Evaluating Business Process Maturity Models

Amy Van Looy

Faculty of Economics and Business Administration, Department of Business Informatics & Operations Management, Ghent University (Belgium), amy.vanlooy@UGent.be

Geert Poels

Faculty of Economics and Business Administration, Department of Management Information Science and Operations Management, Ghent University (Belgium), Geert.Poels@UGent.be

Monique Snoeck

Department of Decision Sciences and Information Management, KU Leuven, Leuven (Belgium), Monique.Snoeck@kuleuven.be

PREPRINT: This paper has been accepted for the Journal of the Association for Information Systems - Manuscript ID RA-JAIS-15-0068.R5

Please cite as follows:

Van Looy A, Poels, G, Snoeck M, 2016, Evaluating Business Process Maturity Models, Journal of the AIS, vol. accepted.

Evaluating Business Process Maturity Models.

Maturity models have become important aids to support process improvement. The many business process maturity models (BPMMs) are, however, frequently criticized for differing in quality, which creates a demand for comparing and evaluating these models. This paper describes a well-founded, ranked and weighted set of critical criteria for BPMMs that are most important to prospective users. The paper also demonstrates how this set can serve as a standard or guideline for designing BPMMs. The evaluation of the used sample of BPMMs against the set of criteria uncovers gaps and shortcomings of current BPMMs. This allows the identification of implications to raise the quality of BPMMs and highlights future research avenues.

Keywords

Analytical Hierarchy Process; content analysis; Delphi technique; Design Science; evaluation study; maturity model; process improvement; trade-offs

1. Introduction

The present paper investigates maturity models for business processes and Business Process Management (BPM) in general, i.e. a topic that has received considerable attention mainly in practitioner (but less so academic) literature as from the 2000s. In particular, business process maturity models (BPMMs) have become real assets for organizations to increase business (process) performance (Dijkman et al., 2015; McCormack & Johnson, 2001; Skrinjar, Bosilj-Vuksic & Stemberger, 2008). BPMMs originate from quality control models (Crosby, 1979), which have been adapted for software processes (e.g. CMM, CMMI) (Ahern, Clouse & Turner, 2004) and afterwards business processes and BPM. Their practical relevance is widely recognized (Harmon, 2013; Harmon & Wolf, 2014), because improving business processes (or the organizational way of working) increases in importance due to globalization, compliance requirements, IT opportunities, outsourcing, etc. (Harrington, 2006; vom Brocke & Rosemann, 2010). In response to this high demand (Harmon, 2013; Harmon & Wolf, 2014), many scholars, institutions and consultancy firms have developed a BPMM.

Although concrete numbers are missing, estimates mention more than 200 process improvement frameworks (among others BPMMs and standards) (Curtis & Alden, 2007; El Emam & Birk, 2000) or over 150 BPMMs addressing one or more BPM areas (de Bruin and Rosemann, 2007). Inspired by Sheard (2001), we may accurately refer to a quagmire in which practitioners tend to drown. Examples of BPMMs that are frequently cited in the literature are CMMI (Ahern, Clouse & Turner, 2004; SEI, 2009), OMG (2008), the models of de Bruin and Rosemann (2007), Hammer (2007), Harrington (2006), and McCormack and Johnson (2001). We illustrate the notion of business process maturity by means of the maturity levels of OMG (2008), which concerns a generic BPMM that follows the CMMI tradition and is supported by a large industry consortium (Curtis & Alden, 2007). OMG's maturity levels range from (1) initial or ad hoc practices, to (2) managed, departmental practices, (3) standardized, end-to-end practices, (4) predictable or quantitatively managed practices, and (5) innovating practices.

Notwithstanding the high demand for BPMMs (Harmon, 2013; Harmon & Wolf, 2014) and the many BPMMs nowadays (Sheard, 2001; de Bruin and Rosemann, 2007), few studies take the point of view of prospective users who face the challenge of evaluating and selecting one out of many BPMMs for their organization. Instead, the literature on BPMMs is mainly restricted to the perspective on how maturity models should be designed (Becker, Knackstedt & Pöppelbuss, 2009; Maier, Moultrie & Clarkson, 2012) or by designing BPMMs accordingly (de Bruin & Rosemann, 2007). The literature still lacks a set of general selection criteria for BPMMs that transcends the needs of a particular organization to evaluate the strengths and weaknesses in the many BPMMs of today. The latter is of paramount importance as maturity models are typically criticized for oversimplifying the complex reality and differing in quality (Röglinger et al., 2012; Maier, Moultrie & Clarkson, 2009). Yet, the perceived utility of BPMMs in industry is high (Harmon, 2013; Harmon & Wolf, 2014), and scholars have empirically demonstrated a positive relationship between business process maturity and business performance (McCormack & Johnson, 2001; Skrinjar, Bosilj-Vuksic &

Stemberger, 2008). A set of general selection criteria would not only address a contemporary business problem, but also has the opportunity to shed more light on BPMM designs by adding the missing angle of prospective users.

To fill this gap, the current paper develops a set of critical factors or criteria for selecting a BPMM, examines the relative importance of the different criteria, and evaluates existing BPMMs against this set. As such, this paper investigates two research questions to introduce a user perspective to BPMM research, which currently focuses only on the designer perspective:

- RQ1. Which criteria are most relevant for BPMM selection, and what is their relative importance?
- RQ2. How can current BPMMs be evaluated against these selection criteria?

The corresponding research objectives are:

- Developing a comprehensive, ranked and weighted set of selection criteria for BPMMs that are not specific to any organization.
- Evaluating existing BPMMs against this set of selection criteria.

This study introduces the notion of quality of BPMMs from a user perspective as meeting end user expectations, and evaluates the extent to which a representative sample of 69 BPMMs meets the identified selection criteria. By investigating the research questions, the current paper can give evidence of the varying quality among BPMMs, however, without focusing on individual BPMMs. Our user perspective is complementary to the designer perspective in BPMM research as the absence of certain model characteristics that are needed for the user evaluation of BPMMs may inform proposals for improving model designs. Our empirical and interpretive study can be characterized as a meta-study (Bostrom, Gupta & Thomas, 2009; Zhao, 1991) because it intends to generalize differences between BPMMs based on an in-depth analysis and discussion. It does so by synthesizing knowledge of current BPMMs, detecting areas for improvement, and laying a foundation for further BPMM research based on the identified shortcomings. This paper thus makes a novel academic contribution as both

the BPMM literature and particular models could benefit from issues about relevant variances of BPMMs.

We proceed with a review of the research on maturity models in section 2. We then describe the research methods (section 3). The subsequent sections present our data analysis and results. In particular, Section 4 explores the initial criteria for BPMM selection, while section 5 works towards a ranked and weighted set of criteria. Afterwards, in section 6, we apply the findings to existing BPMMs in order to conduct a quality check on their coverage of elements required to evaluate the identified selection criteria. Section 7 discusses the results and presents research implications, followed by conclusions in section 8.

2. Literature review

Although the number of academic publications on maturity models is reasonable (i.e. about 332 articles in the Web of Science until 2015), still a small subset of this literature addresses maturity models for business processes. The found papers mostly focus on particular maturity models (e.g. for project management, knowledge management, business-IT alignment, or specific process types such as software processes) or on model development (see next paragraphs). Papers reporting on maturity model evaluation address the validation or application of a particular maturity model in a specific situation, and provide an assessment or evaluation of the application results rather than a comparison or quality check of existing maturity models (as targeted in this study). Similarly, previous studies used the term ‘selection’ to refer to the choice of specific application areas or case situations rather than the choice for one or another maturity model (which is the user perspective addressed in our study). Thus, the academic literature on the evaluation or selection of BPMMs is still very scarce. While Wendler (2012) confirms that especially reflective publications with theoretical implications on maturity models are scarce and a gap exists in evaluating those models, her literature review is not specific to BPMMs nor to the perspective of prospective model users. In (Blinded1), a decision tool for selecting one BPMM out of a large sample of

existing BPMMs is presented, albeit without generalization of the substantial differences between these BPMMs as aimed in this study.

Röglinger, Pöppelbuss and Becker (2012) propose design criteria specifically for BPMMs and subsequently present a limited BPMM comparison, but without giving advice on how to choose a BPMM that fits a particular organization. Mettler (2009) offers a first attempt to translate design criteria into a user perspective. He does so by considering maturity models in general (i.e. independent of the specific business process context), and without evaluating existing maturity models. Similar to Mettler (2009), it is interesting to look at the design science literature and verify the extent to which the design criteria for BPMMs can be translated to a user perspective.

Particularly, maturity models can be built and tested by following a specific design research cycle (Becker et al., 2009; Mettler & Rohner, 2009), in which each phase should meet specific evaluation guidelines (Becker et al., 2009; Hevner, March, Park and Ram, 2004). Referring to the March and Smith (1995) categorization of design artifacts, Mettler and Rohner (2009) and Donnellan and Helfert (2010) conclude that maturity models generally offer **models** (e.g. maturity levels), **methods** (e.g. best practices to achieve higher levels) and **instantiations** (e.g. documents or websites to be used by organizations). Nonetheless, the literature still lacks a common conceptualization with **constructs** for designing maturity models in general and BPMMs in particular (Mettler & Rohner, 2009).

In order to learn more about the **constructs** that typify maturity models, we launched a search query by combining the keywords “maturity model” and “design science” in 2011. Because the few resulting papers in the Web of Science database mainly involved specific maturity models, except for (Becker et al., 2009; Tapia et al., 2008; van Steenbergen et al., 2010), we did a similar search in Google Scholar in order to find more relevant studies (de Bruin et al., 2005; Maier et al., 2009; Mettler et al., 2009). The resulting papers were analyzed by means of a content analysis that focused on finding common elements or characteristics in the design of maturity models, as illustrated in Table 1.

Table 1. Common elements in the design of maturity models (until 2011, i.e. before our study).

	Becker et al (2009)	de Bruin et al (2005)	Maier et al (2009)	Mettler et al (2009)	Tapia et al (2008)	van Steen-bergen et al (2010)
Assessors – WHO						
*** Lead assessor**	X	X	X	-	X	X
*** Other** assessors and respondents	X	X	X	X	X	X
Assessment method – HOW						
*** Data collection technique** to obtain information to assess	X	X	X	X	-	X
*** Calculation** to interpret the collected data as lifecycle levels	X	X	X	-	X	X
*** Representation** to visualize lifecycle levels	-	X	X	X	X	X
Improvement method – WHAT						
*** Capability areas** to be assessed and improved	X	X	X	X	X	X
*** Lifecycle levels**	X	X	X	X	X	X
*** Architecture** (i.e. road map or improvement path) to link capability areas to levels by step-by-step improvements	X	X	X	X	X	X

Table 1 shows that three original groups of functionalities seem to be present in maturity model design (Becker et al., 2009; de Bruin et al., 2005; Maier et al., 2009; Mettler et al., 2009; Tapia et al., 2008; van Steenbergen et al., 2010). In particular, Table 1 reveals that a maturity model may act as an assessment method and an improvement method, and may involve elements referring to the assessors as a third group. This third original group ('Assessors') covers all design elements related to **who** measures maturity, while the 'Assessment method' specifies **how** maturity is measured (e.g. how data is collected and analyzed) and the 'Improvement method' groups elements describing **what** is measured as maturity (e.g. the capability areas and their improvements necessary to reach each consecutive level).

Since the papers in Table 1 are not specific to the context of a BPMM, we verified the extent to which the common design elements and original groups of functionalities are mentioned in the definition of a BPMM. We therefore relied on a comprehensive definition that was derived from a comparison of current BPMM definitions (Blinded2: pp. 1132-1133), in which a BPMM is defined as '*a model to assess and/or to guide best practice improvements in organizational maturity and process capability, expressed in lifecycle levels, by taking into account an evolutionary road map regarding (1) process modeling, (2) process deployment, (3) process optimization, (4) process management, (5) the organizational culture, and/or (6) the organizational structure*'. This BPMM definition explicitly refers to two important functionalities, namely by defining a BPMM as an assessment method (i.e. '*to assess*') and/or as an improvement method ('*to guide best practice improvements*'). The specification of the six elements in the definition (i.e. process modeling, deployment, optimization, management, culture and/or structure) is a translation of the capability areas to be assessed and improved by a BPMM. Furthermore, the keywords in this definition, like 'maturity lifecycle levels' and 'capability lifecycle levels', may serve as major constructs in a common conceptualization for BPMMs.

Despite the fact that the identified common elements are mentioned in most of the papers of Table 1 and that some of them are addressed in the definition of a BPMM (Blinded2: pp. 1132-1133), not all common design elements are detailed in concrete options nor translated to the context of BPMMs. This paper takes the perspective of prospective model users, as a new angle in the literature, to develop such conceptualization required to compare and evaluate different BPMMs and to identify strengths and weaknesses in existing BPMMs.

3. Research method

The current paper presents a meta-study (Bostrom, Gupta & Thomas, 2009; Zhao, 1991) by producing detailed and standardized constructs allowing an objective comparison and evaluation of existing BPMMs. A meta-study typically examines a particular problem encountered in a discipline, such as the increasing number of BPMMs without evaluation

from a user perspective, and ends with prescriptions for resolving the problem. More specifically, the paper concerns a meta-data-analysis, since it is based on empirical research and analyzes a large collection of individual BPMMs for the purpose of understanding and integrating the findings (Zhao, 1991).

Our overall approach is summarized in Figure 1, indicating different research phases with inputs and outputs.

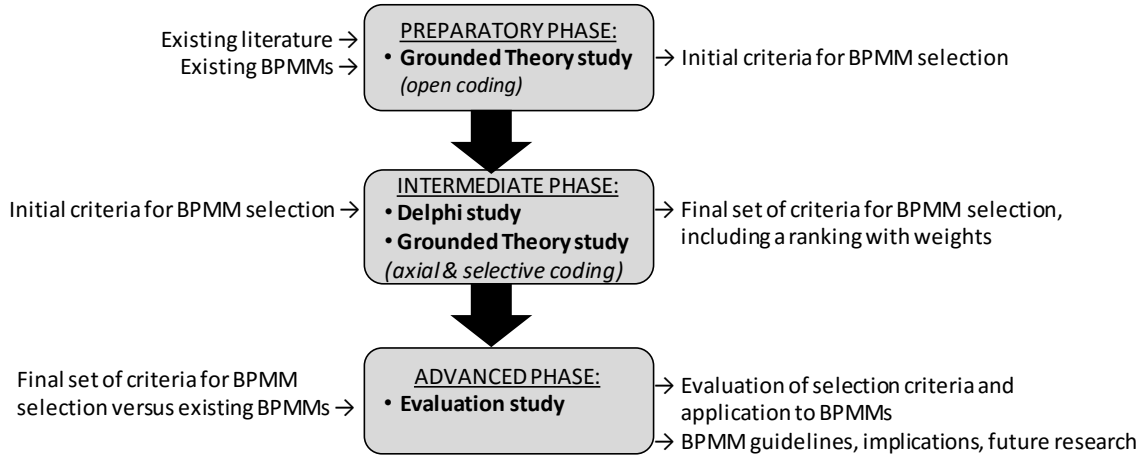

Figure 1. The research approach for this paper.

In particular, our research approach combined recognized methods in order to drive research rigor: (1) Grounded Theory and (2) Delphi. First, a Grounded Theory study (Glaser & Strauss, 1967) was used to explore the problem situation. It particularly aimed at thoroughly familiarizing ourselves with the phenomenon of interest. Further on, a Delphi study (Dalkey & Helmer, 1963), extended by the Analytical Hierarchy Process or AHP (Saaty, 1990), was used for problem-solving. More specifically, the Delphi study allowed identifying the most important (i.e. ranked and weighted) criteria for BPMM selection based on consensus-seeking and multi-criteria decision-making (RQ1). In order to respond to RQ2, a sample of BPMMs was evaluated against this general set of selection criteria (i.e. with a case-independent ranking and weights) to shed light on the degree to which these criteria can be applied to current BPMMs.

3.1. Grounded Theory study

We conducted a content analysis of existing BPMMs in line with the Grounded Theory (Glaser & Strauss, 1967), since the literature reviewed in section 2 largely focuses on maturity models in general, and few papers address maturity models for business processes in particular. Moreover, this way of working allowed us to specify the concrete options per common design element. We analyzed the descriptions in the available documentation (i.e. manuals or descriptive documents) of a representative sample of 69 BPMMs (Blinded1). Thirty-two of these BPMMs address specific process types (24 for supply chains and 8 for collaboration processes to include end-to-end value chains), whereas 37 BPMMs are generic and can be applied to any process type. The BPMMs in this sample were identified by searching in the Web of Science, the BPM Journal, the Google and Google Scholar search engines with the combined keywords of 'process' and 'maturity'. Also the references in the identified papers were traced for other relevant sources (i.e. a snowballing search strategy). Models were included if they: (1) contain maturity levels and/or capability levels, and (2) primarily focus on improving generic business processes, supply chains or collaboration processes instead of other organizational assets. Sampling was done in 2010, and its comprehensiveness was confirmed in 2012 when a resampling effort did not result in additional BPMMs (taking into account some limitations regarding the accessibility of documents, the language, and the keywords). By including different process types (i.e. generic, supply chains and collaboration), the sample suggests versatility to enable transferability of the findings to other process types (e.g. software processes).

Conform to the Grounded Theory terminology (Glaser & Strauss, 1967), the existing BPMMs were analyzed to identify model characteristics as possible selection criteria and options (i.e. values per criterion) through three coding stages, namely (1) initial (open) coding, (2) intermediate (axial) coding, and (3) advanced (selective) coding. Table 2 shows a coding example to illustrate how an extract from BPMM descriptions led to the induction of a pattern, with codes representing possible selection criteria with different options.

Table 2. A coding example.

Extract from BPMM descriptions	Codes
"Conformance with the BPMM is evaluated in appraisals led by an authorized Lead Appraiser, who has been trained extensively."	[Assessment] [Assessor Type - Lead assessor] [Training]
"Conformance is evaluated by using the following forms of evidence: review of artifacts, interviews, and quantitative data describing the performance of a process, its outcomes, and business results."	[Data collection techniques – document review] [Data collection techniques – interview] [Rating scale – qualitative data] [Rating scale – quantitative data]

First, during initial (open) coding, we read the collected documentation of existing BPMMs by going back and forth to identify possible criteria and options. For instance, different BPMMs mentioned that maturity can be measured by collecting data from document reviews, questionnaires, interviews, focus groups or observations. These different options were grouped in a criterion referring to 'data collection techniques'. As another example, different BPMMs presented information on the assessment items that are used to collect data, i.e. open questions or questions with nominal, ordinal, discrete, interval and ratio scales. Hence, 'rating scale' was coded as another possible criterion. After this stage, the Delphi study started to discuss and find potential criteria and options.

Secondly, during intermediate (axial) coding, the criteria and options obtained through the initial coding were reconsidered based on the feedback from the Delphi study, and linked to the common design elements of maturity models (section 2). For instance, the criterion 'data collection techniques' was linked to the assessment method, and rearranged as 'objective' and 'subjective' data collection techniques. Also the criterion 'rating scale' was linked to the assessment method and rearranged as 'qualitative' data and 'quantitative' data.

Afterwards, during advanced (selective) coding, we reread the documentation of existing BPMMs to record how these models actually cover the identified criteria and options. We note that this final coding stage was executed once the final set of selection criteria in the Delphi study was known.

3.2. Delphi study

During the Delphi study, we aimed at validating and extending the set of criteria from the first research phase in order to obtain a comprehensive, ranked and weighted set of selection criteria. To introduce a user perspective, the initial criteria were supplemented by criteria obtained from peer feedback at a conference on enterprise information systems and a pilot study with Business Process Management (BPM) scholars at our faculty. Regarding the pilot, two PhD students working on BPM research were engaged to focus on the formulations of the criteria, whereas two BPM professors also checked for missing criteria and biases. The resulting set of criteria was reviewed by independent subject-matter experts. The experts were also asked for additional criteria that are relevant for BPMM selection in an international Modified Delphi study (Table 3).

Table 3. The different Delphi types, based on Hasson, Keeney & McKenna (2000).

Delphi Type	Purpose
Classical Delphi	For consensus-building with anonymity, starting with open questions
Forecast Delphi	Classical Delphi to combine opinions on trends, i.e. the likelihood and time scale of developments in science, technology, business, etc.
Real-time Delphi	Classical Delphi with real-time calculation and aggregation of group responses (i.e. online Delphi conference)
Modified Delphi	For consensus-building with anonymity, also including closed questions to orient people to the topic
Decision Delphi	For consensus-building on social developments with quasi-anonymity (i.e. experts are mentioned by name, but answers remain anonymous)
Policy Delphi	For dissensus-building with anonymity, to elicit opposing views or alternatives

A Delphi study is an established consensus-seeking decision-making method using '*a series of sequential questionnaires or rounds, interspersed by controlled feedback, that seek to gain the most reliable consensus of opinion of an expert panel*' (Dalkey & Helmer, 1963: pp. 458). Different types of Delphi studies exist (see Table 3). We opted for the Modified Delphi approach because we included the criteria from the Grounded Theory study to orient diverse subject-matter experts to the research topic.

In November 2011, the Delphi study started with 22 BPM experts: 11 academics and 11 practitioners from five continents. The academics had credible BPM(M) publications in

academic journals, and the practitioners had experience in designing a BPMM, applying BPM(M), or were genuinely interested in BPMM selection. All practitioners were consultants or managers with decision power in large to medium-sized organizations, both profit and non-profit. The selection procedure that we applied conforms to (Okoli & Pawlowski, 2004), introducing different backgrounds to minimize bias and allow normative discussions. Per round, the responses were anonymously analyzed by four coders, of which one independent coder was from another university.

The coders stopped iterating a particular criterion when one of the three stopping conditions was reached (Hasson et al., 2000; Okoli & Pawlowski, 2004): (1) consensus was reached to include a criterion for BPMM selection, (2) results for the criterion became stable before reaching consensus, to exclude a criterion, or (3) the majority of experts were no longer willing to continue iterating, to exclude all remaining criteria. Consensus conditions were defined for a 7-point Likert scale (Hasson et al., 2000): (1) 50% of the experts must agree on the two most extreme scores (i.e. 6-7), (2) 75% must agree on the three most extreme scores (i.e. 5-6-7), (3) the interquartile range must be 1.50 or less, and (4) no opposite extreme score given by any expert (i.e. 1). Stability was measured by Spearman's rho, Kendall's tau-b, and the Cohen's Kappa level of agreement on a recoded 3-point scale to examine opinion changes (i.e. between unimportance 1-2-3, neutral 4 and importance 5-6-7).

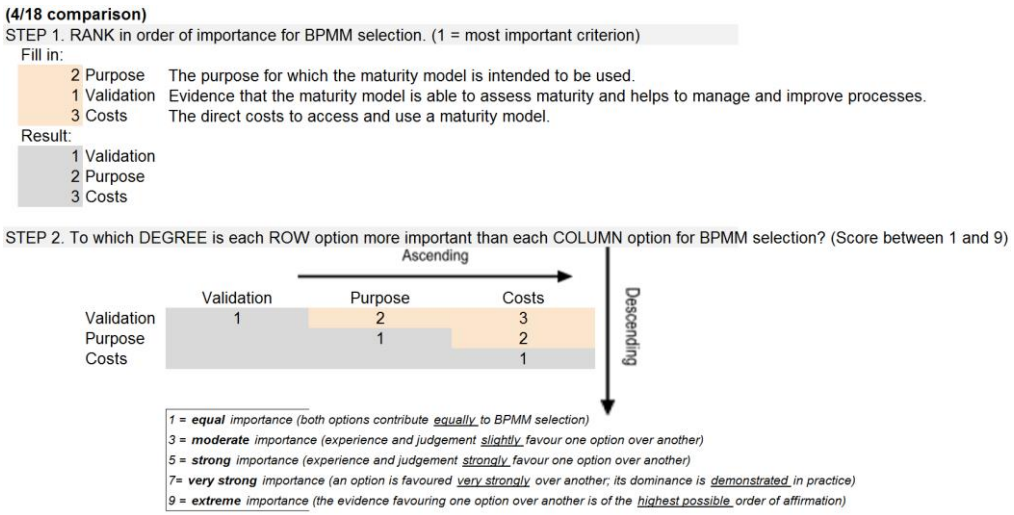

Figure 2. Extract from a completed Delphi questionnaire (round 4) with judgment matrix.

Once the selection criteria with consensus were elicited, the experts were asked to determine which are more important in pairwise comparisons. Therefore, we applied the Analytical Hierarchy Process (AHP) (Saaty, 1990). The experts were asked to complete judgment matrices by using the typical AHP 9-point scale (1/9=extremely less important; 1=equally important; 9=extremely more important) to describe how much more important each row item is compared to each column item, as exemplified in Figure 2. AHP then calculates a priority vector (or principal Eigen vector) and a consistency ratio (CR) per matrix. Afterwards, the expert opinions were aggregated by geometrically averaging only non-random judgments (CR≤0.1). As such, the aggregated priorities or relative importance of selection criteria and their options were obtained.

3.3. Evaluation study

Thirdly, the evaluation study aimed at demonstrating the applicability of the developed set of selection criteria to compare and evaluate BPMMs by means of a general ranking with weights. The set of criteria was evaluated for its utility, effectiveness, and efficiency based on predefined requirements, as shown in Table 4.

Table 4. The requirements for the developed set of selection criteria for BPMMs.

	Requirements	Requirement satisfaction tests
Utility	The ranked and weighted set of selection criteria should enable a comparison and evaluation of BPMMs.	• **Criteria utility.** Based on a textual description of the selection criteria, other BPMMs than those involved in our sample can be evaluated by people not involved in our research. The Cohen's Kappa represents a significant level of agreement or interrater reliability (0.4=<kappa=<1; p<0.05). • **Ranking utility.** Based on descriptive statistics, comparative tables can be generated to visualize the weights of individual selection criteria along with the support in current BPMMs. • **Clarity.** In the comparative tables, the main differences between BPMMs can be highlighted with respect to the ranked and weighted selection criteria.
Effectiveness	The ranked and weighted set of selection criteria should represent relevant BPMM characteristics that prospective BPMM users	Prospective users are satisfied with (1) the selection criteria in the developed set, (2) their descriptions of trade-offs, and (3) the sequence in which they appear based on a

	consider when selecting a BPMM.	general ranking with weights (i.e. 50% for scores 5-6-7 on a 7-point Likert scale, and no opposite extreme score of 1).
Efficiency	The ranked and weighted set of selection criteria should lead to an easier comparison and evaluation of BPMMs compared with an ad hoc way of working.	Based on the comparative tables, relevant differences among BPMMs can be detected by following a standard way of working.

To evaluate the applicability of the selection criteria, existing BPMMs (i.e. input of the first research phase) were evaluated against the ranked set of selection criteria for BPMMs (i.e. output of the second research phase). At the same time, this evaluation provides information about the fit of each BPMM with the selection criteria. The size of the sample including 69 maturity models allowed us to compare and evaluate a comprehensive and unique dataset and allowed drawing general conclusions.

The fact that the Delphi study and the content analysis were performed by the same researchers introduced a danger of bias for the content analysis. To avoid the content analysis being biased by knowledge about the ranking and weighting of the selection criteria, the content analysis was performed before finalization of the Delphi study. Reversely, results from the content analysis were kept hidden from the experts participating in the Delphi study.

4. Preparatory phase: initial criteria for BPMM selection

The criteria that resulted from the Grounded Theory study (as explained in section 3.1) are shown in Table 5. These criteria were used as input for the Delphi study to have their rank determined and their trade-offs discussed by the expert panel. Table 5 also indicates which criteria were proposed by the experts in addition to the criteria that resulted from the Grounded Theory study (see next section).

Table 5. The list of all criteria considered in the Delphi study.

Criterion	Criterion description
(1) Number of assessed organizations	The number of organizations (i.e. autonomous legal entities) that are included in the assessment.
(2) Lead assessor	Whether the assessment is led by an external (quasi-) independent person, i.e. third party.
(3) Number of assessors	The number of assessors who are required to conduct an assessment.

(4) Functional role of respondents	The explicit recognition to include people from outside the assessed organizations as respondents.
(5) Business versus IT respondents	The explicit recognition to include IT people and/or business people as respondents in the assessment.
(6) Data collection technique	The way information is collected during an assessment.
(7) Number of assessment items	The maximal number of questions to be answered during an assessment.
(8) Assessment duration	The maximal duration of a particular assessment.
(9) Rating scale	The type of data that is collected during an assessment.
(10) Presence of capabilities	The capabilities to be assessed and improved.
(11) Number of business processes	The number of business processes to be assessed and improved: (1) one, (2) more or (3) all.
(12) Type of business processes	Whether the BPMM is generic (i.e. for business processes in general) or domain-specific (e.g. for business processes in supply chains or collaboration situations).
(13) Architecture type	The possibility to define a road map per capability, a road map for overall maturity, or both (i.e. a staged architecture with maturity levels and/or a continuous architecture with capability levels).
(14) Number of lifecycle levels (i.e. maturity levels or capability levels)	The number of maturity levels or capability levels that are defined.
(15) Level calculation	The way the resulting maturity levels or capability levels are calculated.
(16) Level representation	The way the resulting maturity levels or capability levels are displayed.
(17) Labeling of levels	The way maturity levels or capability levels are labelled, i.e. what they indicate or explicitly refer to.
(18) External view of levels	The extent to which maturity levels or capability levels take into account possible relationships between individual organizations.
(19) Architecture details	The degree of guidance that a BPMM gives on your journey towards higher maturity (i.e. with descriptive or prescriptive improvements).
(20) Methodology used to create a BPMM (p)	The way the BPMM was created.
(21) Methodology used to validate a BPMM (p)	Whether or not empirical evidence is given that the BPMM helps to enhance the efficiency and effectiveness of business processes.
(22) Direct costs to access and use a BPMM (p)	The direct costs to access and use a BPMM.
(23) Purpose of BPMM use (*)	The purpose for which a BPMM is intended to be used (i.e. only raising awareness or also benchmarking or certification).
(24) Assessment availability (*)	Whether the assessment questions and corresponding level calculation are publicly available (instead of only known to the assessors).

(*) Additional criteria proposed by the experts; (p) Criteria obtained from prior peer feedback and pilot study

5. Intermediate phase: ranked and weighted set of BPMM selection criteria

The Delphi study aimed at building consensus in order to validate and extend the set of criteria from the first research phase. After three rounds, we stopped iterating as 75% of the experts indicated they were no longer willing to continue iterating (Figure 3). This is similar to other Delphi studies, which typically take three to four rounds (de Bruin & Rosemann, 2007; Hasson et al., 2000). On the right, Figure 3 shows that 14 out of 24 criteria reached consensus of being important for BPMM selection. They constitute our final set of selection criteria for BPMM. These selection criteria and their trade-offs (derived from the Delphi study) are detailed in Appendix 1. Other criteria had no trend towards consensus due to condition 4 (i.e. at least one expert with an opposite extreme score in multiple rounds). In all rounds, the response rates exceeded the minimum value of 70%, enhancing research rigor and validity (Hasson et al., 2000). Moreover, 95% of the respondents in the third round (N=17) agreed that the set of final selection criteria is very to extremely important for BPMM selection (i.e. scores 6 or 7; median = 6; interquartile range = 0).

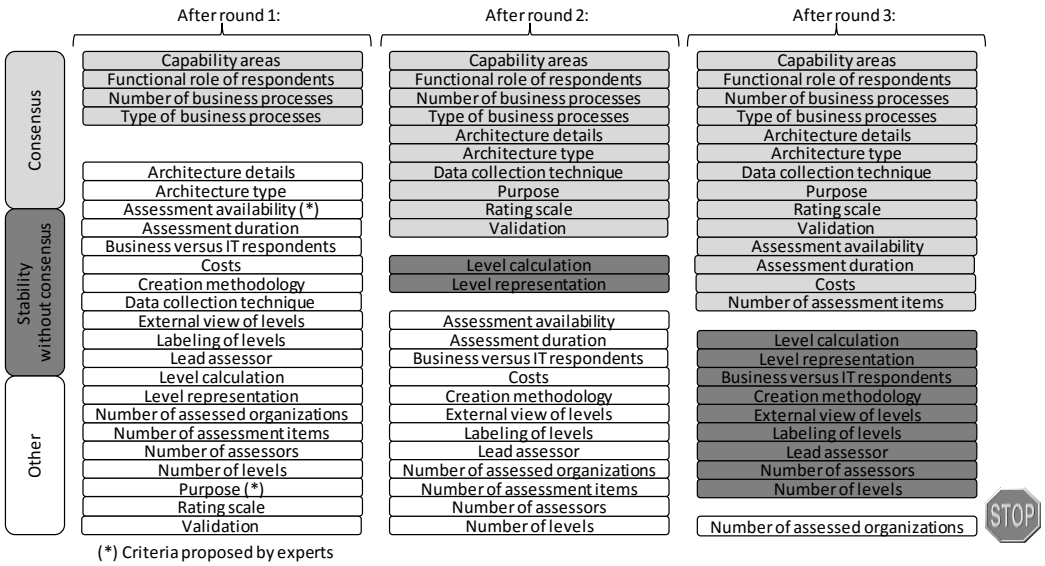

Figure 3. An overview of the criteria throughout the consensus-seeking Delphi rounds.

The 14 selection criteria that reached consensus were grouped and weighted in a fourth Delphi round to obtain a ranked set. AHP uses a hierarchical approach to first assign weights to groups of criteria, then to the criteria per group, and afterwards to the options per criterion. Section 2 already mentioned three original groups in the literature on maturity models, i.e.

‘Assessors’, ‘Assessment method’, and ‘Improvement method’. Nonetheless, since most criteria regarding the assessors and respondents did not reach consensus of being most important for BPMM selection, the remaining criterion (i.e. ‘functional role of the respondents’) of the group of ‘Assessors’ was added to the group of ‘Assessment method’. This choice is justified as AHP requires groups with multiple criteria, and preferably of almost similar sizes to avoid a bias or overestimation. Hence, the ‘Assessment method’ and the ‘Improvement method’ were chosen as possible groups for our study. Some selection criteria could not be classified in the ‘Assessment method’ or the ‘Improvement method’, but rather involve general or practical considerations of a BPMM (e.g. the costs to access and use a BPMM, or the degree to which a BPMM has been validated). The latter concern criteria that were grouped as ‘Contextual’ criteria (i.e. in the sense that they are rather model-independent because they do not belong to the assessment method or improvement method of a BPMM). Hence, the remaining selection criteria were grouped as follows (Figure 4): (1) assessment method criteria, i.e. how maturity is measured and by whom, (2) improvement method criteria, i.e. what is measured as maturity, particularly the capability areas and their improvements to reach successive levels, and (3) contextual criteria (e.g. costs).

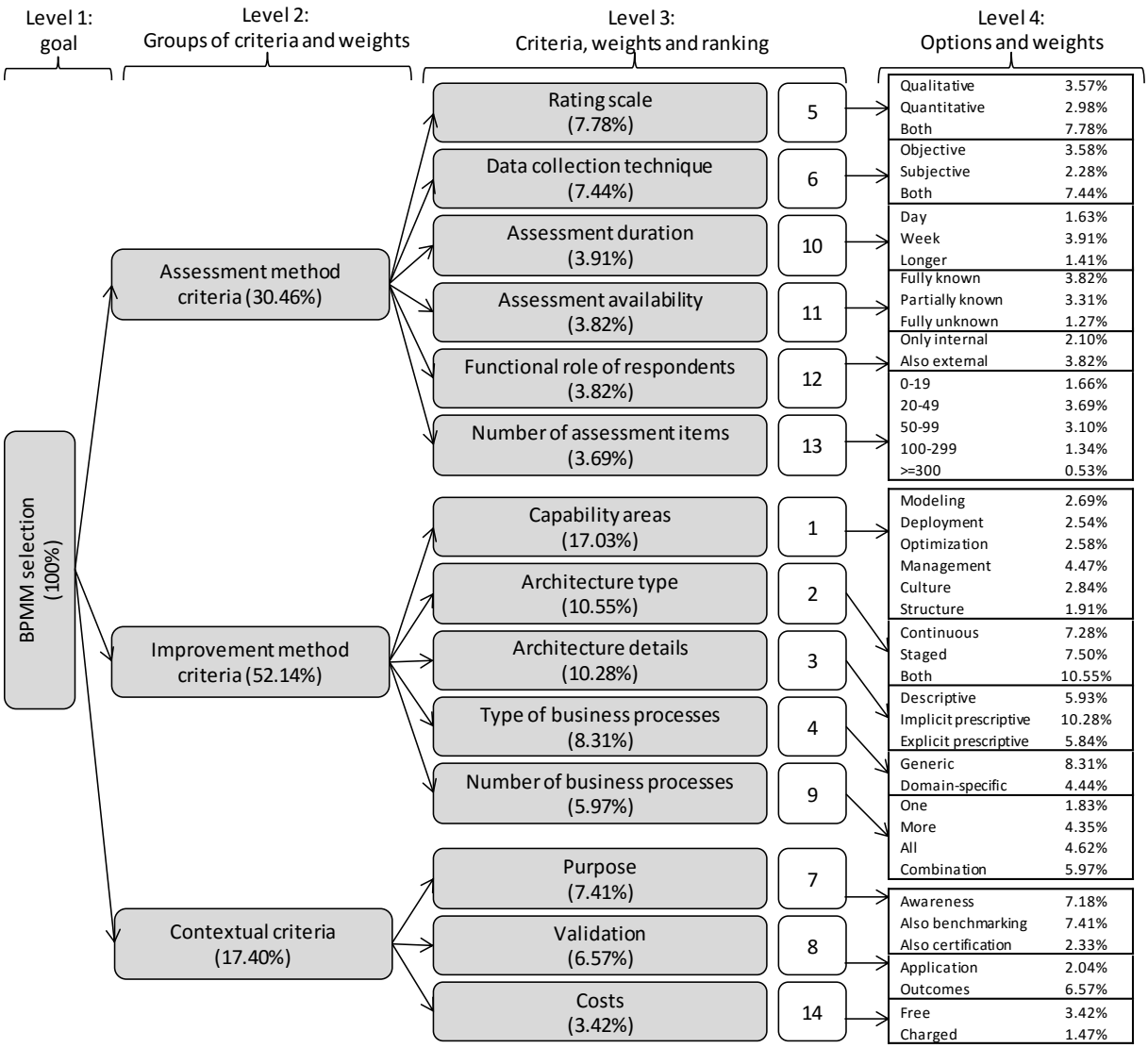

Figure 4. The hierarchical AHP model for BPMM selection.

The resulting weights are shown in Figure 4. The experts argued that the improvement method criteria should be more decisive for BPMM selection (i.e. representing a higher overall weight in Figure 4) than the assessment method criteria or the contextual criteria. This finding reflects that levels given by BPMMs are not an end goal, but capability improvements and performance improvements are. Similarly, organizations that merely strive for the highest (not optimal) levels are rather misusing BPMMs (Dijkman et al., 2015; Maier et al., 2009). Hence, the 'capability areas' criterion received the highest weight as it ultimately represents what is being measured and improved. The 'costs' criterion has the lowest weight to avoid an organization from selecting a free model that measures the wrong scope of capabilities for that organization (and becomes useless as such).

In a fifth and final Delphi round, 95% of the respondents (N=20) were satisfied with the obtained weights (i.e. scores 5, 6 or 7; median = 6; interquartile range = 0). Consequently, the vast majority confirmed the resulting ranking and weighting by AHP.

In the previous paragraphs, we explained how we managed the internal validity and objectivity for identifying and weighting selection criteria in the best possible way. Due to the small sample size, Delphi results cannot be repeated. Nevertheless, reliability was taken care of by careful expert selection and multiple coders. In order to further enhance external validity, our methodology could be repeated, e.g. for other types of maturity models.

6. Advanced phase: evaluation of the selection criteria and application to existing BPMMs

6.1. Evaluation of the selection criteria

The evaluation study aimed at demonstrating the applicability of the ranked and weighted set of criteria for comparing and evaluating BPMMs, based on the requirements of Table 4.

First, for evaluating the requirement of 'criteria utility', we asked two former master students who recently graduated in business informatics to look for additional BPMMs by using the same search query as described in section 3.1. The students were unbiased in the sense that they were unfamiliar with our research on business process maturity models and were only provided with the list of 69 BPMMs that we previously found and used in our research. In order to control for possible biases based on the origin of maturity models, we asked one student to look for an additional BPMM in an academic database (e.g. Web of Science or Google Scholar) and the other student to search for a BPMM created by consultants and described on the WWW. Then, we asked them to individually evaluate each of the two newly found BPMMs (i.e. based on the documents found) using the descriptions of the selection criteria and their options as given in Appendix 1. The students mapped both BPMMs to the criteria in a similar way for 10 out of 14 criteria for the academic model, and for 12 out of 14

criteria for the non-academic model (0.4<kappa=0.571<1; p=0.002<0.05), suggesting that our way of working can be repeated for other BPMMs and by other people.

Regarding the 'effectiveness' requirement, we asked seven prospective BPMM users to rate their satisfaction with the criteria, the descriptions of trade-offs between options of criteria, and the sequence in which the criteria and their options appear (as an equivalent to the ranking and weights) on a 7-point Likert scale (1 = very dissatisfied; 7 = very satisfied). The users were strategically chosen to cover different business scenarios, namely:

- four practitioners who enrolled for the BPM course of a postgraduate training program and who represented different organizational sizes (micro, small, medium, large);
- two managers who wanted to use a BPMM in their organization and who represented different organizational sectors (non-profit, private);
- one scholar who wanted to use a BPMM in her research.

It turned out that five out of seven prospective users were satisfied with the selection criteria, their descriptions and sequence (i.e. scores 5, 6 or 7; median = 6; min.=4; max.=7). The two remaining users gave a neutral score of 4 because they deemed themselves too unexperienced as practitioners to evaluate the set of criteria, while all other prospective users gave positive scores. This finding suggests that the selection criteria should rather be used by managerial roles.

The other requirement satisfaction tests (i.e. for 'ranking utility', 'clarity, and' 'efficiency') of Table 4 are demonstrated by applying the selection criteria to the sampled BPMMs (N=69), as described in the next section.

6.2. Application of the selection criteria to BPMMs

We mapped each of the BPMMs from our sample (N=69) to the selection criteria, and investigated to what extent the importance attributed to individual criteria corresponds to the effective support in the BPMMs. Effective support means that the elements important to the evaluation are covered by current BPMMs, and that missing elements correspond to what

experts deem less important. As a result of this evaluation, information on the fit of the BPMMs with the selection criteria was obtained.

This section provides a global overview of the evaluation of the BPMMs against the selection criteria. A detailed report per BPMM is out-of-scope, but can be obtained upon request.

6.2.1. Evaluation of BPMMs against the assessment method criteria

Table 6 summarizes the results of the content analysis (i.e. last column) along the options and the weights for selection criteria pertaining to a maturity assessment. The highest values for weight and number of BPMMs are indicated in **bold** and underlined. When the highest valued option for a criterion does not match the largest number of BPMMs offering this option (i.e. when the bold values are not on the same row), a quality gap for a particular criterion can be uncovered.

Table 6. Selection criteria pertaining to the assessment method of a BPMM.

Assessment method criteria	Options	Weight (%)	Number of BPMMs (N=69)
Rating scale	Qualitative	3.57	**41**
	Quantitative	2.98	1
	Both	**7.78**	14
	(Missing value)		*(13)*
Data collection technique	Objective	3.58	4
	Subjective	2.28	**41**
	Both	**7.44**	11
	(Missing value)		*(13)*
Assessment duration	Day	1.63	**18**
	Week	**3.91**	9
	Longer	1.41	7
	(Missing value)		***(43)***
Assessment availability	Fully known	**3.82**	**31**
	Partially known	**3.31**	23
	Fully unknown	1.27	15
	(Missing value)		*(0)*
Functional role of respondents	Only internal	2.10	**46**
	Also external	**3.82**	9
	(Missing value)		*(14)*
Number of assessment items	0-19	1.66	**17**
	20-49	**3.69**	**16**
	50-99	**3.10**	6
	100-299	1.34	7
	>=300	0.53	3
	(Missing value)		*(20)*

For only two of the six assessment method criteria, the expert preferences for the different options per criterion are almost similar to what the sampled BPMMs cover (i.e. the highest

values in the last two columns of Table 6 cover similar rows). It concerns the criteria called ‘assessment availability’ and ‘number of assessment items’. These two criteria express an appreciation for the assessment availability of most BPMMs, without an excessive list of assessment items. In particular, the two highest weights for the ‘assessment availability’ criterion are addressed by 54 out of 69 sampled BPMMs, while the highest weight of 3.82 is addressed by almost half of the models. Regarding the ‘number of assessment items’ criterion, the two highest weights are addressed by almost a third of the sampled BPMMs, while the three highest weights are addressed by the majority of the evaluated models.

On the other hand, Table 6 shows that the experts prefer a combination of qualitative and quantitative scales, and of objective and subjective data collection techniques to obtain results that are closer to reality. Nevertheless, the large majority of BPMMs only cover qualitative scales with subjective techniques, which are easier for respondents to assess (e.g. the perceived instead of actual performance). Another gap concerns the expected duration of a maturity assessment. The experts prefer an assessment that takes between one day and one week, as it must be seriously undertaken. On the other hand, many existing BPMMs, according to their documentation, allow assessing a business process or a set of business processes within a single day, responding to the busy life of managers. We must, however, note that 43 out of 69 sampled BPMMs do not explicitly mention the expected assessment duration in their documentation, indicating a considerable planning uncertainty for practitioners. Finally, we look at the criterion called ‘functional role of respondents’. While the large majority of BPMMs (i.e. 46 out of 69 models) are restricted to internal respondents, the experts recommend the inclusion of external respondents to allow a complete 360-degree feedback. Information from external parties may add perspectives for (future) cross-organizational collaboration and policy acceptance.

6.2.2. Evaluation of BPMMs against the improvement method criteria

Table 7 highlights the strengths and weaknesses in actual BPMMs according to the selection criteria pertaining to the aspect of ‘Improvement method’ of a BPMM.

Table 7. Selection criteria pertaining to the improvement method of a BPMM.

Improvement method criteria	Options	Weight (%)	Number of BPMMs (N=69)
Capability areas	Modeling	2.69	56
	Deployment	2.54	**68**
	Optimization	2.58	**68**
	Management	**4.47**	**67**
	Culture	2.84	57
	Structure	1.91	30
	(Missing value)		*(0)*
Architecture types	Continuous	7.28	14
	Staged	7.50	**31**
	Both	**10.55**	24
	(Missing value)		*(0)*
Architecture details	Descriptive	5.93	21
	Implicit prescriptive	**10.28**	**30**
	Explicit prescriptive	5.84	18
	(Missing value)		*(0)*
Type of business processes	Generic	**8.31**	**37**
	Domain-specific	4.44	32
	(Missing value)		*(0)*
Number of business processes	One	1.83	7
	More	4.35	**36**
	All	4.62	24
	Combination	**5.97**	2
	(Missing value)		*(0)*

For three of the five improvement method criteria, the expert preferences are almost similar to what BPMMs actually offer (Table 7), namely ‘capability areas’, ‘architecture details’, and ‘type of business processes’. The identified capability areas were grouped according to (Blinded3), namely along the phases of the traditional business process lifecycle (Weske, 2010) which are influenced by the cultural and structural aspects in an organization (vom Brocke & Rosemann, 2010). Existing BPMMs seem to address these capability areas in line with the expectations of the experts. Almost all sampled BPMMs cover the areas related to the traditional business process lifecycle (i.e. 56 sampled BPMMs cover modeling, all models cover deployment and optimization, and 67 cover management) (Weske, 2010), plus culture (i.e. 57 out of 69 BPMMs) (vom Brocke & Rosemann, 2010). In line with the literature, structural reconfigurations in the organization chart or the formal introduction of a competence center, called Centre of Excellence (vom Brocke & Rosemann, 2010), are relatively less frequently addressed by BPMMs, and are also considered as less relevant by the experts (i.e. indicated by the lowest weight for this criterion). In general, good management by a professional process owner or process manager is considered essential

for improving business (process) performance. Secondly, regarding the 'architecture details' criterion, most BPMMs only implicitly offer a prescriptive roadmap (Maier et al., 2009). Also the experts are of the opinion that BPMMs must offer sufficient guidance, while allowing organizations to make their own choices instead of being too restrictive. Thirdly, the experts rated generic BPMMs higher than domain-specific BPMMs, as the former can be applied to any process type. Also the BPMM sample contains slightly more generic models (i.e. 37 generic versus 32 domain-specific BPMMs), suggesting that generic models are easy to find.

In strong contrast is the number of business processes to be assessed and improved. Table 7 shows that most BPMMs cope with multiple processes in a certain business domain (i.e. referring to the option 'more', such as in supply chains), followed by models that cope with all processes in an organization. Although the experts agree that a combination of particular and all processes would be beneficial to organizations, only two sampled BPMMs cover this combination. Finally, regarding the architecture types, Table 7 indicates that most BPMMs offer a staged architecture with maturity levels, possibly in combination with a continuous architecture containing capability levels, whereas the experts rather recommend combining both maturity levels and capability levels in order to obtain a refined overview. In particular, a maturity level expresses the overall state of a business process or a set of business processes, while capability levels can add information on individual capability areas needed for realizing specific organizational objectives (de Bruin & Rosemann, 2007; Tapia et al., 2008). For instance, separate capability levels for capability areas related to process deployment and process optimization may add information on a specific customer service objective or the capability area of process modeling can help realize an objective for obtaining grants or quality labels by means of a capability level that measures the efforts.

6.2.3. Evaluation of BPMMs against the contextual criteria

Table 8 focuses on the selection criteria pertaining to the contextual factors of a BPMM.

Table 8. Selection criteria pertaining to the contextual factors of a BPMM.

Contextual criteria	Options	Weight (%)	Number of BPMMs

			(N=69)
Purpose	Awareness	**7.18**	**46**
	Also benchmarking	**7.41**	20
	Also certification	2.33	3
	(Missing value)		*(0)*
Validation	Application	2.04	**25**
	Outcomes	**6.57**	19
	(Missing value)		*(25)*
Costs	Free	**3.42**	**31**
	Charged	1.47	13
	(Missing value)		*(25)*

For two of the three contextual criteria, the expert preferences correspond to what most sampled BPMMs offer (Table 8). Regarding the 'purpose' criterion, the experts agree that BPMMs aim at improving the internal way of working, but the experts also advise BPMM designers to enable benchmarking (i.e. by comparing maturity levels against competitors, across departments, industries or regions). Similarly, 46 out of 69 sampled BPMMs can be used for just raising awareness, while 20 BPMMs also allow benchmarking. Certification for external recognition is perceived as rather an unnecessary effort. Similarly, only few existing BPMMs issue a certificate. Furthermore, regarding the 'costs' criterion, many BPMMs can be accessed and used free of charges, which is also appreciated by the experts. On the other hand, the experts highly value the validation of outcomes in the 'validation' criterion, whereas this is only guaranteed by a minority of BPMMs (i.e. 19 out of 69 BPMMs or 27.5%). Strikingly, 25 sampled BPMMs do not mention any validation efforts in their documentation.

7. Discussion

7.1. Discussion about the research questions

The research in this paper has investigated selection criteria or BPMM desiderata, without focusing on particular BPMMs. It contributes to the literature by distilling a ranked and weighted set of BPMM selection criteria. The use of strict research methods allows to posit that the 14 developed criteria can be considered as being the most important ones for BPMM selection (RQ1). They constitute an objective (or rather inter-subjective) set of criteria that intends to identify the main differences among BPMMs for prospective users. In other words, our research efforts ensure that we can consider the proposed set as valid and if only a few selection criteria are covered to a satisfying extent by specific BPMMs, this could be seen as

an indication that these maturity models are flawed. For instance, a well-designed BPMM incorporates a number of assessment items, techniques for data collection, activities for improving capability areas, etc.

As a result of examining to what extent the selection criteria are applicable to current BPMMs (RQ2), taking into account the ranks and weights, the present study provided empirical evidence of the differences (and so the varying quality) among BPMMs (section 7.2). As such, the paper demonstrated the applicability of this ranked set to assess the quality of BPMMs by satisfying the initial requirements (Table 9), allowing reflections on the selection criteria themselves (section 7.3).

Table 9. Evidence for the applicability of our set of selection criteria for BPMMs.

	Requirement satisfaction tests	Evidence
Utility	• **Criteria utility.** Based on a textual description of the selection criteria, other BPMMs than those involved in our sample can be evaluated by people not involved in our research. The Cohen's Kappa represents a significant level of agreement or interrater reliability (0.4=<kappa=<1; p<0.05). • **Ranking utility.** Based on descriptive statistics, comparative tables can be generated to visualize the weights of individual selection criteria along with the effective support in current BPMMs. • **Clarity.** In the comparative tables, the main differences between BPMMs can be highlighted with respect to the ranked and weighted selection criteria.	• **Criteria utility.** kappa=0.571 (p=0.002). • **Ranking utility.** Table 6, Table 7, Table 8 • **Clarity.** Highlights in Table 6, Table 7, and Table 8 (i.e. bold and underlined for the highest weights and number of BPMMs; italics for the number of missing values)
Effectiveness	Prospective users are satisfied with (1) the selection criteria in the developed set, (2) their descriptions of trade-offs, and (3) the sequence in which they appear based on a general ranking with weights (i.e. 50% for scores 5-6-7 on a 7-point Likert scale, and no opposite extreme score of 1).	Satisfaction of 5 out of 7 prospective users (or 71.4%); no negative scores
Efficiency	Based on the comparative tables, relevant differences among BPMMs can be detected by following a standard way of working.	A difference exists between the general weights and the support in BPMMs for: • 4 out of 6 assessment method criteria • 2 out of 5 improvement method criteria • 1 out of 3 contextual criteria

7.2. Discussion about BPMMs

In the evaluation study, we noticed discrepancies between the experts' opinions and what BPMMs actually offer. Therefore, we take a closer look at some major issues, starting with the assessment method criteria. In particular, the use of qualitative versus quantitative

scales, and of objective versus subjective data collection techniques is a first point of discussion. The fact that only one BPMM in our sample merely uses quantitative scales and only four BPMMs merely use objective techniques for data collection emphasizes the importance of an organizational setting. For instance, some capability areas (like management or culture) are hard to measure with numbers and document reviews alone. Many BPMMs still have opportunities to enhance the accuracy of their assessments by applying data triangulation and combining the advantages of different rating scales and data collection techniques.

Secondly, in terms of the improvement method, a strong difference exists in terms of the number of business processes to be assessed and improved. This discrepancy indicates a substantial gap in existing BPMMs, which can be explained by the historical evolution of BPMMs. As from the 1980s, the CMM(I) tradition has inspired many maturity models to cover particular processes (Ahern et al., 2004). The focus on the BPM mastery of an organization is only introduced in the 2000s (McCormack & Johnson, 2001, de Bruin & Rosemann, 2007; Hammer, 2007). A next logical step in the BPMM evolution might be to simultaneaously examine particular and all business processes in an organization.

Thirdly, the preferences for contextual criteria can be explained by existing theories. For instance, the results showed that BPMMs would generally profit from introducing more benchmarking possibilities. Also in the literature, benchmarking is associated with best practices that can lead to organizational performance and learning, as a substitute for market and stakeholder forces (Van Helden & Tillema, 2005). If led by a third-party, the assessment may acquire more external recognition (i.e. credibility), but at higher cost (Mettler, 2009). Following (Grabowski & Lee, 1993), organizations must decide on how much money they are willing to spend in relation to their strategy and budget. Finally, another great challenge for most BPMMs was uncovered for the ‘validation’ criterion in section 6.2.3. In line with (Soh & Markus, 1995), organizations must recognize that business value can only be created with a time delay between design (or selection), appropriate use and outcomes. In other words,

validating the outcomes of a BPMM takes time, but maturity models can at least provide some information on their (ongoing) validation efforts in their documentation.

The discussion also gives rise to three concrete areas for improving existing BPMMs. These areas represent calls to action for both BPMM designers (i.e. who should include the identified selection criteria in the documentation of their models) and users (i.e. who should consider the selection criteria).

First, Figure 5 illustrates that the improvement method criteria are addressed by all sampled BPMMs, whereas BPMMs generally do not discuss the assessment method criteria and the contextual criteria. This particularly applies to the assessment duration (with 43 missing values out of 69 BPMMs), validation and costs (both with 25 missing values). The improvement method criteria represent the direction of a BPMM, independent of a practical assessment, and thus received the highest weights. Another explanation for the missing values is that the design studies presented in section 2 each address less than half of our selection criteria (Mettler, 2009; Röglinger et al., 2012), which may impact on the decisions made by BPMM designers so far. This leads to a first area for improvement in BPMMs.

Suggested improvement 1. Designers (e.g. scholars or consultancy firms) should provide information on all identified selection criteria to facilitate the user's choice.

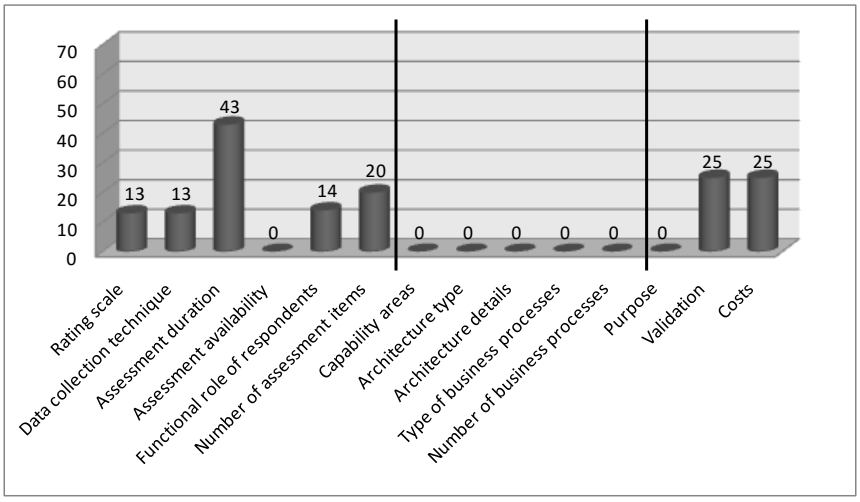

Figure 5. The number of BPMMs with missing values per selection criterion (N=69).

This first suggested improvement is not a trivial matter in practice, since many of the existing BPMMs can be enhanced by addressing the criteria discussed in our study. Figure 6 gives further evidence for the varying quality of BPMMs by looking at the rankings. Based on the relative weights obtained through AHP in the international Delphi study, a selection score (from 0 to 100) was calculated per sampled BPMM. Per criterion, a BPMM scores the weight that corresponds to the option to which it applies.

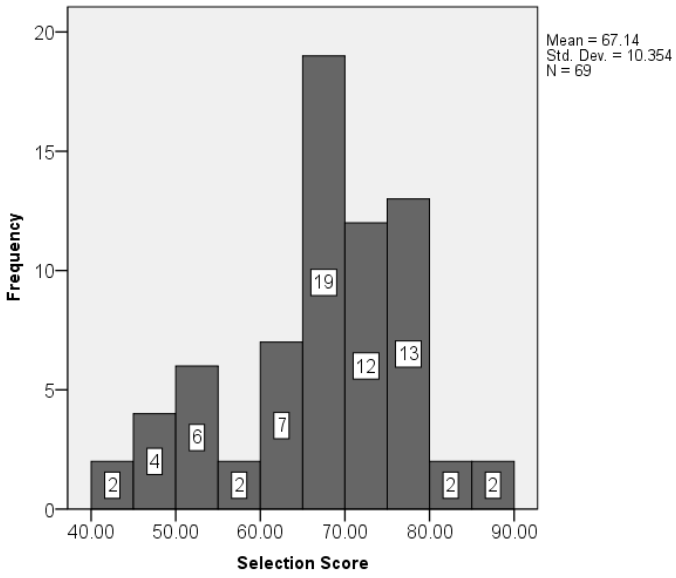

Figure 6. The frequency distribution of selection scores across BPMMs (N=69).

Figure 6 shows that the selection scores in the BPMM sample range from 40 to 90, with a mean of 67.14 and a standard deviation of 10.35. Fifty-one BPMMs have a score between 60 and 80, whereas four BPMMs have a higher score and 14 BPMMs have a lower score. Hence, no BPMM in the sample covers all the ideal options, as stipulated by the experts. This is not a shortcoming per se, as the preferences of specific organizations may still differ from the assigned weights. Although we aim for general conclusions (instead of focusing on particular BPMMs), we note that the most mentioned BPMMs in the literature are not mediocre. For instance, the models of CMMI (SEI, 2009), OMG (2008) and Harrington (2006) have a score between 65 and 70, whereas the models of de Bruin and Rosemann (2007) and Hammer (2007) score around 75. McCormack and Johnson's model (2001) is in the forefront with almost 82 points. A direct comparison, however, remains difficult as a fit with an organization's particular needs is paramount. For instance, the number of business

processes addressed in these BPMMs differ. Instead, we generalize that BPMMs with the lowest selection scores (i.e. with many missing values and/or a combination of less preferred options) may require an improvement.

Suggested improvement 2. Designers should consider the different options per selection criterion, and possibly make their design more flexible by providing alternative options to offer a better fit for purpose.

The third area for improvement refers to the wide range of selection scores.

Suggested improvement 3. Organizations, as potential BPMM users, should develop a critical attitude towards the fit for purpose of BPMMs, instead of taking quality for granted.

We do not assert that all BPMMs are flawed and none should ever be selected, given that the practical value of using a maturity model is positively perceived (Harmon, 2013; Harmon & Wolf, 2014) and empirically demonstrated (McCormack & Johnson, 2001; Skrinjar, Bosilj-Vuksic & Stemberger, 2008). Instead, the evaluation study indicates ample opportunities for designers to improve their BPMMs, and indicates that organizations should not choose just any BPMM. Our findings motivate why we do not advocate a single BPMM. The perfect BPMM (i.e. which satisfies all needs of all organizations) seems non-existent, and is perhaps unrealistic to design. Hence, organizations must make concessions on selection criteria that are less relevant to them. As this study emphasized the importance of capability areas, concessions should never be made on this criterion.

7.3. Discussion about the selection criteria for BPMMs

We now reflect on a generalization of the findings regarding the distilled set of selection criteria themselves, the knowledge contributions, and a possible refinement of the options per criterion in future research.

First, our research offers standardized steps to compare other or new BPMMs:

- Search for BPMM documentation by combining the keywords ‘process’ and ‘maturity’, and verify whether it: (1) contains maturity levels and/or capability levels, and (2) primarily focuses on business processes instead of other organizational assets.
- Perform a content analysis of the identified BPMM documentation to analyze how the selection criteria of Table 5 and Appendix 1 are applied.
- Compare the results of the content analysis with the general weights of Figure 4.
- Evaluate the BPMM documentation by counting the missing values (Figure 5) and the number of times the model does not apply the selection criteria options as preferred by the experts (i.e. the options with the highest weights) (Figure 6).

Furthermore, we demonstrated that the criterion ‘type of business processes’ allows specifying domain-specific BPMMs for supply chains and collaboration processes. Although not investigated, other domain-specific BPMMs may be similarly covered, like those specifically for software processes. Further research could extend our approach to other maturity models than BPMMs, like for business-IT alignment, e-government, project management, etc. Many selection criteria seem generic for any maturity model, particularly those related to the assessment method and the contextual criteria. Moreover, our methodology is reusable for other maturity model types. Opportunities also exist for fellow scholars to examine BPMMs during their use, and the impact these designs may have on the ability of individuals to understand the models.

Next, we discuss the degree to which the developed set of selection criteria goes beyond the current body of knowledge. Although diverse elements in the design of maturity models (section 2, Table 1) were considered in the Delphi study as possible BPMM desiderata, it turned out that some are less decisive for BPMM selection. This proves that selection criteria or BPMM desiderata differ from currently known design criteria, and can supplement the latter. At first sight, when contrasting our findings with the comparative work of Mettler (2009) and Röglinger et al. (2012), mentioned in the literature review, it seems that each study addresses less than half of the criteria we identified. Both design studies neither discuss

trade-offs nor alternatives, e.g. different ratings scales, data collection techniques, or ways to calculate and represent lifecycle levels. As many BPMM desiderata are not addressed by the presented design studies, they contribute to the Business Process Management discipline as new knowledge of BPMMs. Nonetheless, a direct comparison remains difficult as the design criteria of Mettler (2009) and Röglinger et al. (2012) are at different levels of analysis. For instance, many design criteria of Mettler (2009) seem to refer to decisions that organizations must make, independent of a BPMM, e.g. whether improvement projects are launched, whether they are led by consultants, and the link between process improvements and other organizational targets. Furthermore, Röglinger *et al.* (2012) present a criterion that they have never seen in any BPMM. It concerns a decision calculus (or cost-benefit analysis) to verify which improvements best fit a particular organization. This criterion might supplement our 'costs' criterion in future research. Surprisingly, the presence of capability areas (which are core to all maturity models) is only mentioned by Röglinger et al. (2012), albeit without identifying the capability areas, and not by Mettler (2009). As our study has assigned top priority to capability areas, we encourage further research on their theoretical understanding. Particularly, business process capability areas have been underpinned by only a few scholars (de Bruin & Rosemann, 2007; Blinded3). More research is needed to verify how they can be improved in a more dynamic way instead of a static, one-size-fits-all roadmap presented in current BPMMs (Pöppelbuss et al., 2011), and to refine capability areas with a contingency approach depending on organization-specific characteristics (Plattfaut, Niehaves, Pöppelbuss & Becker, 2011; vom Brocke et al., 2015). This final criticism is not only valid for BPMMs, but generally for most maturity models (Donnellan & Helfert, 2010).

This brings us to future research opportunities. For instance, a similar evaluation of BPMMs can be performed within five or ten years to examine any progress or evolution made in the BPMM documentation. We also strongly encourage further research on how capability areas can become more dynamic and contingent upon organization-specific factors (like organization size, market competitiveness, etc.). For instance, a new Delphi study or case

studies could generate ideas that help solve such problems in order to further theorize about BPMMs and the Business Process Management discipline. Given the observed shortcomings, current BPMMs may help raise awareness and serve as benchmarking tools, without being the ultimate solution to all implementation problems. Furthermore, another Delphi study could collect data to elaborate on the trade-offs of other selection criteria. Particularly, the 'costs' criterion could be detailed with price levels to develop a cost-benefit analysis or decision calculus. Meanwhile, organizations could benefit from customization by changing and combining existing BPMMs to create a BPMM that better fits their organizational needs (possibly supported by a more advanced decision support tool than presented in (Blinded1)).

8. Conclusion

In this study, we have contributed to research on Business Process Management and maturity models by performing an in-depth analysis of the constituents of BPMMs with the purpose of identifying selection criteria. This paper has discussed a range of weighted criteria that impact on the selection of a BPMM (based on an international Delphi study with strict consensus conditions and coder triangulation) in order to provide evidence of the strengths and weaknesses of BPMMs (based on a content analysis of a large BPMM sample for generic processes, supply chains and collaboration processes). The findings shed light on issues in the documentation of BPMMs, relevant to managers and executives in their investment decisions when choosing a particular BPMM. Moreover, the evaluation of the applicability of the selection criteria provided an empirical analysis of deficiencies in BPMMs. To the best of our knowledge, it is the first time that a profound evaluation or quality check of BPMMs is carried out from a user perspective. The results also confirm that obtaining the highest maturity levels is not an end goal of BPMMs, but capability improvements and performance improvements are. In sum, our study has uncovered a rich understanding of normative user behavior associated with BPMM selection, supplementing existing design studies. It has also provided evidence for the varying quality of many BPMMs that go around

these days, resulting in calls to action for both designers and organizations. Avenues for future research include a generalization towards other types of maturity models and the actual use of BPMMs (to supplement selection, which comes first). The BPMM selection criteria can also be refined, for instance by investigating how the process capability areas can become more dynamic and contingent (i.e. organization-dependent) and by investigating how BPMMs can facilitate a cost-benefit analysis for organizations to make accurate investments with respect to improvement initiatives. Other Delphi studies might be conducted to further explore and overcome these limitations.

ACKNOWLEDGMENTS

To be inserted after the blinded review process.

REFERENCES

Ahern, D. M., Clouse, A., & Turner, R. (2004). *CMMI Distilled.* Boston: Pearson Education.

Becker, J., Knackstedt, R., & Pöppelbuss, J. (2009). Developing maturity models for IT management. *Business & Information Systems Engineering, 1*(3), 213-222.

Blinded1: authors of this paper are not mentioned during the double-blind review.

Blinded2: authors of this paper are not mentioned during the double-blind review.

Blinded3: authors of this paper are not mentioned during the double-blind review.

Bostrom, R. P., Gupta, S., & Thomas, D. (2009). A meta-theory for understanding information systems within sociotechnical systems. *Journal of Management Information Systems, 26*(1), 17-47.

Crosby PB (1979) Quality is free. McGraw-Hill, New York

Curtis, B., & Alden, J. (February 2007). The business process maturity model: what, why, and how. *BPTrends*.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458-467.

de Bruin, T., Freeze, R., Kulkarni, U., & Rosemann, M. (2005). Understanding the main phases of developing a maturity assessment model. *ACIS Proceedings* (pp. 1-10) Sydney: AISEL.

de Bruin, T., & Rosemann, M. (2007). Using the Delphi technique to identify BPM capability areas. *ACIS Proceedings* (pp. 642-653). Toowoomba: AISEL.

Donnellan, B., & Helfert, M. (2010). The IT-CMF: A practical application of design science. *DESRIST Proceedings* (pp. 550-553). St.Gallen: Springer.

El Emam, K., & Birk, A. (2000). Validating the ISO/IEC 15504 measure of software requirements analysis process capability. *IEEE of Transaction Software Engineering, 26*(6), 541-566.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of Grounded Theory: strategies for qualitative research.* New Jersey: Transaction Publishers.

Grabowski, M., & Lee, S. (1993). Linking information systems application portfolios and organizational strategy. In R. D. Banker, R. J. Kauffman & M. A. Mahmood (Eds.), *Strategic Information Technology Management* (pp. 33-54). Pennsylvania: IGI Global.

Hammer, M. (2007). The process audit. *Harvard Business Review, 4*, 111-123.

Harmon, P. (2013). Case studies in how organizations become more mature. *BPTrends, 11*(11), Email Advisor.

Harmon, P., & Wolf, C. (2014). The state of business process management 2014. *BPTrends.*

Harrington, H. J. (2006). *Process Management Excellence.* California: Paton Press.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32*(4), 1008-1015.

Hevner, A. R., March, S.T., Park, J., & Ram, S. (2004). Design science in information systems research, *MIS Quarterly, 28*(1), 75-105.

Maier, A. M., Moultrie, J., & Clarkson, J. P. (2009). Developing maturity grids for assessing organizational capabilities: practitioner guidance. *ICMC Proceedings, Academy of Management* (pp. 1-29). Vienna: Management Consulting Division.

Maier, A. M., Moultrie J., & Clarkson, J. P. (2012). A review of maturity grid based approaches to assessing organizational capabilities. *IEEE Transactions on Engineering Management, 59*(1), 138-159.

March, S. T., & Smith, G.F. (1995). Design and natural science research on information technology, *Decision Support Systems, 15*(4), 251-266.

McCormack, K., & Johnson, W. C.(2001). *Business process orientation.* Florida: St. Lucie Press.

Mettler, T. (2009). *A design science research perspective on maturity models in information systems* (Report BE IWI/HNE/03). St. Gallen: Institute of Information Management.

Mettler, T., & Rohner, P. (2009). Situational maturity models as instrumental artifacts for organizational design. *DESRIST Proceedings* (pp. 1-9). Malvern: ACM.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design constructions and applications. *Information & Management, 42*, 15-29.

OMG (2008). *Business Process Maturity Model (BPMM). Version 1.0.* Retrieved December 2, 2009 from http://www.omg.org/spec/BPMM/1.0/PDF

Peffers, K., Tuunanen, T., Rothenberger, M.A., & Chatterjee, S. (2007). A design science research methodology for information systems research, *Journal of Management Information Systems, 24*(3), 45-77.

Plattfaut, R., Niehaves, B., Pöppelbuss, J., & Becker, J. (2011). Development of BPM capabilities. *ECIS Proceedings* (pp. 1-12). Helsinki: AISEL.

Pöppelbuss, J., Plattfaut, R., Ortbach, K., Malsbender, A., Voigt, M., Niehaves, B. et al. (2011). Service innovation capability: proposing a new framework. *FCCSIS Proceedings* (pp. 545-551). Szczecin (Poland): IEEE Explore.

Dijkman, R., Lammers, S. V., & de Jong, A. (2015). Properties that influence business process management maturity and its effect on organizational performance, *Information Systems Frontiers, 1-18*. DOI 10.1007/s10796-015-9554-5

Röglinger, M., Pöppelbuss, J., & Becker, J. (2012). Maturity models in business process management. *Business Process Management Journal, 18*(2), pp. 7-7.

Saaty, T. L. (1990). An exposition of the AHP in reply to the paper "remarks on the analytical hierarchy process". *Management Science, 36*(3), 259-268.

SEI (2009). *CMMI for services. Version 1.2.* Retrieved February 11, 2010 from http://www.sei.cmu.edu/reports/09tr001.pdf

Sheard, S. A. (2001). Evolution of the frameworks quagmire. *IEEE Computer, 34*(7), 96-98.

Skrinjar, R., Bosilj-Vuksic, V., & Stemberger, M. I. (2008). The impact of business process orientation on financial and non-financial performance. *Business Process Management Journal, 14*(5), 738-754.

Soh, C., & Markus, L. M. (1995). How IT creates business value: a process theory synthesis. *ICIS Proceedings* (pp. 1-12). Amsterdam: AISEL.

Tapia, R. S., Daneva, M., van Eck, P., & Wieringa, R. (2008). Towards a business-IT alignment maturity model for collaborative networked organizations. *EDOC Workshop Proceedings* (pp. 70-81). Munich: IEEE.

Van Helden, J. G., & Tillema, S. (2005). In search of a benchmarking theory for the public sector. *Financial Accountability and Management, 21*(3), 337-361.

van Steenbergen, M., Bos, R., Brinkkemper, S., van de Weerd, I., & Bekkers, W. (2010). The design of focus area maturity models. *DESRIST Proceedings* (pp. 317-332). St.Gallen: Springer.

vom Brocke, J., Rosemann, M. (2010). Foreword. In J. vom Brocke & M. Rosemann (Eds.), *Handbook on BPM 2* (pp. vii-ix). Berlin Heidelberg: Springer.

vom Brocke, J., Zelt, S., & Schmiedel, T. (2015). On the role of context in business process management. *International Journal of Information Management* (in press). doi:10.1016/j.ijinfomgt.2015.10.002

Wendler, R. (2012). The maturity of maturity model research: A systematic mapping study. *Information and Software Technology, 54*(12), 1317-1339.

Weske, M. (2010). *Business Process Management.* Berlin Heidelberg: Springer.

Zhao, S. (1991). Metatheory, metamethod, meta-data-analysis: What, why, and how? *Sociological Perspectives, 34*(3), 377–390.

APPENDIX 1. BPMM selection criteria

A. Assessment Method Criteria

- **Rating scale: the type of data that is collected during an assessment.**
  - Quantitative data (discrete, interval or ratio scales) can be statistically analyzed and compared, independent of the assessors' interpretation. Qualitative data (nominal or ordinal scales) provide more in-depth descriptions, but depend more on the assessors' skills. Also a combination is possible, depending on the available data and skills.
- **Data collection technique: the way information is collected during an assessment.**
  - Objective techniques involve document reviews, and give an idea of how organizations work, without interrupting individuals or activities. They minimize biased results of (particularly internal) assessors and respondents. Subjective techniques gather information about how organizations actually work, e.g. by questionnaires, interviews or observations. As it concerns personal beliefs, some precautions can be taken, e.g. a third party lead assessor, multiple assessors and respondents, data collection training, or a combination with objective techniques.
- **Assessment duration: the maximal duration of a particular assessment.**

- Some BPMMs only take one day (e.g. a quick scan within 15 minutes), whereas other BPMMs present a more profound analysis of one week or longer. As time is money, the user must consider how much time he wants to spend on the assessment alone.

- **Assessment availability: whether the assessment questions and corresponding level calculation are publicly available (instead of only known to the assessors).**
  - BPMMs do not always provide full details. This particularly counts for non-academic models, e.g. in consultancy. The user must decide whether this limited availability is an issue for the organization. For instance, fully known BPMMs (i.e. either free or charged) can be used for educating process team members or for earning credibility.

- **Functional role of respondents: the explicit recognition to include people from outside the assessed organizations as respondents.**
  - If only internal respondents (managers and/or staff) are questioned, the user assumes they fully know the stakeholders' needs. By involving stakeholders, an organization recognizes the need for an outside-in perspective by explicitly listening to them.

- **Number of assessment items: the maximal number of questions to be answered during an assessment.**
  - More questions provide more insight to develop a road map, but may be less feasible and/or take longer. Less than 20 questions are rather used as a teaser or a quick scan.

B. Improvement Method Criteria

- **Capability areas: the capabilities to be assessed and improved.**
  - BPMMs differ in the capabilities they address, varying from basic capabilities related to the traditional business process lifecycle (i.e. modeling, deployment, optimization, and management) to the addition of organizational capabilities (i.e. to create a process-oriented culture and structure). In theory, all presented capability areas are required for fully mature business processes. However, in practice, an organization can opt for a subset, e.g. depending on the degree of top management support, IT background of the user, prior BPM experience, organization size, etc. For instance, organizations with local, bottom-up initiatives or with limited BPM experience might wish to start with the basic capability areas, limited to the traditional business process lifecycle. Additionally, the culture capability area requires a minimum level of management support to promote business processes and granting (financial) rewards to process performance. Finally, structural configurations inherently require top management support. The latter is recommended if you already have some BPM experience or if your ambition is to standardize processes across large departments or divisions. The user must select a set of capability areas that best fits its organizational needs.

- **Architecture type: the possibility to define a road map per capability, a road map for overall maturity, or both.**
  - It concerns linking (maturity of capability) levels to capability areas in a step-by-step plan, which explains how to reach each consecutive level. A continuous architecture provides capability levels per capability area, i.e. one road map per area. It allows organizations to assess and improve each capability area separately, and thus to improve areas at a different pace or to limit their scope to only those capability areas they are interested in. As not all capability areas are necessarily taken into account, there is a risk for suboptimal optimizations (in terms of overall maturity). On the other hand, a staged architecture provides maturity levels linked to all capability areas together, i.e. one road map for overall maturity. The emphasis is on simultaneous advancements, instead of individual capability advancements.

- **Architecture details: the degree of guidance that a BPMM gives on your journey towards higher maturity.**

- It concerns the extent to which the road map (step-by-step plan) explains which criteria (goals and best practices) must be satisfied before reaching each particular level: (1) descriptive, (2) implicit prescriptive or (3) explicit prescriptive. A descriptive road map is limited to high-level descriptions. As it gives less support, it is suited for organizations wishing to become acquainted with BPMMs, or for organizations which are highly experienced with process improvements. An implicit prescriptive road map has criteria interwoven in the assessment questions, i.e. with an ordinal scale or a matrix, that explain all capability areas per level. Assessors can derive the criteria from the assessment questions. Finally, an explicit prescriptive road map gives most guidance by separately listing criteria from the assessment questions.

- **Type of business processes: whether the BPMM is generic (i.e. for business processes in general) or domain-specific (e.g. for business processes in supply chains or collaboration situations).**
  - The terminology used in generic BPMMs, e.g. in the assessment questions, is likely more holistic. Benchmarking becomes possible across business domains. Domain-specific BPMMs use terminology adapted to their domain, which might be less abstract and thus better understandable. However, benchmarking remains limited to organizations within the same domain. This choice requires strategic considerations.

- **Number of business processes: the number of business processes to be assessed and improved: (1) one, (2) more or (3) all.**
  - For BPMMs focusing on a single business process, the process boundaries must be defined by the user, e.g. whether a business process is assessed and improved as a sub process or as a separate process. BPMMs can also focus on more than one, but not all business processes within the assessed organizations. Assessment questions then deal with a particular business domain or value chain and all its (sub) processes. Furthermore, BPMMs can cope with all business processes in the assessed organizations. As such, assessment questions take a management perspective by focusing on how organizations deal with business processes in general, without focusing on particular processes. By improving this BPM mastery, it is likely that particular processes are indirectly improved too.

C. Contextual Criteria

- **Purpose: the purpose for which a BPMM is intended to be used.**
  - The basic purpose of any BPMM is assessing and identifying process improvements, i.e. raising awareness. The key is recognizing deficiencies, creating willingness to act and to follow-through on the findings. Besides raising awareness, BPMMs can allow benchmarking with other organizations (for comparing with competitors and sharing best practices) or certification (for external recognition of assessment results).

- **Validation: whether or not empirical evidence is given that the BPMM helps to enhance the efficiency and effectiveness of business processes.**
  - Most BPMMs do not provide any proof of validity (or success). If they do, evidence is frequently limited to enumerating other organizations applying the model. Only few BPMMs give evidence for the performance outcomes. The user must decide whether some proof of validity is required, depending on the planned investments. However, we discourage the use of non-validated BPMMs. They can result in frustrations, time and money losses, i.e. if they appear to be flawed or unusable after you start using it.

- **Costs: the direct costs to access and use a BPMM.**
  - Not all BPMMs are free of charges. Particularly non-academic models may ask a one-off access fee or a required training to be followed. Recurring costs rather serve to pay a third party lead assessor, certification or benchmarking. The user must decide which budget can be spent, and adapt his expectations accordingly: you often get what you pay for. Academic models can be free if they use your data for research enhancements.