks in real-life processes; Table 5 gives the results (rating scale of Type 2) for each alternative. The most frequent relationships between tasks are sequences, followed by start conditions (applicable to any form of relationship) and process splits. Loops and alternative tasks are rare.

Table 5. Frequency of observations that apply to tasks in processes (N= 130, na=0)

Answer	μ	σ	Rank
Tasks follow in strict sequence	2.32	.72	1
Condition must be satisfied in order to start a task	2.40	.64	2
Task has more than one immediate successor (split)	2.47	.61	3
Task has more than one immediate predecessor (merge)	2.58	.64	4
Task is repeated till some criterion is satisfied (looping)	2.82	.78	5
Tasks are alternative to each other	3.16	.82	6

According to Question (Q9) and Table 5, start conditions are common. We also inquired into the nature of these conditions **(Q10)**. As reported by the companies (N=129; na=1), start conditions relate to:

- The results of adjacent tasks (92/71%),
- The overall state of the process (82/64%),
- The task only (e.g. the availability of resources) (70/54%)
- Information external to the process (60/47%)
- Time (50/39%)

Splits, which are the third most important relationship between tasks (see Question (Q9) and Table 5), raise the situation that a task has several successors. Table 6 summarizes the companies' answers (rating scale of Type 2) about what happens with the successors **(Q11)**. Mostly, all of the successors are executed, either concurrently (with synchronization) or independent of each other (without synchronization). If not all succeeding tasks are executed, the selection mostly depends on a condition (see Question (Q10) for the nature of conditions). In line with the results of Question (Q9), exclusive choices are not very frequent.

Table 6. Execution of succeeding tasks

Answer	N/na	μ	σ	Rank
All (AND) … and concurrently	108/22	2.64	.76	1
….independent of each other	109/21	2.97	.70	2
Not all tasks are executed (OR)	106/24	3.25	.85	3
Which tasks are executed depends on…				
… a random selection	99/31	4.09	.87	3-IV
… some selection condition is satisfied	107/23	2.55	.73	3-I
… subjective experience	106/24	3.22	.82	3-II
As soon as one task has been executed, no other task is performed (XOR)	98/32	3.29	.92	3-III

In addition to tasks and their relationships, persons, events and resources are important to describe processes (see Question (Q8), Table 4). In the questionnaire, these concepts were investigated in more detail by the Questions (Q12) to (Q14):

Table 7 lists the resources companies wish to include in process models (Question **(Q12)**, rating scale of Type 1). Information is most frequently needed, followed by personnel and documents. Note that 'information' (e.g., a database listing the availability of items in store) and 'document' do not necessarily coincide.

Table 7. Resources needed in process models (N= 130, na=0)

Answer	μ	σ	Rank
Information	1.53	.61	1
Personnel	1.72	.86	2
Documents	1.77	.70	3
Software systems	1.98	.95	4
Material	2.75	1.11	5
Machines, appliances	3.04	1.13	6

If persons are used in process models, the companies would like to represent the following details (Question **(Q13)**, N=130; na=0):

1. Required capabilities, roles, positions (110/85%)
2. Organizational units the persons belong to (94/72%)
3. Names (46/35%)
4. None information at all (5/4%)

When asked to mention other details wished to be included, (2/14%) of the respondents (Σc_o =10) mentioned the availability of the person executing the task.

Finally, information is not only an important resource (see Question (Q12), Table 7), but also the most important trigger of processes or tasks – Table 8 shows the results of Question **(Q14)** on a rating scale of Type 3. Other common event types are timing and human intervention.

Table 8. Triggers of processes or tasks (N= 130, na=0)

Answer	μ	σ	Rank
Company internal *information*	2.58	.87	1
Information from business partners	2.62	.81	2
Timing (date, cycles)	2.82	.89	3
Human judgment and intervention	2.84	.79	4
State of process or task	3.10	.87	5
Deviations from targets or failures	3.45	.92	6
Signals from machines and sensors	3.91	.81	7

The Questions (Q8) to (Q14) dealt with constructs provided by most process modeling languages. We also wanted to know **(Q15)** whether the companies need additional concepts to describe their processes or tasks. As it can be seen in Table 9, the execution status (started, cancelled, etc.) is the concept most needed for the task, followed by priorities, goals, cost and planning status, whereas goals are the most important information for processes. The open Question **(Q16)** aimed at getting examples for process goals. Among the Σc_o =127 answers obtained, timing goals (19/15%) were the most frequently mentioned category. Examples of timing goals are the reduction of cycle time, throughput time etc. Timing goals were followed by customer satisfaction (8/6%) and quality-related goals (7/5%). Other named goals included the resolution of problems (6/4%) as well as the reduction of error rates (5/4%).

Finally, Table 10 summarizes how the sampled companies currently document their processes **(Q17)**: Most companies combine text (55.9%) and some (modeling) language (altogether 55.9%), but also tables are widespread (31.5%). Among the process modeling languages, the BPMN dominates, followed by the Unified Modeling Language (UML) and Event-driven Process Chains (EPC). Other ways (Σc_o =124) of describing processes are value chains (3/2%), flow charts (3/2%) and company-specific notations (4/2%).

Table 9. Additional concepts needed in process models (N=111, na=19)

Answer	Execution status		Priorities		Goals		Cost		Planning status	
	Task	Process	Task	Process	Task	Process	Task	Process	Task	Process
Needed (c_i)	71	68	61	54	54	80	48	58	44	56
Percentage responses (c_i / N)	64%	61%	55%	49%	49%	72%	43%	52%	40%	50%
Rank	1	2	2	5	3	1	4	3	5	4

Table 10. Documentation of processes (N = 127; na = 3)

Answer		Count c_i	Percentage responses (c_i / N)
As text		71	55.9%
As tables		40	31.5%
With languages	BPMN	27	21.3%
	UML	19	15.0%
	EPC	16	12.6%
	BPEL	5	3.9%
	IDEF	4	3.1%
Other		14	11.02%

4. DISCUSSION

4.1 Comparison with other Empirical Results

In this section we compare our findings concerning the concepts needed to describe processes with other empirical results that have been obtained for the BPMN: Based on an analysis of 120 BPMN models, it was revealed that the most often used BPMN constructs are the following (ordered by decreasing occurrence frequency): sequence flow, task, end/start/general event, pool (to express organizational information or roles) and gateway [32]. Our Table 4 agrees with these results in the (naturally) high importance of the construct 'task' (Rank 1 in Table 4 as opposed to Rank 2 in [32]), but ranks organizational information (persons, roles) higher than events. Moreover, there is also some agreement concerning the relationships between tasks: Though the occurrence frequency of the gateway types (AND occurs more frequently than XOR [32]) depends on the particular processes modeled, our results are analogous (see Table 6). Loops of activities play a subordinate role in process models, both according to our results (Table 5) and according to [32].

The contradictions stated above probably result from the fact that the ranking obtained in [32] also reflects the BPMN syntax and not only process modeling requirements. For example, the sequence flow is the most frequent construct in BPMN models [32] because it is needed to connect tasks and events, which, however, have to be defined before. Moreover, the comparatively low importance of events in our results can be explained by the fact that we inquired about the constructs needed to describe processes irrespective of any modeling language (also textual descriptions were allowed); thus, our results are free from any syntactic restrictions that require the use of events (e.g., start and end events).

Interestingly, the dominance of the task in our results cannot be derived from the modeling style since most companies use the interactions between departments as the starting point to model processes (Q7).

By testing hypotheses derived from the BWW representation model, another research [19] also aimed at identifying needed or excess (and, thus, unused) BPMN constructs. The following was discovered [19]: First, there is only moderate need for business rules, which supports our results of Question (Q8) where business rules were mentioned only two times as 'other' concepts. Secondly, data objects turned out to be the fourth most useful of the hypothetically 'excess' BPMN constructs, which agrees with the fourth rank of documents in our Table 4. Thirdly, activity looping rated comparatively high in perceived usefulness, which contradicts both our results in Table 5 and [32]. Altogether, the results of [19] are restricted by the BWW representation model and the BPMN, and our research is not subject to these restrictions; thus, a detailed comparison is difficult.

Our research augments the existing results for the BPMN by a language-neutral view, and language neutrality explains most of the deviations from the construct rankings observed by other researchers. Whether or not construct rankings are important for selecting among process modeling languages is discussed in the next section.

4.2 Comparison with Process Modeling Languages and Evaluation Frameworks

Our empirical investigation (see Section 3.1) aimed at discovering the concepts that must be represented by process modeling languages to describe real-life processes. The first column of Table 11 summarizes our results. The order of the concepts in bold print corresponds to the ranks from Table 4. Most of these concepts were investigated in more detail by further questions (e.g., Question (Q9) for relationships; see Section 3.3), whose numbers are given and whose rankings determine the order of the more detailed concepts. Concepts with a rounded mean rank below three are omitted in Table 11.

The final goal of our research (see Section 1) was to support companies in selecting 'the best' process modeling language. Our empirical investigation enables an assessment of process modeling languages by the required expressiveness. Valid references for expressiveness must be consistent sets of independent, atomic statements [15]. For that reason the first column of Table 11 differs as follows from the tables in Section 3.3:

- Personnel and documents are no subtypes of the concept 'resource' (see Table 7), but distinct concepts (see Table 4).
- Information (see Table 7) is not explicitly listed as a resource because it is provided by software systems, documents or events, which are separate concepts.
- Time information for tasks (see Table 9) does not appear in Table 11 as it can be expressed by time events for durations and deadlines (see also [14]).
- The concept ‚material' (Table 7) can be seen as a special type of ‚object' and is, thus, omitted in Table 11.

Altogether, the first column of Table 11 is a valid reference for measuring expressiveness. The expressiveness of the process modeling languages that were most frequently used in our sample (see Table 10 in Section 3.3) is apparent from Table 11: We count a line of the reference (*reference item*) as expressible (count: 1) by a process modeling language if it is directly or indirectly supported. *Direct support* means that the process modeling language provides a construct that readily represents the semantics of the reference item, whereas *indirect support* requires the combination or appropriate definition of existing constructs. Indirect support by adapting more general constructs is typical for the reference items representing detailed concepts in Table 11. For example, the general constructs 'pool' and 'lane' of the BPMN or 'partition' of UML Activity Diagrams can be used to express organizational units, capabilities or even names. Moreover, the generic construct 'event' of the EPC can be used to represent all specific event types such as messages, timing etc.

Table 11. Expressiveness of prominent process modeling languages and evaluation frameworks

Our empirical results Concepts needed to describe processes (Q8)	**Existing process modelling languages**			**Standards**		**Evaluation frameworks**				
						Newly proposed			**Widespread**	
	BPMN [14]	UML Activity [13]	EPC for ARIS [23]	WfMC [28]	XPDL [30]	[11]	[24]	[9]	BWW [5]	WF-Pattern [22] [26]
Tasks	✓	✓	✓	Activity	Activity	Activity	Activity	Activity	Trans-formation	♦
Relationships between tasks (Q9)	✓	✓	✓	Transition	Transition	Relation	Depen-dency	Control flow	State law	♦
Sequence	✓	✓	✓	Transition	Transition	Relation	✓	Control flow	♦ State law	WFCP-1
Start Condition (of task)	♦	✓	♦	Pre-condition	TC	(Behavior)	♦	—	Lawful trans-formation	WFDP-33, 34, 35, 38, 39
Split (Q11)	✓	✓	✓	TC	Route-A	Behavior	♦	✓	— (s. GrRo)	WFCP-2, 6
AND	✓	✓	✓	TC	Route-A	Behavior	Fork	✓	—	WFCP-2
OR	✓	✓	✓	TC	Route-A	Behavior	Selection	✓	—	WFCP-6
XOR	✓	✓	✓	TC	Route-A	Behavior	Selection	✓	—	WFCP-4
Merge	✓	✓	✓	TC	Route-A	Behavior	✓	Join	—	WFCP-3, 5, 7, 8, 9, 30 to 38
Loop	✓	✓	♦	TC	Activity	Behavior	—	♦	—	WFCP-10, 21
Persons or Roles (Q13)	✓	✓	♦	Role	Pool, Participant	Agent	Role	Role	Thing	R-RBA
Capability	♦	♦	♦	—	Participant	—	♦	♦	Property	R-CBA, R-DE
Organizational Unit	♦	♦	✓	—	Pool, lane	—	♦	✓	Thing	R-OA
Name	P	♦	♦	—	Participant	—	(Actor)	—	Property	R-DBAS
Documents or objects	♦*	✓	✓	WF-rele-vant data	(Data Object)	Entity, In-formation	Informa-tion	Data Object	Thing	WFDP-2, 3, 5, 6
Events (Q14)	✓	♦	✓	— See Text	✓	✓	✓	✓	✓	WFDP-37
Information	✓	♦	✓	—	Message	✓	✓	♦	♦	WFDP-1, 5, 6, 8 to 14, 16 to 18, 20 to 22, 24, 25, 38
Timing	✓	—	♦	P	Timer	♦	✓	♦	♦	(WFEH Deadline)
Human intervention	♦	♦	♦	—	—	♦	—	♦	♦	WFDP-7, 14 to 25 (WFCP-16; WFEH External trigger)
State of process/taks	♦	♦	♦	—	♦	♦	✓	♦	State	(WFDP-38;WFEH)
Deviations	✓	✓	♦	—	Error	♦	—	♦	♦	WFEH (except for External Trigger)
(Additional) Resources** (Q12)	✓	✓	✓	—	♦	✓	✓	✓	Thing	R-DA, -FBA, -RA, -CH, -RF, -HBA, -DBOS, -DBOM, -RMA, -RRA, -SHQ, -ED, -LD, -D, -E, -SD, -AR
Software Systems	♦	♦	✓	Invoked application	Applica-tion	Informa-tion	♦	Applica-tion	Thing	♦
Machines, Appliances	♦	♦	✓	—	Participant	♦	♦	♦	Thing	♦
Additional information for a task (Q15)										
Execution status	✓	♦ S	—	✓	♦ (Extension Attribute)	—	State	—	State	WFDP-1, (WFDP-3)
Priority	♦ P	♦ S	—	P		—	—	—	Property	
Goal	♦ P	♦ S	✓	—		—	✓	✓	Property	
Cost	♦ P	♦ S	—	—		—	—	—	Property	
Excess constructs or elements provided by the process modeling languages or frameworks										
	≈ 25	8	—	—	9	—	3	≈ 15	≈ 22	≈ 57
Expressiveness Core/Total Reference										
	1 / 1	0.96 / 0.96	1 / 0.89							

Legend: ✓ Directly supported (count: 1) ♦ Indirectly supported (count: 1) —: Not supported (count: 0)
* No influence on control flow.
** Only if not already listed before; rearrangement due to the calculation of expressiveness (see text).

Abbreviations: P/S: Can be defined by a construct's properties/stereotypes, Route-A: Route Activity, R: Workflow resource pattern.
TC: Transition condition, WFCP/DP/EH:Workflow control pattern/data pattern/exception handling pattern

Expressiveness is calculated by relating the sum of all expressible reference items (irrespective of their order/rank) to the sum of all items of the reference [15]. The *total reference* of Table 11 consists of 28 items[2]; if the additional information for a task is excluded, 24 items remain in the *core reference*. For the core reference, the BPMN and EPCs are equally expressive (see Table 11); UML Activity Diagrams slightly lag behind because of their inability to express timing. For the total reference, the BPMN is the most expressive language; both the BPMN and UML Activity Diagrams benefit from the languages' extension mechanisms such as freely definable attributes or stereotypes.

Altogether, the differences in expressiveness are marginal, especially within the core reference. For the companies this result implies that the BPMN, EPCs and UML Activity Diagrams are interchangeable. Technically, this interchangeability is (except for events that represent human intervention) supported by the XML Process Definition Language (XPDL) [30], a standard for the exchange of process models (diagrams) between tools; see Table 11. The XPDL support even comprises the total reference as additional information for tasks can be defined by 'extension attributes'.

If we measure complexity by the count of constructs of a process modeling language that do *not* represent reference items ('excess constructs' in Table 11), our findings support the larger complexity of the BPMN compared to UML Activity Diagrams as reported in [21].

To sum up our results, the most widespread process modeling languages equally satisfy the companies' requirements concerning process modeling. Thus, our evaluation of process modeling languages is far more positive than the existing ones (see Section 2). This can be explained as follows: First of all, expressiveness always depends on the purpose ('What must be described?'), which also dictates the evaluation basis (reference). Distinct references will naturally provoke different assessments of process modeling languages and their expressiveness. The references used in the field of process modeling languages (see Section 2) are given in Table 11. On the one hand, they are very generic (e.g., [11], [5]) and, thus, do not adequately cover the requirements of process modeling. On the other hand, the number of excess elements hints at 'over-engineering' of the references (especially [9]) or at the inclusion of purposes beyond process modeling, namely workflow execution and the corresponding tool support [22]. The findings of our empirical investigation suggest that modeling business processes requires far less than the workflow patterns, but more than the WfMC reference model [28], which in fact corresponds to the 'least common denominator'. For example, though 'events' are mentioned several times in the document of the standard, they are no part of the process definition [28].

4.3 Limitations of our Research

Our results can be assumed to be representative for large companies that operate worldwide in any sector. However, the list 'Forbes Global 2000' we used suffers from three limitations: First, it disregards large non-American companies that don't have commercial relations with the USA. Secondly, ranking companies based on sales, profit and market value favors sectors where borrowed capital is important (e.g., banking and insurance companies). Thirdly, non-profit organizations (i.e., public administration, universities) are completely excluded. These limitations might well affect the validity of our results.

The ratings we have obtained from the companies reflect the subjective experiences of our participants, which is a common limitation of such surveys (e.g., [19]). Our rating scales (see Section 3.2) cover perceived need, importance or frequency – especially in the latter case we cannot expect that the number of tasks in all process models of a company was counted. This could easily be mistaken for a disadvantage, but also the representativity of process models in a sample cannot be guaranteed [32].

From a methodical point of view, the calculation of mean ratings is only valid if the ratings of the underlying scale are equidistant (interval scale). This assumption is generally made [8]. Additionally, our main conclusions rely on these means only to exclude two items (signals, planning status) from the reference in Table 11; thus, the effect of this methodical assumption is negligible.

Process modeling can serve different purposes; the most important ones for companies are Business Process Reengineering, documentation, understanding and communication [7], [16]. These purposes need expressiveness concerning description, which we have measured here. Nevertheless we acknowledge that distinct purposes (e.g. workflow execution by BPM tools) may require other expressiveness, i.e., another reference, which will lead to deviating evaluation results.

Finally, gathering requirements is usually not done by questionnaires. For that reason we currently conduct qualitative interviews with selected companies from the sample.

5. CONCLUSIONS AND FUTURE RESEARCH

From an inquiry of 130 public companies from all over the world we gathered a common set of requirements that reflect the concepts needed to describe real-life business processes. We used these requirements as a reference to assess the expressiveness of the currently dominating process modeling languages, i.e. the BPMN, UML Activity Diagrams and EPCs. On the whole, the expressiveness of these process modeling languages is equal and, thus, they can be used interchangeably – if the process models are created for documentation, understanding, and communication or to support Business Process Reengineering.

The required expressiveness also depends on the purpose of modeling. Since our survey gathered the reasons to describe processes, we will conduct a more detailed analysis on the correlation between needed language constructs and process model usage.

If we accept the core reference we have found as the common notion of 'business process', the results of [18] indicate that the evaluated process modeling languages should be equally understandable. This is another topic for future research.

6. APPENDIX

In the following, we give the wording of the questions that are discussed in this paper as they appeared in the questionnaire:

(Q1*) To what are your processes related?* Product of our company, Customer contact, Administration, Systems integration; data transformation, Systems development, Emergency procedure, Other.

[2] The line ‚Additional information for a task' is a heading and, thus, not counted.

(Q2) *List at least one "title" or short description of typical processes in your company/department:*

(Q3) *Please assess the following statements about the distribution of the processes in your company!* (Scale Type 3) Processes are executed within one department by one person, The processes are executed within one department but involve more than on responsible person, The processes are executed within the company by more than one department, The execution of the processes involves also other companies.

(Q4) *How do you measure the average run time of the processes in your company?* (Scale Type 3) In days, In weeks, In months, In years.

(Q5) *How often do you execute the processes in your company?* (Scale Type 3) Several times a day, Several times a week, Several times a month, Several times a year.

(Q6) *Please estimate the following numbers for an average process in you company (wherever it applies)!* The number of persons from the same department involved in the execution of an average process, The number of departments involved in the execution of an average process, The number of application systems involved in the execution of an average process, The number of tasks or activities of an average process.

(Q7) *How did you proceed in describing your processes?* We started from the products of our company and concentrated on their transformation from an initial to a final form, We started from some goal and grouped the activities to achieve it, We started from the interactions (including data flow) between departments or responsible persons.

(Q8) *Which entities do you need to describe the processes in your company/department?* (Scale Type 1) The tasks that have to be done in the process, The relationship between the tasks, The persons or roles who execute the tasks, Resources needed to fulfill the tasks (e.g. machines, material, documents, software systems, etc.), Time information related to the task (e.g. durations, start/end time points), Timely distances between the end of a task and the start of the following task.

(Q9) *How often do the following observations apply to the tasks in your processes?* (Scale Type 2) The tasks follow each other in a strict sequence. In other words, each task has at the most one predecessor and one successor, A task has more than one immediate successor (process splits), A task has more than one immediate predecessor (process merge), Some condition (other than the end of the preceding task) must be satisfied in order to start a task, A task (or a group of tasks) is repeated, till some criterion is satisfied, Tasks are alternatives to each other.

(Q10) *If the start or the selection of a task depends on a condition, the condition refers to:* The tasks only, e.g., the availability of resources, The results of adjacent tasks, The overall state of the process, Information external to the process, Time, Other.

(Q11) *If more than one task can be started, they are performed:* (Scale Type 2) All and concurrently, All independent of each other and in any order, Not all tasks are executed, Which task(s) is/are chosen for execution, depends on: -a random selection, -whether or not some condition is satisfied, -the subjective experience of some person in charge, As soon as one task has been executed no other tasks from the ones that can be started is performed.

(Q12) *Which resources do you (wish to) include in the models or descriptions of your processes?* (Scale Type 1) Machines, appliances, Personnel, Material, Documents, Information, Software Systems.

(Q13) *Which information about the persons executing the task do you (wish to) include in the models or descriptions of your processes?* None, The required capabilities (roles, positions), Names of the responsible persons, Organizational units they belong to, Other.

(Q14) *What triggers the execution of your processes or their tasks?* (Scale Type 3) Information sent from business partners (customers, suppliers, etc.), Company-internal information, Timing (date, cycles), Signals from machines or sensors, Deviations from targets or failures, The state of some process or task, Human judgment and intervention.

(Q15) *What kind of additional information do you want to gather in your models?* Needed for the task & Needed for the process: Priorities, Cost, Goals, Execution status (e.g., cancelled…), Planning status (e.g., plan/actual; strategic/tactical).

(Q16) *Could you give us at least one example of goals you wish to express?*

(Q17) *How do you describe the processes in your company/ department?* As text in normal language, As tables, By using a (process) modeling language, e.g., BPEL, BPMN, EPC, IDEF, UML, Other (process) modeling language, please specify.

7. REFERENCES

[1] Becker, J., Kugeler, M. and Rosemann, M. 2003. *Process Management – A Guide for the Design of Business Processes*. Springer, Berlin et al.

[2] Economist Intelligence Unit (in cooperation with the IBM Institute for Business Value): *E-readiness rankings 2008: Maintaining momentum.* A white paper from the Economist Intelligence Unit. London et al., 2008. http://www.eiu.com/site_info.asp?info_name=ibm_ereadiness&page=noads&rf=0.

[3] Forbes: *The Forbes Global 2000*. http://www.forbes.com/lists/2008/18/biz_2000global08_The-Global-2000_Rank.html.

[4] Gravetter, F. and Wallnau, L. 2009. *Statistics for the Behavioral Sciences*. 8th ed., Wadsworth, Belmont.

[5] Green, P. and Rosemann, M. 2000. Integrated process modeling: An ontological evaluation Information Systems. In *Proceedings of the 11th International Conference of Advanced Information Systems Engineering*. (Stockholm, Sweden, June 05 - 09, 2000). CAiSE 2000. Elsevier. Vol. 25, No. 2, 73-87. DOI= http://dx.doi.org/10.1016/S0306-4379(00)00010-7.

[6] Holt, J. 2009. *A Pragmatic Guide to Business Process Modeling*. 2nd ed., BCS, Swindon.

[7] Indulska, M., Recker, J., Rosemann, M. and Green, P. 2009. Business Process Modeling: Perceived Benefits. In *Proceedings of the 28th International Conference on Conceptual Modeling* (Gramado, Brazil, November 09 - 12, 2009). ER 2009. LNCS, Vol. 5829, 458-471. Springer, Berlin.

[8] Jackson, S. 2009. *Research Methods and Statistics: A Critical Thinking Approach*. 3rd ed., Wadsworth, Belmont.

[9] List, B. and Korherr. B. 2006. An evaluation of conceptual business process modelling languages. In *Proceedings of the ACM Symposium on Applied Computing* (Dijon, France,

April 23 – 27, 2006). ACM, New York, NY, 1532-1539. DOI= http://doi.acm.org/10.1145/1141277.1141633.

[10] LimeSurvey. Version 1.85. http://www.limesurvey.org/.

[11] Lin, F., Yang, M. and Pai, Y. 2002. A Generic Structure for Business Process Modeling. *Business Process Management Journal*. 8, 1 (2002), 19-41.

[12] Mendling, J., Reijers, H.A. and Cardoso, J. 2007. What Makes Process Models Understandable? In *Proceedings of the 5th International Conference on Business Process Management* (Brisbane, Australia, September 24 – 29, 2007). BPM 2007. LNCS Vol. 4717, 48-63. Springer, Berlin.

[13] Object Management Group (OMG): *OMG Unified Modeling Language (OMG UML), Superstructure, Version 2.2.* OMG Document Number: dtc/2010-06-05, http://www.omg.org/technology/documents/formal/uml.htm.

[14] Object Management Group (OMG): *Business Process Model and Notation (BPMN) Version 2.0*. OMG Document Number: dtc/2010-06-05, http://www.omg.org/cgi-bin/doc?dtc/10-06-04.

[15] Patig, S. 2004. Measuring Expressiveness in Conceptual Modeling. *In Proceedings of the 16th International Conference of Advanced Information Systems Engineering*. (Riga, Latvia, June 07 – 11, 2004). CAiSE 2004. Springer, Berlin et al. LNCS Vol. 3084, 127-141.

[16] Patig, S., Casanova-Brito, V. and Vögeli, B. 2010. IT Requirements of Business Process Management in Practice – An Empirical Study. In *Proceedings of the 8th International Conference on Business Process Management* (Hoboken, USA, September 13-16, 2010). BPMN 2010.

[17] Recker, J. 2008. *BPMN Modeling – Who, where, how and why. BP Trends, March 2008*. http://www.bptrends.com/publicationfiles/05-08-ART-BPMN%20Survey-Recker-JR%20final.pdf.

[18] Recker, J. and Dreiling, A. 2007. Does It Matter Which Process Modeling Language We Teach or Use? An Experimental Study on Understanding Process Modelling Languages without Formal Education. In *Proceedings of the 18th Australasian Conference on Information Systems* (Toowoomba, Australia, December 05 - 07, 2007). ACIS 2007.

[19] Recker, J., Indulska, M., Rosemann, M. and Green, P. 2006. How Good is BPMN Really? Insights from Theory and Practice. In *Proceedings of the 14th European Conference on Information Systems* (Göteborg, Sweden, June 12 – 14, 2006). ECIS 2006.

[20] Recker, J.; Rosemann, M. and Krogstie, J. 2007. Ontology- Versus Pattern-Based Evaluation of Process Modeling Languages: A Comparison. In *Communications of the Association for Information Systems*. 20, Article 48 (2007). http://aisel.aisnet.org/cais/vol20/iss1/48.

[21] Recker, J., zur Muehlen, M., Siau, K., Erickscon, J. and Indulska, M. 2009. Measuring Method Complexity: UML versus BPMN. In *Proceedings of the 15th Americas Conference on Information Systems* (San Francisco, California, August 06 – 09 2009). AMCIS 2009. AISeL, Paper 541. http://aisel.aisnet.org/amcis2009/541.

[22] Russell, N., ter Hofstede, A.H.M., van der Aalst, W. and Mulyar, N. 2006. *Workflow control-flow patterns – A revised view.* BPM Center Report BPM-06-22, BPMcenter.org. 2006. http://www.workflowpatterns.com/documentation/documents/BPM-06-22.pdf

[23] Scheer, A.-W., Thomas, O. and Adam, T. 2005. Process Modeling Using Event-driven Process Chains. In *Process-Aware Information Systems*. Dumas, M., van der Aalst, W. and ter Hofstede, A.H.M., Ed. Wiley, Hoboken, New Jersey, 119-145.

[24] Söderström, E., Andersson, B., Johannesson, P., Perjons, E., and Wangler, B. 2002. Towards a Framework for Comparing Process Modeling Languages. In *Proceedings of the 14th International Conference of Advanced Information Systems Engineering*. (Toronto, Canada, May 27 - 31, 2002). CAiSE 2002. LNCS Vol. 2348, 600-611. Springer, Berlin.

[25] van der Aalst, W. and van Hee, K. 2004. *Workflow Management: Models, Methods and Systems*. MIT Press, Cambridge.

[26] van der Aalst, W. and ter Hofstede, A. 2010. *Workflow Patterns Home Page* http://www.workflowpatterns.com/patterns/index.php.

[27] Weske, M. 2007. *Business Process Management: Concepts, Languages, Architectures*. Springer, Berlin.

[28] The Workflow Management Coalition (WfMC). 1995. *The Workflow Reference Model*, Document Number TC00-1003, Issue 1.1., 19 November 95. http://www.wfmc.org/standards/docs/tc003v11.pdf.

[29] The Workflow Management Coalition (WfMC). 1999. *Terminology & Glossary*, Document Number WFMC-TC-1011, Issue 3.0, February 99. http://www.wfmc.org/standards/docs/TC-1011_term_glossary_v3.pdf.

[30] The Workflow Management Coalition (WfMC). 2010. *Process Definition Interface - XML Process Definition Language (XPDL)*, Document Number WFMC-TC-1025, 10 October, 2008, Version 2.1a. http://www.wfmc.org/xpdl.html.

[31] Wolf, C. and Harmon, P. 2010. *The State of Business Process Management 2010*. BPTrends Reports, February 2010. http://www.bptrends.com/surveys_landing.cfm.

[32] zur Muehlen, M., Recker, J. and Indulska, M. 2008. How Much Language is Enough? Theoretical and Practical Use of the Business Process Modeling Notation. In *Proceedings of the 20th International Conference of Advanced Information Systems Engineering*. (Montpellier, France, June 16 - 17, 2008). CAiSE 2008. LNCS Vol. 5074, 465-479. Springer, Berlin. DOI= http://dx.doi.org/10.1007/978-3-540-69534-9_35