e-Informatica Software Engineering Journal, Volume 9, Issue 1, 2015, pages: 57–77, DOI 10.5277/e-Inf150104

An Approach to Assessing the Quality of Business Process Models Expressed in BPMN

Małgorzata Sadowska*
*Faculty of Computer Science and Management, Institute of Informatics, Wrocław University of Technology
m.sadowska@pwr.edu.pl

Abstract

Introduction: The quality of business process models is important in the area of model-based software development. To the best knowledge of the author there is no working practical model for quality assessment of BPMN 2.0 Process Diagrams, which measures the actual models and automatically interprets the measured values.
Objectives: Propose a metamodel for assessing the quality of BPMN 2.0 process models and a working solution – a model for quality assessment of process models in BPMN (called MAQ) and a tool that implements MAQ.
Methods: The metamodel was built upon the information presented in ISO/IEC 25010 (2011) standard. The methodology of MAQ was driven by its essential elements. Quality characteristics were selected through a systematic literature review. Quality metrics were identified through a literature review restricted by questions that every relevant literature had to affirmatively answer. Quality metrics were implemented in the tool and quality criteria were proposed based on the interpretation of the results of measuring a repository of BPMN models. Finally, quality functions were proposed and the complete MAQ was implemented in the tool.
Conclusions: MAQ was preliminary evaluated for usefulness through a survey-based experiment. The results showed that the model works in most cases and is needed in general.

Keywords: BPMN quality model

1. Introduction

Working with models has become a common practice in model-based software development. Models play an important role in the entire development process. In order to model business processes, various notations and languages are used, such as BPMN, UML Activity Diagram, UML EDOC Business Processes, IDEF, ebXML BPSS, Activity-Decision Flow (ADF) Diagram, RosettaNet, LOVeM and Event Process Chains (EPCs). In this paper BPMN was chosen because it is a standard notation used to model business processes [1] and it was created in a way that it is readily understandable by all business users, while still being able to represent complex process semantics [2].

Nowadays the need for achieving high quality BPMN models seems to be undeniable [1, 3, 4]. To start with, quality has an impact at ease of early detection and therefore correction of BPMN models. Early discovery of defects in software artifacts is cheaper than repairing consequences of modelling errors in later design phases [3,5]. Also, poor quality of process models can result in poor information systems [6]. Next, models of good quality are claimed to have a positive influence in reducing software maintenance costs [7]. Finally, all the mentioned aspects may lead to economic benefits through satisfaction of the user requirements for BPMN models and the resulting software.

The desire to ensure high quality in the actual BPMN models is a background underlying the idea of the model for assessing the quality of BPMN models (called MAQ). MAQ is designed for quality assessment of BPMN 2.0 Process Diagrams, but MAQ was not intended to calculate other types of BPMN models (Collaboration Diagrams or Choreography Diagrams) what can be considered as a limitation of the model. MAQ considers every graphical construct for process models defined in the standard, thus it is able to calculate actual models. By "an actual model" the author understands a Process Diagram which is not limited to a truncated subset of BPMN graphical elements, represents complex process semantics and is able to graphically represent the actual business process. Supporting the quality of the so understood actual models in the opinion of the author is essential and is an aim of this paper.

A solution that seeks to help modellers in verifying the quality of actual models in an effective automated way cannot be abstract and should be easy to be directly used on actual BPMN models. Therefore, the focus of this paper is on developing a model, and more importantly, a method for assessing the quality of business process models in BPMN.

Measuring business process models is a relatively new discipline [8] even though the first version of BPMN 1.0 was released already in 2004 and the final adopted specification of BPMN 1.0 was finalized in 2006. Currently, BPMN has already been evaluated both empirically and analytically [9]. The literature describes many metrics that can be potentially used for assessing the quality of business process models. Please notice that MAQ is designed for quality assessment of the actual models which use a full range of the BPMN Process Diagram graphical constructs. The need for supporting all constructs triggered the need to apply a list of assumptions or changes in the selected metrics. This is caused by the fact that many of the original versions of metrics were designed to calculate only models with a truncated subset of elements, not actual models, which represent a complex process semantics.

The rest of the paper is organized as follows: Section 2 presents related works, Sections 3-5 define a quality metamodel, the developed MAQ and the implemented tool, Section 6 summarizes preliminary evaluation of MAQ, Sections 7 and 8 present threats to validity and conclusions.

2. Related Work

The model for assessing the quality of business process models in BPMN was only found to be directly related to the findings of two other papers: [10] and [11].

The contribution of [11] known as the 3QM-Framework, provides quality marks, metrics and measurement procedures which mainly focus on evaluating quality of handwritten BPMN models. The overall quality in 3QM-Framework is based on aggregation of metrics and measurement procedures and its result may vary depending on the project context. Therefore, user groups have to derive weighting of measurement. This paper differs by proposing a model for an instant and automatic assessment of quality, which is aimed to be helpful also for non-expert users. Quality marks from 3QM-Framework are referred to Section 4.1 among other findings from systematic literature review.

Makni et al. [10] implemented a tool which can provide the results of measuring some of BPMN metrics chosen from the literature by its authors. The tool is aimed to help designers to choose a subset of metrics corresponding to design perspectives. Interpretation of the results of measurements is left to users.

A systematic literature review in the area of model quality was conducted by Mohagheghi, Dehlen and Neple [5]. The focus of this paper is set only on business process models created with the use of BPMN. The classification of model quality goals developed in [5] is referred among other literature references to Section 4.1.

Sánchez-González et al. [4] presented a systematic review of measurements for business processes. The metrics from the review were taken into consideration while developing the MAQ. Additional help in choosing relevant met-

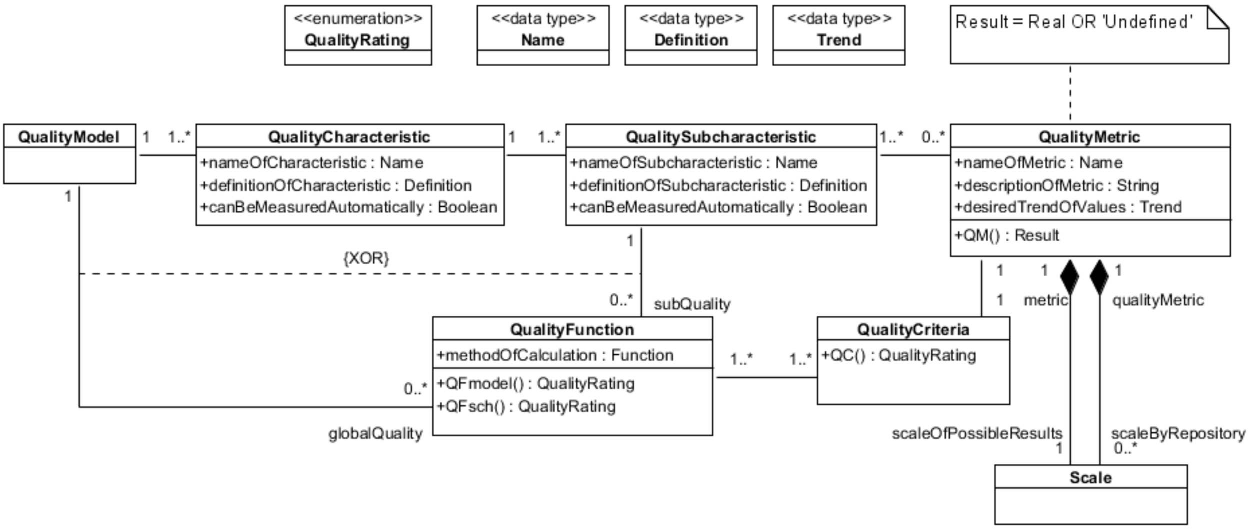


Figure 1. Metamodel for assessing the quality of business process models in BPMN.

rics came from [12]. Nonetheless, the final list of quality metrics used in the MAQ was extracted from the literature based on the proposed selection criteria listed in Section 4.2. The metrics in some cases were additionally adjusted by the author to calculate actual models in BPMN.

MAQ and the metamodel for assessing the quality of business process models in BPMN was initially developed in the author's master thesis [13]. This article presents a reanalyzed approach to the information contained in the thesis and the improved version of MAQ, the tool that implements MAQ and the metamodel. There were many major changes and improvements in the metamodel, and some changes in MAQ and the tool. The most important change in MAQ was removing the indicators for the syntactic quality of BPMN models. All the definitions and descriptions presented in the article were rethought and reanalyzed from its initial proposition.

3. Metamodel for Assessing the Quality of Business Process Models in BPMN

This section introduces the proposed metamodel for assessing the quality of business process models in BPMN. The structure of the metamodel is presented in Figure 1. An example instantiation of the metamodel is MAQ (described in Section 4).

The metamodel is built upon the information presented in ISO/IEC 25010 [14] in conjunction with ISO/IEC 14598 [15] standard. Following [14], a quality model is a "defined set of characteristics, and of relationships between them, which provides a framework for specifying quality requirements and evaluation quality." The hierarchical decomposition is the main idea of the model which is aimed to decompose quality down to a level which can be measured and thus the quality can be evaluated.

By quality characteristics, the quality of actual BPMN model can be described and evaluated. Quality characteristics are further decomposed into related quality subcharacteristics. The role of the subcharacteristics is to specify the general characteristics more concretely. Quality characteristics and subcharacteristics in the metamodel are suggested to be named and defined in a natural language.

In order to talk about measurement, both terms "metric" and "measure" are often used interchangeably by researchers [16]. In this paper the term "quality metric" is adopted. Quality metrics are aimed for measuring quality subcharacteristics. One metric may be assigned to more than one subcharacteristic and as it is suggested in the standard more than one quality metric may be used to measure a quality subcharacteristic. A scale defines mathematically a theoretically possible results that a potential BPMN model may obtain for a specific quality metric, as a result from the calculation of the mathematical equation. The scale is also used to specify a scale of results obtained by a repository of actual models in BPMN – this scale is a subset of the scale of theoretically possible results. Each quality metric also owns a desired trend of values that are favorable for a metric. The trend can be described in a natural language, e.g. the lower obtained value of the quality metric by an actual BPMN model, the better quality of the model.

Quality rating defines rating levels for the measured values. In the metamodel, quality criteria are used to determine the rating levels associated with the results. The results are the obtained values on the scale of quality metric for a specific BPMN model. Finally, quality functions are used to assess the quality of quality subcharacteristics or the overall quality of the actual BPMN model. Quality functions base on the quality criteria and either quality subcharacteristic or a quality model, what is represented by a XOR constraint in the metamodel.

4. MAQ

The metamodel presented in Section 3 defines the structure of MAQ. The metamodel may be used to produce other models for assessing the quality of business processes models in BPMN, MAQ is only one of the possible instantiations of the metamodel. The following subsections are aimed to present how the essential elements of MAQ were obtained.

4.1. Selection of Quality Characteristics and Quality Subcharacteristics

The set of quality characteristics and quality subcharacteristics subsequently determined and gathered together constitute a hierarchical structure of MAQ. In order to identify characteristics, a systematic literature review (SLR) was conducted. Following [17], "a systematic literature review (...) is a means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest." The goal of this review was to provide definitions of characteristics based on analysis of previous literature in the area. The synthesis of the literature was conducted using a well-defined methodology and search strategy with the specified two research questions being addressed: RQ1: *"What quality characteristics of models exist?"* and RQ2: *"Which of the identified quality characteristics are suitable to the developing model?"*.

In order to make the process replicable, the search strategy consisted of several steps as outlined in [17]. At first, keywords were identified in order to minimize the effect of differences in terminologies. The following are the keywords that were formulated from the terms used in the domain and research questions; or their synonyms, alternate words and meaningful combinations: "BPMN," "business process models," "model-driven engineering," "conceptual modelling," "quality," "model quality," "quality characteristics," "quality goals," "quality of business process models." The keywords were used to build search queries in order to obtain relevant articles. Six queries were based on the Boolean AND to join keywords: 1) "model quality" AND "business process models" 2) "quality characteristics" AND "business process models" AND BPMN 3) "quality goals" AND "business process models" AND BPMN 4) "quality of business process models" 5) "conceptual modeling" AND "model quality" 6) "model-driven engineering" AND quality.

The search was conducted within the following electronic databases: ACM Digital Library, SpringerLink, ScienceDirect, Emerald, Academic Search Complete, Elsevier/ICM and ProQuest. Strategy for searching was conducted in two phases. In the first phase, the chosen publication channels were searched for. After eliminating duplicates and reading titles and abstracts in all of the found papers, the literature was chosen for further reading based on the studies selection criteria. Two inclusion criteria were applied: "paper describes quality characteristics of models" and "paper must contain the search keywords." In spite of that, two exclusion criteria were used: "paper describes quality characteristics of software products" and "paper does not relate to Software Engineering/Development." Finally, the full body of the filtered literature was read and the literature relevant for this systematic literature review was identified. After the first phase of searching, the second phase was initiated in order to obtain a more representative set of studies. In this phase, the reference list of all the selected literature was scanned in order to discover more papers. Lastly, if the literature was claimed to be relevant, it was found in the electronic databases.

The first phase of searching resulted in 10 papers and second search phase additionally included 7 papers. The final list of primary studies included in systematic literature review contains the following papers: [3, 5, 11, 18–30]. The description of the studies with identified quality characteristics can be found in [13].

To summarize, the primary studies describe 14 sets of quality characteristics of models. The obtained results concentrate mostly around characteristics of UML models (8 papers). The other resulting papers refer to characteristics for: conceptual models (3 papers), collaborative modelling including models (1 paper) and information models (2 papers). Only 3 papers were found which directly discuss quality characteristics of business process models, however, the studies are rather recent, from the years between 2010–2012.

In order to answer RQ1 and RQ2, definitions of characteristics from findings, explicitly relevant to BPMN models, were gathered together and compared against each other. The selected and systematized characteristics from the literature particularize the area of quality in order to be relevant for business process models in BPMN. Definitions of the quality characteristics and quality subcharacteristics of MAQ:

1. **"Correctness"** – in accordance with an analysis in regards to making correct statements about the domain AND following BPMN notation according to the specification, e.g. not violating rules and conventions (well formedness and syntactic correctness).
   a) **"Syntactic correctness"** – model in BPMN is syntactically correct if all terms are used in accordance with the syntax rules of the BPMN notation.
   b) **"Semantic correctness"** – model in BPMN is semantically correct if it corresponds to the domain and the reality of the analysed situation.
2. **"Integrity"** – description of all and only relevant elements of the domain, business process and purpose of modelling.
   a) **"Informational completeness"** – a correct scope of the BPMN model (does model in BPMN include all and only relevant features of the domain).
   b) **"Consistency"** – no contradictions in the model and the domain concepts are adequately represented in the model.
   c) **"Accordance with purpose"** – is when the BPMN model meets the original goals for why it was created.
3. **"Modifiability"** – ability of the BPMN model to be modified or changed AND supporting reusability and extensibility.
   a) **"Changeability"** – support for changes or improvements.
   b) **"Reusability and extensibility"** – support for the model to be used in the creation of new models or extended with new terms.
4. **"Complexity"** – related to the complexity of the BPMN model with the goal of simplicity and minimalism.
   a) **"Minimality and simplicity"** – if the BPMN model contains the minimum possible constructs.
5. **"Understandability"** – satisfaction of the users and their comprehensibility AND an aesthetics of BPMN model.
   a) **"User comprehensibility"** – being understandable by users – both human and tools.

b) **"Aesthetics of model"** – when the organization of the BPMN model is pleasing and improving the look in order to ease its understanding.

Important remarks to MAQ:

(1) MAQ considers only models with the ensured correctness. A BPMN model must be syntacticaly and semantically correct before the model quality can be assessed. This decision is motivated by the fact that it is useless to assess further quality factors if the model does not adhere to the syntax rules of the BPMN notation (syntactic correctness) and does not correspond to the reality of the analysed situation (semantic correctness). Similarly, as it is useless to discuss the quality of the software which does not fulfil customer requirements.

(2) MAQ only provides but does not take into further consideration characteristics with *canBeMeasuredAutomatically=false*. Investigation of these characteristics is an area of further research.

4.2. Selection of Quality Metrics

Having outlined the quality subcharacteristics (Section 4.1), an important question arose such as how to measure the subcharacteristics and how to interpret the measured values. In order to evaluate the quality of subcharacteristics, a set of metrics to measure models of business processes was required. The relevant metrics were identified by performing a literature review restricted by a proposed set of selection criteria. The focus of literature search was to obtain quality metric(s) for each of the quality subcharacteristics, the knowledge of how to calculate each metric and what its values mean.

For each quality subcharacteristic, a metric or a set of metrics were chosen from the literature and rationalized. The choice was based on the selection criteria, which were questions that relevant literature describing metric had to affirmatively answer. Selection criteria for the literature were as follows:

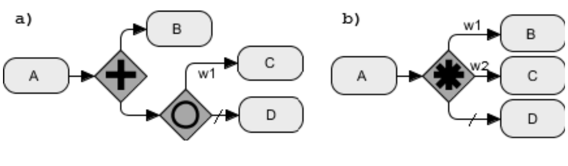
1. Is the metric useful (or can the metric be useful after changes or adjustments) for the context of modelling of business processes in BPMN?
2. Is it possible to calculate the metric for business process models in BPMN (directly or after changes or adjustments)?
3. Is the method of calculating the metric well described in the literature (or is it possible to propose the method of calculating the metric logically)?
4. Is there a general trend that identifies a good or bad value of the metric known in the literature?
5. Do not the metrics limitations exclude it from being applicable to the relevant subcharacteristic(s)?

Only a few of the selected metrics were created for BPMN models, e.g. Control-flow Complexity metric [31] or Cross-connectivity metric [32]. Many more metrics were originally proposed in the related to BPMN areas, e.g. to measure UML models [5, 30], to measure business processes modelled in the YAWL language [33], or were adjusted to a new purpose from mature metrics used in software engineering, especially used in object-oriented software engineering, e.g. [1, 8, 34].

The metrics selected from the literature are listed below. As previously explained, for some metrics it was necessary to introduce and apply additional assumptions, adjustments or changes in order to be able to use the metrics on the actual models in BPMN. One general assumption to the method of calculation of metrics from the literature was adopted in this paper. There are 5 different types of gateways in BPMN. These five types of gateways are: Exclusive Gateway, Event-Based Gateway, Inclusive Gateway, Parallel Gateway and Complex Gateway. In MAQ both Data-Based Exclusive Decision/Merge Gateway and Event-Based Exclusive Decision/Merge Gateway are considered as XOR gateways wherever XOR gateway is stated in the definition of the metric. The distinction between data-based and event-based XOR gateways is based on whether the information required to make the decision is available within the process (data-based is used) or comes from an external source (event-based is used). But both XOR represents a decision to take exactly one path in the flow so from the point of view of metrics in MAQ they are considered as XOR gateways and calculated in the same way.

Details of the selected quality metrics and the applied changes to their original definitions are presented below. Due to the fact of the limited space, in order to get familiar with a method of calculating of metrics, please refer to the indicated literature references.

1. **Coupling metric (CP)** [8]
   Short description: CP metric calculates the degree of coupling, which is related to the number of interconnections among the tasks of a process model.
   Desired values: Low CP values are desired. The higher coupling value of the process, the more difficult it is to change the process and the higher probability that there will be errors in the process.
   Assumptions or changes: The original metric [8] considers AND, OR, XOR gateways and does not provide the method of how one should calculate *connected* function between gateways. In order to be able to calculate actual examples of BPMN models two additional assumptions to the method of calculating the metric were required:
   t_1, t_2 – activities; $g_1, \ldots, g_n$ – gateways
   a) $\mathit{connected}(t_1, t_2) = 0$, if $(t_1 \rightarrow g_1 \rightarrow \ldots \rightarrow g_n \rightarrow t_2) \wedge (g_1 \neq \ldots \neq g_n) \wedge (t_1 \neq t_2)$
   b) $\mathit{connected}(t_1, t_2) = 0$, if $(t_1 \rightarrow$ Complex Gateway $\rightarrow t_2) \wedge (t_1 \neq t_2)$
   Following examples illustrate situations where the assumptions are needed:



2. **Control-flow Complexity metric (CFC)** [31, 35]
   Short description: CFC is an additive metric. In order to calculate the complexity of a process, one should add the control-flow complexity value of all split and join constructs.
   Desired values: Low CFC values are desired. The greater the overall structural complexity of a process is, the higher value of the CFC will be obtained.
   Assumptions or changes: CFC metric distinguishes between AND, OR, XOR gateways. In order to allow calculating actual BPMN models with all possible constructions, in MAQ split for Complex Gateways results in value 0 so that it does not change the result of calculations of CFC metric for the whole BPMN model. Further research may consider changing this method of calculating Complex Gateway.
3. **Cross-connectivity metric (CC)** [32]
   Short description: Let a process model be given by a set of nodes and a set of directed arcs. Each arc goes from a source node to a destination node. CC metric is designed to measure the strengths of arcs between model nodes. It aims to capture the cognitive effort to understand the relationship between every pair of process model elements.
   Desired values: High CC values are desired. The more difficult it is to understand the model or the model is more likely to include errors; the lower the CC value is assigned to the model.
   Assumptions or changes: CC metric is very sensitive to the syntactic correctness of the BPMN model. The following is a list of additional assumptions that were applied to CC metric so that the algorithm could calculate the actual BPMN models:
   a) The original CC metric [32] does not address the problem of BPMN models with events and how events influence the results. In BPMN it is hard to model without events. For different types of events (start events, end events and intermediate events) the weight of node is assumed to be equal 0. Further research should provide information if this value should be different or if there should be a spectrum of weights for different event nodes.

b) The original CC metric [32] does not consider type of gateway other than AND, OR, XOR. In order to allow calculating actual BPMN models, metric result for the models with Complex Gateways is currently set as "Undefined." Further research may consider stating this value.
c) CC metric cannot calculate business process models smaller than two elements (e.g. two tasks), however it seems to be of minor importance because the majority of models are more complex.
d) Following Vanderfeesten et al. [32], CC metric is based on the assumption that tasks in a model have at most one input and output arc while connectors can have multiple input and output arcs. Therefore, the metric result for other BPMN models is currently set as "Undefined."

4. **Imported Coupling of a Process metric (ICP)** [1]
Short description: ICP is a coupling metric that focuses on process if it is highly dependent on external services offered by other processes.
Desired values: Low ICP values are desired. The higher ICP value, the more dependent the process is on the services offered by other processes, what might increase delays, costs and error probability.
Assumptions or changes: Metrics in MAQ ought to provide a value for the whole BPMN model. The original ICP metric by Khlif et al. [1] calculates a result for each single task or sub-process in BPMN model. In MAQ, ICP metric for the whole business process model is defined as the greatest ICP value obtained by any of its tasks or sub-processes. Additionally, in order to properly calculate ICP values for models in BPMN, associations and data associations should also be included. Therefore the changed metric counts the sent message flows, sequence flows, associations and data associations.
5. **Exported Coupling of a Process metric (ECP)** [1]
Short description: ECP is a coupling metric that focuses on process and its influence on the whole model based on how many other processes dependent on its services.
Desired values: Low ECP values are desired. The higher ECP value, the more other processes depend on the services of the process, what might increase delays, costs and error probability.
Assumptions or changes: The assumptions for ECP are nearly the same as the assumptions for ICP, but for the fact that the changed metric counts the received message flows not the sent message flows.
6. **Fan-in/fan-out metric (FIO)** [36]
Short description: FIO metric can be used to analyse the complexity of business process model based on the modular structure. Modular modelling is supported in BPMN by sub-processes. The metric is similar to the metric proposed by Khlif et al. [1], however, it does not include length.
Desired values: Low FIO values are desired. The higher structural complexity of a model or sub-model according to the FIO value, the more difficult it is to use the model and there is more likelihood that it is badly designed.
Assumptions or changes: Metrics in MAQ ought to provide a value for the whole BPMN model. The original FIO metric [36] counts only sub-processes (it does not count tasks) and calculates a separate result for each single sub-process in BPMN model. In MAQ, FIO metric for the whole business process model is defined as the greatest FIO value obtained by any of its sub-processes. Due to the fact that from the definition of the metric it is not clear what should be adopted if the model does not have a modular structure, if the model does not have sub-processes, in MAQ FIO value is assumed to be equal zero.
7. **Number of Activities, Joins and Splits (NOAJS)** [37]
Short description: Splits in BPMN do not necessarily have corresponding joins. NOAJS complexity metric can measure such not well structured processes based on counting activities, joins and splits together.
Desired values: Low NOAJS values are desired.

Assumptions or changes: None.

8. **Interface complexity of an activity metric (IC)** [37]
Short description: IC metric can be used to evaluate the complexity of processes.
Desired values: Low IC values are desired.
Assumptions or changes: From the original definition of the metric [37], it is not clear how *Length* should be calculated for BPMN models. In MAQ, it is calculated as follows: *Length*=1 for a task element and *Length*=3 for a sub-process (representing sub-processes as a collection of activities).
Metrics in MAQ ought to provide a value for the whole BPMN model. The original IC metric [37] calculates a result for each single activity. In MAQ, IC metric for the whole model is defined as a sum of all IC values obtained by all activities in the model. These will reduce a limitation of the original metric, which can give the result zero as the value of complexity if an activity has no external interactions, e.g. for the end activities of the process.
Assumption for the following points 10, 11 and 12 describing Halsted-based Process Complexity metrics: The following refinement of the metrics is used in MAQ:
n_1 – number of unique activities, splits, joints and control-flow elements.
n_2 – number of unique data objects, data inputs, data outputs and data stores (duplicates removed).
N_1 – number of unique types of activities and control-flow elements used in BPMN model, e.g. task, sub-process, XOR gateway, OR gateway, etc.
N_2 – number of unique data types used in the BPMN model – data objects, data inputs, data outputs and data stores.

9. **Halsted-based Process Difficulty metric (HPC_D)** [37]
Short description: HPC_D is a quantitative measure of complexity and is aimed to calculate a difficulty of the process.
Desired values: Low values are desired.
Assumptions or changes: HPC_D has a limitation, because its value cannot be calculated if n_2 equals 0. In MAQ, the result of such calculation is set as "Undefined."

10. **Halsted-based Process Length metric (HPC_N)** [37]
Short description: HPC_N is a quantitative measure of complexity and is aimed to calculate a length of the process.
Desired values: Low values are desired.
Assumptions or changes: HPC_N metric can be calculated only if (n_1>0 and n_2>0), otherwise it cannot be calculated because the log value is undefined. In MAQ, the result of such calculation is set as "Undefined."

11. **Halsted-based Process Volume metric (HPC_V)** [37]
Short description: HPC_V is a quantitative measure of complexity and is aimed to calculate a volume of the process.
Desired values: Low values are desired.
Assumptions or changes: HPC_V metric can be calculated only if (n_1+n_2 >0), otherwise it cannot be calculated because the log value is undefined. In MAQ, the result of such calculation is set as "Undefined."

12. **Sequentiality metric (S(G))** [12]
Short description: S(G) is a structural metric. The sequentiality ratio is the number of arcs between none-connector nodes divided by the number of arcs.
Desired values: High S(G) values are desired. The higher S(G) value, the less likely it is to have errors in the overall model.
Assumptions or changes: None.

13. **Number of Nodes metric (Sn(G))** [12]
Short description: Sn(G) is a structural metric that calculate the number of nodes of process model.
Desired values: Low Sn(G) values are desired. The higher Sn(G) value, the more likely it is to have errors in the overall model.
Assumptions or changes: None.

14. **Number of Activities metric (NOA)** [37]
Short description: NOA metric sums up activities in a business process model. It is a simple and popular metric that can be used to measure complexity.
Desired values: Low NOA values are desired.

Assumptions or changes: None.

15. **Coefficient of Connectivity metric (CNC(G))** [12]
 Short description: CNC(G) is a structural metric. The coefficient of connectivity gives the ratio of arcs to nodes in BPMN models.
 Desired values: Low CNC(G) values are desired. The higher CNC(G) value, the more likely it is to have errors in the overall model.
 Assumptions or changes: None.
16. **Cognitive complexity measure (W)** [33]
 Short description: Cognitive complexity measure is a cognitive weight proposed to measure the effort needed for comprehending the model.
 Desired values: Low W values are desired. The higher W value, the more difficult it is to understand the model.
 Assumptions or changes: Cognitive weights of business process model elements in [33] were proposed for YAWL language. Based on the analogy with BPMN language, the adequate cognitive weights for BPMN models are proposed in Table 1. In MAQ, the cognitive weight of a BPMN model is defined as a sum of the cognitive weights of its individual elements.
17. **Density metric (D(G))** [12]
 Short description: D(G) is a structural metric that calculates the ratio of the total number of arcs to the maximum number of arcs.
 Desired values: Low D(G) values are desired. The higher D(G) value, the more likely it is to have errors in the overall model.
 Assumptions or changes: None.

The chosen quality metrics were assigned to quality subcharacteristics of MAQ based on information derived from the literature. Some metrics are useful for more than one subcharacteristic (please refer to metamodel in Fig. 1).

Rationale for assigning metrics to subcharacteristics is as follows:

Quality subcharacteristic: **"Changeability"** has the following metrics assigned:

- CP – The lower value of coupling, the easier to change the process [8].
- CFC – Models with a reasonable complexity are easier to be modified and maintained. The metric may help to develop simpler processes when it is possible [35]. Following [38], CFC metric is suitable to measure changeability.
- D(G) and S(G) – In [4], conducted experiments showed that Density and Sequentiality metrics are closely connected with modifiability.
- HPC_D, HPC_V and HPC_N – Metrics can predict maintenance effort [37].
- ECP and ICP – Business process models that have high coupling metric are difficult to be changed or maintained because they have a high level of informational dependency between activities [1].

Quality subcharacteristic: **"Reusability"** and extensibility has the following metric assigned:

- FIO – In accordance with [36], FIO metric detects poor modularization. If modularization is used in a reasonable way, dividing a model in modular sub-models can lead to smaller, reusable models.

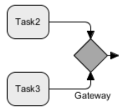
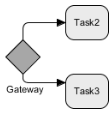
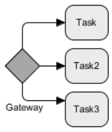
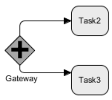
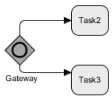
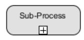
Quality subcharacteristic: **"Minimality and Simplicity"** has the following metrics assigned:

- FIO – Following [36], if the examined sub-process in the model has both large fan-in and fan-out, this may indicate that the model does not have an appropriate size or was not partitioned into modules in a sensible way. Redesigning in this situation could improve the sub-process.
- NOA and NOAJS – These simple metrics may show models that are badly designed with an excessive number of activities [1].

Quality subcharacteristic: **"User comprehensibility"** has the following metrics assigned:

- CFC – It is easier to understand and maintain business process models which have low complexity. Business processes should minimize their complexity in order to be helpful to the various stakeholders [35]. Following [38], CFC metric is suitable to measure understandability.
- CC – Models with a high cross-connectivity can facilitate understanding of business processes among various stakeholders [32].

Table 1. Proposition of cognitive weights for BPMN models in Cognitive Complexity Measure.

BPMN structure	*BPMN symbol*	*Cognitive weight*
Single consecutive step in a work-flow		**1**
All joins. In [33], the metric was originally defined only for business process models that are well-structured. In BPMN, corresponding joins are not necessary. The weight of join elements is considered as equal to the cognitive weight of sequence elements.		**1**
XOR-split (exactly one of two branches is chosen)		**2**
XOR-split (exactly one of more than two branches is chosen)		**3**
AND-split		**4**
OR-split or Complex Gateway		**7**
Sub-process (can be used for decomposing BPMN models)		**2**
Start or End event		**2**
Intermediate event (both intermediate events attached to the boundary of activities and intermediate events within the normal flows)		**3**

- NOA and NOAJS – The metrics provide some information about the understandability of designs [37].
- W – The measure can state whether models are easy or difficult to comprehend [33, 38].
- FIO – The metric was developed for analysing the modularization; modular sub-processes can help to make the model easier to comprehend [36].
- Sn(G), CNC(G) and S(G) – In [4], conducted experiments showed that the metrics are closely connected with user's understandability.
- CP – High complexity in a process may result in bad understandability, therefore, process complexity should be kept in low level [8].
- IC – The metric is a measure of complexity of process models and complexity measures the understandability of a design [1].

Quality subcharacteristic: **"Aesthetics of diagrams"** has the following metrics assigned:

- CP – Business process models with high CP metric have complicated connections, which can be reflected in the organization of BPMN models [8].
- ICP and ECP – The organization of BPMN models with high ICP or ECP metrics may not be clear and thus difficult to understand. The coupling metrics detect models in which multiple processes depend on each other, which may influence the look of the whole design.

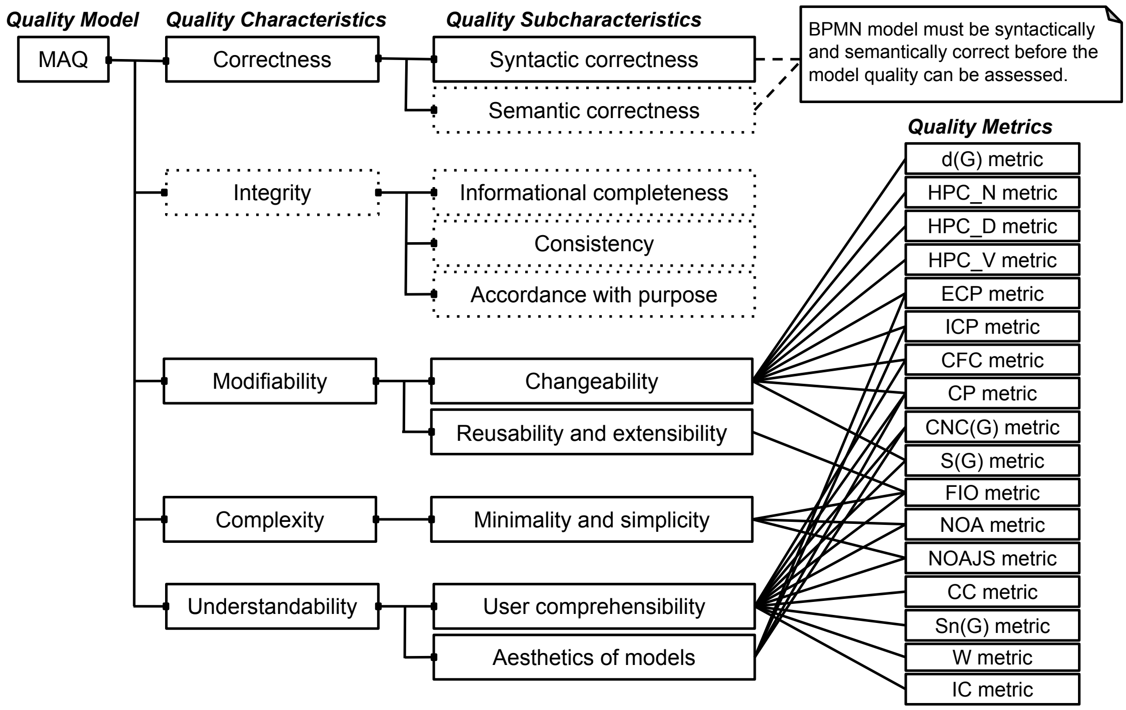


Figure 2. A schema of the hierarchical structure of MAQ.

- CNC(G) – In formal esthetics the coefficient of network complexity measure is considered with the notion of elegance [37].

Figure 2 presents a schema of the hierarchical structure of the extracted quality characteristics, quality subcharacteristics and quality metrics.

4.3. Selection of Quality Criteria

Very rarely does it happen that the literature indicates which values of metrics are good or bad, and an accurate analysis of the results is mostly left to the user. This is not a problem if the user is an expert and a quick analysis of multiple metrics and models is not required. The purpose of MAQ is to automate the process of model quality assessment. Therefore, it is very important to define exact quality criteria as functions that appraise the results of quality metrics.

More often than specific numbers, the authors of metrics indicate the general trend of metrics' results, e.g. the lower (higher) value of a metric, the better model. Therefore, one of the selection criteria for the literature in Section 4.2 rejected literature and metrics for which this trend was not clear. With this knowledge in mind, quality criteria for metrics based on the results obtained from measuring models from a pre-prepared BPMN repository. The repository contained 57 business process models in BPMN; collected from five different Internet sources [13]. The identified BPMN models had varying quality in different sources because they were created by users with different levels of experience in BPMN. The repository contained officially correct BPMN examples given by the OMG, models from master and doctoral theses and models created by various individuals, who had less experience with BPMN. This variety of models helped to define which results of metrics were obtained by high and low quality models. An effort was made to collect models of diverse quality, however it poses a threat to validity. In order to be able to examine the repository of BPMN models and to propose quality criteria, two tools were needed:

1. A tool that implements all the chosen quality metrics from Section 4.2. This tool was created and is reviewed in Section 5.
2. Additional statistical software which contains tools for clustering. In this case Weka

software was chosen and simple k-means function was selected for clusterization.

The algorithm of k-means clustering was developed by Hartigan and Wong [39]. In MAQ, in the k-means clustering function, the k value was declared as equal 4 or 2, based on the results of metrics used on the repository. The seed value for each metric was chosen individually as integer value without rounding from the equation: maximal metric's value minus minimal metric's value divided by 2.

Quality rating in MAQ is defined as an ordinal scale that describes whether the result of the metric is of good or bad quality on a scale of *Class A* (highest) through *Class E* (lowest). The chosen scale is ordinal since the quantitative levels of quality had varying distances between them for each metric. For example, the range of values obtained from the BPMN repository for CP metric had a range of 0 to $0.\overline{3}$ and the CFC metric had a range from 0 to 16. Clearly, the ranges between metrics were different in practice and not easily comparable without the use of an ordinal scale. The ordinal scale was created using results of measurements of the repository. The values for each entity of the ordinal scale were based on an interval of values which were relevant to each metric. For example, the CFC metric had an observed range from 0 to 16. The trend of values for each metric suggests what are good or bad values in terms of quality for each metric. This information was used to assign intervals to quality ratings. For example, in the case of CFC metric, the value should be low in order to attain a good quality. Clusterization of the metric was then used to create intervals of ordinal elements, e.g. zero to one for *Class A*, from one to four for *Class B*, etc.

The example calculations of quality criteria presented below, are based on CP metric. A full list of the defined quality criteria is available in [13].

Summary of CP metric:

- Type of measurement method: *Objective*,
- Scale of theoretically possible results: *Real from zero to infinity*,
- Scale of results obtained by models from the repository: $[0.0, 0.333333]$,
- *Low CP values are preferable* for high quality models.

Weka software settings: Simple k-means function, Number of clusters: 4, Seed: 0. Clusters obtained through the use of the software were as follows: *cluster 0:* 0.114167, *cluster 1:* 0.176786, *cluster 2:* 0.064361, *cluster 3:* 0.003934.

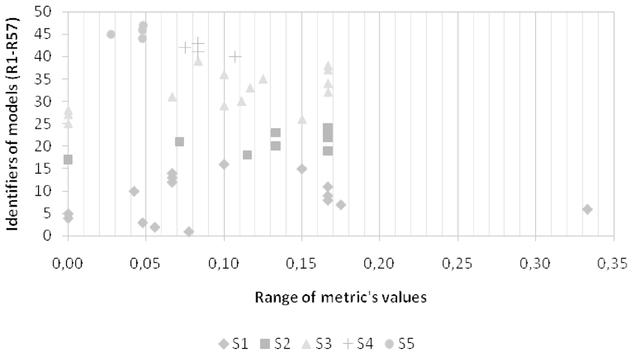


Figure 3. Result of measures of CP metric on the repository of BPMN models.

Table 2. Assignment of results to quality ratings

Range of results	*Quality Rating*
[0.0, 0.003934)	Class A
[0.003934, 0.064361)	Class B
[0.064361, 0.114167)	Class C
[0.114167, 0.176786)	Class D
[0.176786, ∞)	Class E

The obtained quality criteria for CP metric:

$$QC() = \begin{cases} Class\ A, \text{if } CP \in [0.0, 0.003934) \\ Class\ B, \text{if } CP \in [0.003934, 0.064361) \\ Class\ C, \text{if } CP \in [0.064361, 0.114167) \\ Class\ D, \text{if } CP \in [0.114167, 0.176786) \\ Class\ E, \text{if } CP \in [0.176786, \infty) \end{cases}$$

4.4. Selection of Quality Functions

Quality functions combine the results of quality criteria for both quality subcharacteristics and the overall quality of actual BPMN models. In this way, they indicate whether the model quality is good or bad. More specifically, the result of quality function for e.g. *"Minimality and Simplicity"* subcharacteristic has to combine results of quality criteria for FIO, NOA and NOAJS metrics. Based on this example, the quality function determines what should be stated as an overall quality if, for example, criteria for FIO results in *Class B*, NOA in *Class C* and NOAJS in *Class B*.

There are many possible interpretations for how to propose quality functions. For example, the function for a subcharacteristic could be calculated as follows:

– The best quality rating obtained by any of metrics assigned to the subcharacteristic QF_{sch} = min {QualityRating}
– The ceiling of the mean quality rating obtained by metrics assigned to the subcharacteristic. Let *Class A*=1, *Class B*=2, *Class C*=3, *Class D*=4, *Class E*=5 $QF_{sch} = (\text{QualityRating}) \lceil \sum_{m=1}^{M} QR_m/M \rceil$ where M is a number of the assigned metrics
– etc.

In MAQ, quality functions were proposed taking into account the following issues:

a) Not every quality metric can be calculated for each actual BPMN model. Some metrics may result in an "Undefined" value. Hence the need for differentiation in interpretation between metrics that always result in a real value (ECP, ICP, d(G), Sn(G), HPC(V), CNC(G), CFC, W, FIO, NOAJS, NOA, IC) and metrics that may result in an "Undefined" value (HPC(D), HPC(N), S(G), CC, CP).

b) Metrics that result in an "Undefined" value for an actual BPMN model should be excluded from the calculation of quality and only metrics that result in a non-undefined result should have influence on the quality.

c) The proposed quality functions use a Fibonacci sequence and the ceiling function. The Fibonacci sequence may help in addressing the differences between results of metrics whose values are not easy to be directly compared. The distance between quality ratings varies depending on quality criteria. Fibonacci sequence seems to be relevant since the direct comparing of quality rating as ratios of each other (e.g. *Class E* as half of the quality of *Class D*) would overestimate the result. It seems to be relevant also because when the quality is low, it is important that this is clear to the user. The Fibonacci sequence increases rapidly from the initial value of 1 to 8. As a result, the quality rating for bad quality model can be represented using higher values such as 8 in order to make the whole quality function more sensitive to bad quality. In order to make the rating more sensitive to a bad quality, the Fibonacci sequence is used starting from the third value. To summarize, Fibonacci sequence and ceiling function are chosen because they are more informative to have results that are sensitive to a low quality. These more sensitive results show clearly when the quality is low. Nevertheless, at this stage of research in the field it is difficult to assess if this interpretation is acceptable. Further research should investigate which interpretation of quality function is best for combining metrics for models in BPMN.

Quality functions for quality subcharacteristics are defined as follows (quality function for the whole BPMN model is analogical):

$$QF_{sch}() = (\text{QualityRating}) \lceil \frac{\sum_{m=1}^{M} \text{QRvalue}(m)}{M} \rceil$$

where:

$$QRvalue(m) = \begin{cases} 1, \text{ if } QC(m) = ClassA \\ 2, \text{ if } QC(m) = ClassB \\ 3, \text{ if } QC(m) = ClassC \\ 5, \text{ if } QC(m) = ClassD \\ 8, \text{ if } QC(m) = ClassE \end{cases}$$

m – Quality metric assigned to the quality subcharacteristic, which produced a non-undefined result for the measured BPMN model.
M – Number of the assigned quality metrics.

(QualityRating) – The result of the equation is transferred into the adequate quality rating with casting so that the lower (worse) value of quality rating is assigned. An example: *ClassA*=1, *ClassC*=3, *ClassE*=8 results in: $QF_{sch}() = (QR)\lceil \frac{1+3+8}{3} \rceil = (QR)\lceil 4 \rceil = Class\ D$

5. BPMN Quality Tool

BPMN Quality Tool[1] is a plug-in implemented in Java language to Business Process Visual ARCHITECT (Simulacian) – well-known software

[1] The plug-in is available online and can be found in: <https://sourceforge.net/projects/bpmn-quality/>.

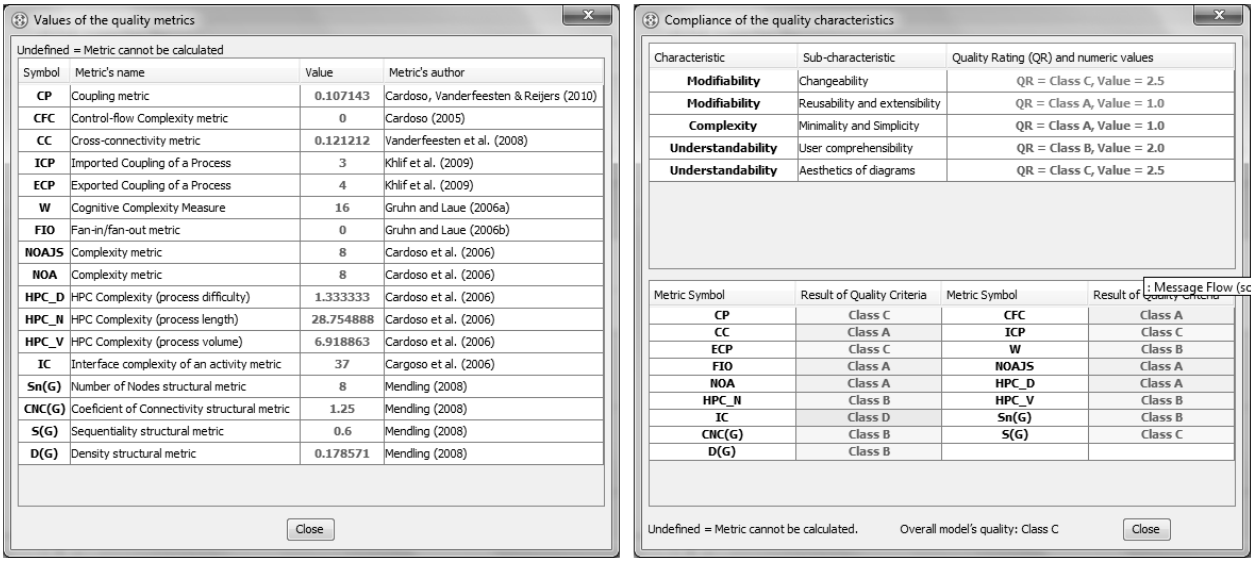


Figure 4. Example of measures of MAQ on a BPMN model.

for modeling in BPMN (tested in the 4.0 version of the software).

5.1. Initial Functionality of the Tool

BPMN Quality Tool in its initial functionality allows for measuring and displaying values of quality metrics for the actual BPMN models (example is presented on the left side of Figure 4). This functionality was used to gather data needed to propose quality criteria for metrics (described in Section 4.3).

An additional functionality of the plug-in, available through a pop-up menu, shows relationships of BPMN elements in actual models. "Show Relationship of Element" option lists all relationships of the chosen element (example can be found in Fig. 5). Especially it provides information about:

- the type of flow going to or from the chosen element (it distinguishes between sequence flow and message flow),
- the name of the flow (if the flow has a name, "Unnamed" otherwise),
- the direction of the flow (if flow goes "To" or "From" the chosen element),
- the icon of the BPMN model's element to or from which the flow goes.

This functionality may help to get more information about elements in complex BPMN models or in models with bad aesthetics, e.g. where arches cross. The analysis of the relationships of elements was a base to implementation of quality metrics presented in Section 4.2.

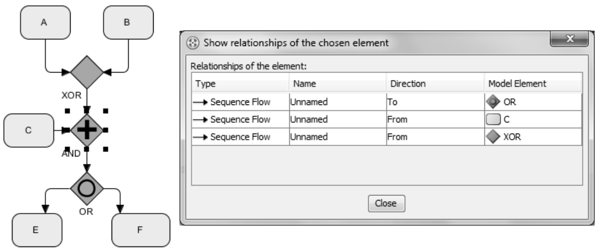


Figure 5. Example use of "Show Relationships of Element" option for the AND gateway element.

5.2. Implementation of MAQ

The final functionality of *BPMN Quality Tool* allows for assessing the quality of actual business process models in BPMN. "Quality Assessment" option is an implementation of the developed MAQ. The option shows a quality of sub-characteristics and an overall quality (example is presented on the right side of Fig. 4).

6. Preliminary Evaluation of MAQ

This section describes the process of how the preliminary evaluation of MAQ was conducted by a survey-based experiment and what results were obtained. The survey can be found in [13] in Appendix E.

6.1. Survey Study Design

Research question: The survey and survey-based experiment was aimed to provide an answer to the question **"Is the developed model for assessing the quality of business process models in BPMN considered as useful?"**.

Types of questions: The survey was based on the questionnaire which consisted of closed questions. This form was chosen because it allowed for a more quantitative feedback. In some surveys the respondents left additional comments. The comments mostly added details to the questions of the survey. The most interesting suggestions were about some additional functions the tool for assessing the quality of the models should have so that it would be useful for modelers. And there were also very important comments which helped to improve the definitions of the characteristics. These responses were later analyzed in a qualitative manner.

Population of the survey: The survey population consisted of experts on BPMN who used BPMN in work, research or for private purposes. The author identified potential experts to be contacted, however the final classification if someone is or is not an expert was based on how high the respondent rated his or her knowledge of BPMN (additional question to the survey).

The initial question in the survey was: "Please select a number which best describes the level of your knowledge of BPMN notation" on a five point scale, where 1 meant a novice and 5 meant an expert. Experts in the survey were respondents who declared the level of their BPMN knowledge as 3 or more, as well as ticked that they used BPMN in their work, research or for private purposes.

In total 14 expert responses were obtained from 125 potential experts contacted.

6.2. Survey-Based Experiment

Objective and Design: The aim was to evaluate practical usefulness of the MAQ model. This was done by comparing assessment given by the tool that implemented MAQ and the assessment given by the expert respondents. It was checked if the respondent's evaluation of the quality of BPMN models agreed with the results determined by the tool. This indicated how useful MAQ and the tool for automatic assessment of the quality of business process models in BPMN were. The respondents assessed three BPMN models based on the identified quality subcharacteristics. The subcharacteristics for each BPMN model were assessed using the previously introduced quality rating from *Class A* to *Class E*. The same process was done by the tool. Later, results were compared and analysed. The goal of the experiment was defined according to the goal template in [40] as follows:
analyse the quality of BPMN models
for the purpose of the evaluation of MAQ
with respect to its accuracy in the evaluation of quality of business process models in BPMN
from the point of view of BPMN experts
in the context of the business process modelling domain.
Objects: The objects were BPMN models.
Subjects: Responses from expert respondents.
Independent variables: Independent variables were three models in BPMN. The BPMN models for the survey were chosen in order to be of either good or bad quality. Chosen models of good quality were visually different. The selected models seemed to be relevant since they represented both good and bad quality models. Due to the fact that the choice of the models was based on the author's knowledge and the chosen models could possibly be not representative for the whole population of business process models in BPMN it posed a threat to validity (please refer to Section 7). Besides quality, a number of other concerns were taken into consideration when selecting the BPMN models for the survey. Firstly, the models needed to be non-trivial by representing processes that could be of importance. For example, the three BPMN models

represented a trouble ticket system, a purchase ordering process and a software upgrade process. Secondly, the BPMN models needed to be reasonably complex, so that the respondents could easily understand them. This was seen in the models as they did not consist of more than 50 elements (including flows). Thirdly, the number of BPMN models chosen for the survey had to be limited in order to increase the response rate of the survey – more models could discourage the respondents. All of these reasons contributed to the choice of three non-trivial and appropriately complex BPMN models.

Dependent variables: The assessment by the tool and the assessment by experts that responded to the survey were the two dependent variables of the experiment. Dependent variables of the experiment were selected in order to understand the correlation between the respondent's ratings of the quality versus the tool's using the quality rating scale.

Hypotheses: Null hypothesis H0: The mean of the survey result for each characteristic was equal to the MAQ model's result.

Alternative hypothesis H1: The mean of the survey result for each characteristic was not equal to the MAQ model's result.

The hypotheses were tested using a student's t-test and the 95% confidence interval of the mean of the expert responses.

6.3. Results of the Preliminary Evaluation

The graphical charts that present the comparison of expert responses with tool response are available in Section 10.2 of [13]. In the survey-based experiment, the equality of the mean of expert's assessment of the quality of the surveyed BPMN models with that of the tool was mostly not rejected. The results of the MAQ model fell within the confidence interval for the characteristics of "Changeability" and "User comprehensibility" for all three models indicating that the equality of expert's responses and the responses of the tool cannot be rejected for the models surveyed. The hypothesis of expert and tool agreement was rejected in one of the three models for "Aesthetics of the model." However, "Reusability and extensibility" was rejected in two of the three models and "Minimality and simplicity" was rejected in all models, indicating that they mostly did not equal the response of experts. The reason mostly lay in the fact that currently there is very little research on metrics that can calculate BPMN models and be proper indicators for "Minimality and simplicity" subcharacteristic (only 3 relevant metrics were found) and "Reusability and extensibility" subcharacteristic (only 1 relevant metric was found and the metric takes into consideration only the models with sub-processes). The MAQ is built in a way that it is easily extensible if future research proposes updated or new metrics relevant for the subcharacteristics. Due to the fact that the number of the collected responses to the survey cannot be accepted as representative, the obtained results may be used only as suggestions for the direction of the research, if it is correct and helpful. It is identified as a threat to validity in Section 7.

7. Threats to Validity

The author has identified a number of threats to validity. The following is the explanation of them and their mitigating factors to the developed model. Threats to validity for research in the field of software engineering are presented as consisting of four types, which are: construct validity, conclusion validity, internal validity and external validity [40]. For each type of validity threat, risks that could pose a threat to the validity of MAQ are identified.

Conclusion validity threats are issues which affect the way conclusions are made from treatments.

Reliability of measures: When measurements are not consistently applied it can create a risk that the validity is threatened. The created tool assured that all metrics, quality criteria and quality functions were calculated automatically so that the threat that calculations would be unreliable was mitigated.

Random heterogeneity of BPMN models: Variation poses a risk when the objects under study are heterogeneous. The BPMN models needed to be heterogeneous in terms of quality since good and bad quality was under assessment. In order to mitigate the consequences of heterogeneous BPMN models, the author narrowed the focus to Process Diagrams in BPMN 2.0 notation.

Internal validity threats are concerned with whether the relationship between the treatment and outcome is causal. This means that the relationship between the treatment and outcome cannot be caused by some unknown factor.
Quality assurance of SLR: A threat to validity is posed for the systematic literature review by the fact of lack of quality assurance activities. The activities could include reviewing the selected papers to see if they match the inclusion criteria by another reviewer, developing and verifying the protocol with another reviewer to make sure if the extracted data are correctly interpreted.
Literature review for metrics: The choice of the literature was restricted by the proposed selection criteria for the literature. There is a threat to validity that an important work could have been omitted.
Selection of BPMN models: BPMN models were selected by the author of the article. The models were selected from publicly available sources where the licensing allowed for them to be used for research. This poses a threat to validity since models which are licensed in a research-friendly way could be different than BPMN models in general. In order to mitigate this, 57 models from 5 different sources were collected so that a more general selection could be achieved. Furthermore, a number of models needed to be redrawn, so they could have a file format supported by a tool created by the author. The author tried to rewrite the models without defects, however, rewriting models always possibly may increase defects.
Selection of BPMN models in survey: There were three models chosen to be used in the survey. They were selected based on their varying quality in respect to the quality subcharacteristics measured. By selecting only three models of the 57 models there is a threat to validity that the models which were selected were not representative of the whole population of business process models.
Selection of experts to the survey: Naturally there is a variation in the level expertise in the field of BPMN, and so the experts contacted may not be representative of the whole BPMN expert population. The author tried to mitigate this threat by contacting experts directly and also verifying through the survey that they considered themselves to be experts.

Construct validity threat refers to the theory and the observation if they are related in a causal way.
Lack of metric validation: If the metrics used in the paper have not been validated theoretically or empirically and were used in the MAQ model. Some metrics were additionally adjusted to the need of MAQ by the author. The selected metrics come from scientific research and so that the selection was based on peer-reviewed metrics. Nevertheless, it poses a threat to the validity of the model.

External validity threats are concerned whether the result of the research is generalizable to a larger scope.
Interaction of selection and treatment: This is a threat when the subject of a study is not representative of the general population. Quality characteristics and subcharacteristics were selected using a SLR with a well-defined protocol that followed systematic guidelines [17]. Furthermore, the metrics were selected using a defined set of selection criteria that was applied to the literature. As a result, a well-defined methodology was applied in order to collect quality characteristics and metrics from the population of scientific literature relevant to BPMN models.
Generalization of survey responses: There is a serious risk that the survey respondents are not generalizable to the population of practitioners when the sample used is not representative. The population used cannot be considered as representative for the general population. Therefore, the results obtained from the survey-based experiment may be used only as suggestions as to whether the direction of the research is correct and helpful to practitioners but they cannot be considered as a statistically significant result.
Generalization of BPMN models: Threats to va-

lidity as a result of BPMN models not being generalized to the population pose a risk since the result can only be generalized to an appropriate scope. The repository of BPMN models consisted of only BPMN 2.0 Process Diagrams. This means that the conclusions cannot be generalized to other types of BPMN diagrams.

8. Conclusions

The quality of business process models is important in the area of model-based software development. The need for high quality of models is supported by many arguments both industrial and research based. This paper focuses on a practical proposal of a model for quality assessment of actual models in BPMN, called MAQ. The first part of the paper presents a metamodel of the MAQ. The metamodel defines a structure of the MAQ and is built upon the information presented in ISO/IEC 25010 [14] standard. Later on, all parts of the MAQ are described. The most important MAQ parts are: quality characteristics, quality subcharacteristics, quality metrics, quality criteria and quality functions. Later section shows BPMN Quality Tool which implements MAQ and can be helpful for modelers to ensure that generated actual BPMN models are correct and properly built. The model was preliminary evaluated for usefulness through a survey-based experiment in [13].

MAQ aims to assess only the quality characteristics which can be measured in isolation from any additional information about the domain, but the BPMN model itself. Therefore, MAQ only lists but is not designed to automatically assess subcharacteristic of *"Syntactic correctness"*, characteristic of *"Integrity"* and its subcharacteristics: *"Informational completeness"*, *"Consistency"* and *"Accordance with purpose"*. This might be considered as a limitation of MAQ.

Not all of the metrics chosen from the literature were validated. Additionally, some of the original metrics operated or were tested only on a subset of BPMN elements. Therefore, some additional changes to the original metrics were necessary to be applied in order to be able to measure actual models in BPMN. These changes also have not yet been validated. Future research may validate, change or extend the proposed metrics.

The MAQ and its implementation seem to be a good starting point to further development. The MAQ and *BPMN Quality Tool* can be further extended while new metrics will be introduced, existing metrics will be further developed in order to be able to measure actual models, and new quality criteria or quality functions will be suggested. This may lead to consideration of new perspectives and more compatible correlation between quality characteristics, quality subcharacteristics, quality metrics, quality criteria and quality functions.

References

[1] W. Khlif, L. Makni, N. Zaaboub, and H. Ben-Abdallah, "Quality metrics for business process modeling," *Proceedings of the 9th WSEAS international conference on Applied computer science (ACS'09)*, 2009, pp. 195–200.

[2] O. OMG, "Business process model and notation (bpmn) version 2.0," *Object Management Group*, 2011.

[3] H. A. Reijers, J. Mendling, and J. Recker, "Business process quality management," *Handbook on Business Process Management*, 2010, pp. 167–185.

[4] L. Sánchez-González, F. García, J. Mendling, F. Ruiz, and M. Piattini, "Prediction of Business Process Model Quality Based on Structural Metrics," *Conceptual Modeling – ER 2010, Lecture Notes in Computer Science*, Vol. 6412, 2010, pp. 458–463.

[5] P. Mohagheghi, V. Dehlen, and T. Neple, "Definitions and approaches to model quality in model-based software development – A review of literature," *Information and Software Technology*, 2009, pp. 1646–1669.

[6] T. Rozman, G. Polancic, and R. V. Horvat, "Analysis of most common process modelling mistakes in BPMN process models," 2007. [Online]. http://www.slideshare.net/tomirozman/eurospi2007trozman

[7] I. Dubielewicz, B. Hnatkowska, Z. Huzar, and L. Tuzinkiewicz, "Quality-Driven Software Development for Maintenance," *Emerging technologies for the evolution and maintenance of software models*, 2012, pp. 1–31.

[8] J. Cardoso, I. Vanderfeesten, and H. A. Reijers, "Computing coupling for business process models," 2010. [Online]. http://eden.dei.uc.pt/~jcardoso/Research/Papers/Old%%2020paper%20format/Caise-19th-Coupling-Cardoso-Vanderfeesten.pdf

[9] G. Aagesen and J. Krogstie, "Analysis and design of business processes using BPMN," *Handbook on Business Process Management 1*, 2010, pp. 213–235.

[10] L. Makni, W. Khlif, Z. H. Nahla, and H. Ben-Abdallah, "A tool for evaluating the quality of business process models," 2010. [Online]. http://subs.emis.de/LNI%20/Proceedings/Proceedings177/234.pdf

[11] S. Overhage, D. Q. Birkmeier, and S. Schlauderer, "Quality marks, metrics, and measurement procedures for business process models," *Business & Information Systems Engineering*, Vol. 4, No. 5, 2012, pp. 229–246.

[12] J. Mendling, "Metrics for business process models," in *Metrics for process models: empirical foundations of verification, error prediction.* Berlin, Heidelberg: Springer-Verlag, 2008, pp. 103–133.

[13] M. Sadowska, "Quality of business models expressed in BPMN," M.S. thesis, Wrocław University of Technology, Wrocław, 2013.

[14] "ISO/IEC 25010:2011(E) Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models," 2011.

[15] "ISO/IEC 14598-1:1999(E) information technology – software product evaluation – part 1: General overview," 1999.

[16] S. Wagner, "Quality Models," in *Software product quality control.* Berlin: Springer, 2013, pp. 29–89.

[17] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering, v2.3," *EBSE Technical Report EBSE-2007-01*, 2007, p. 65.

[18] T. Arendt and G. Taentzer, "UML model smells and model refactorings in early software development phases," 2010, pp. 1–5.

[19] J. Becker, M. Rosemann, and C. von Uthmann, "Guidelines of business process modeling," *Business Process Management, Lecture Notes in Computer Science*, Vol. 1806, 2000, pp. 30–49.

[20] F. Fieber, M. Huhn, and B. Rumpe, "Modellqualität als indikator für softwarequalität: eine taxonomie," *Informatik-Spektrum*, Vol. 31, No. 5, 2008, pp. 408–424.

[21] A. A. Jalbani, J. Grabowski, H. Neukirchen, and B. Zeiss, "Towards an integrated quality assessment and improvement approach for UML models," *SDL'09 Proceedings of the 14th international SDL conference on Design for motes and mobiles*, 2009, pp. 63–81.

[22] J. Krogstie and A. Sølvberg, "Conceptual modeling in a quality perspective," 2000, pp. 94–101.

[23] C. F. Lange and M. R. Chaudron, "Managing model quality in UML-based software development," *Software Technology and Engineering Practice*, 2005, pp. 7–16.

[24] O. I. Lindland, G. Sindre, and A. Solvberg, "Understanding quality in conceptual modeling," *Software IEEE*, Vol. 11, No. 2, 1994, pp. 42–49.

[25] J. Mendling, H. A. Reijers, and W. M. van der Aalst, "Seven process modeling guidelines (7PMG)," *Information and Software Technology*, Vol. 52, No. 2, 2010, pp. 127–136.

[26] H. J. Nelson, G. Poels, M. Genero, and M. Piattini, "A conceptual modeling quality framework," *Software Quality Journal*, Vol. 20, No. 1, 2012, pp. 201–228.

[27] R. Schuette and T. Rotthowe, "The guidelines of modeling – an approach to enhance the quality in information models," in *Conceptual Modeling – ER '98*, ser. Lecture Notes in Computer Science, T.-W. Ling, S. Ram, and M. Li Lee, Eds. Springer Berlin Heidelberg, 1998, Vol. 1507, pp. 240–254. [Online]. http://dx.doi.org/10.1007/978-3-540-49524-6_20

[28] S. Si-Said Cherfi, J. Akoka, and I. Comyn-Wattiau, "Conceptual modeling quality – from EER to UML schemas evaluation," *Conceptual Modeling – ER 2002 Lecture Notes in Computer Science*, Vol. 2503, 2003, pp. 414–428.

[29] D. Ssebuggwawo, S. Hoppenbrouwers, and E. Proper, "Assessing Collaborative Modeling Quality Based on Modeling Artifacts," *The Practice of Enterprise Modeling,Lecture Notes in Business Information Processing*, Vol. 68, 2010, pp. 76–90.

[30] B. Unhelkar, "The quality strategy for UML," in *Verification and Validation for Quality of UML 2.0 Models.* Hoboken, NY: Wiley-Interscience, 2005, pp. 1–26.

[31] J. Cardoso, "How to measure the control-flow complexity of web processes and workflows," *Workflow handbook 2005*, 2005, pp. 199–212.

[32] I. T. P. Vanderfeesten, H. A. Reijers, J. Mendling, W. M. van der Aalst, and J. Cardoso, "On a quest for good process models: The cross-connectivity metric," *CAiSE, Springer*, Vol. 5074, 2008, pp. 480–494.

[33] V. Gruhn and R. Laue, "Adopting the Cognitive Complexity Measure for Business Process Models," *5th IEEE International Conference on Cognitive Informatics, 2006. ICCI 2006*, Vol. 1. IEEE, 2006b, pp. 236–241.

[34] G. Muketha, A. Ghani, M. Selamat, and R. Atan, "A survey of business process complexity metrics," *Information Technology Journal*, Vol. 9, No. 7, 2010, pp. 1336–1344.

[35] E. Rolón, J. Cardoso, F. García, F. Ruiz, and M. Piattini, "Analysis and validation of control-flow complexity measures with BPMN process models," *Enterprise, Business-Process and Information Systems Modeling*, 2009, pp. 58–70.

[36] V. Gruhn and R. Laue, "Complexity metrics for business process models," *9th international conference on business information systems (BIS 2006)*, Vol. 85, 2006a, pp. 1–12.

[37] J. Cardoso, J. Mendling, and H. A. Reijers, "A discourse on complexity of process models," *Proceedings of the 2006 international conference on Business Process Management Workshops (BPM'06)*, 2006, pp. 117–128.

[38] L. Sánchez-González, F. G. Rubio, F. R. González, and M. P. Velthuis, "Measurement in business processes: a systematic review," *Business Process Management Journal*, Vol. 16, No. 1, 2010, pp. 114–134.

[39] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol. 28, No. 1, 1979, pp. 100–108.

[40] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Springer, 2012.