Softw Syst Model (2017) 16:759–788
DOI 10.1007/s10270-015-0482-0

REGULAR PAPER

A case study about the improvement of business process models driven by indicators

Laura Sánchez-González[1] · Félix García[2] · Francisco Ruiz[2] · Mario Piattini[2]

Received: 5 February 2015 / Revised: 20 April 2015 / Accepted: 17 June 2015 / Published online: 23 July 2015

Abstract Organizations are increasingly concerned about business process model improvement in their efforts to guarantee improved operational efficiency. Quality assurance of business process models should be addressed in the most objective manner, e.g., through the application of measures, but the assessment of measurement results is not a straightforward task and it requires the identification of relevant indicators and threshold values, which are able to distinguish different levels of process model quality. Furthermore, indicators must support the improvements of the models by using suitable guidelines. In this paper, we present a case study to evaluate the BPMIMA framework for BP model improvement. This framework is composed of empirically validated measures related to quality characteristics of the models, a set of indicators with validated thresholds associated with modeling guidelines and a prototype supporting tool. The obtained data suggest that the redesign by applying guidelines driven by the indicator results was successful, as the understandability and modifiability of the models were improved. In addition, the changes in the models according to guidelines were perceived as acceptable by the practitioners who participated in the case study.

Keywords Business process improvement · Measurement · Indicators · Redesign guidelines

Communicated by Prof. Antonio Vallecillo.

✉ Félix García
Felix.Garcia@uclm.es

Laura Sánchez-González
laurasnchezglez@gmail.com

Francisco Ruiz
Francisco.RuizG@uclm.es

Mario Piattini
Mario.Piattini@uclm.es

[1] INSA, Av. Bellisens, Reus, Tarragona, Spain

[2] Instituto de Tecnologías y Sistemas de Información, University of Castilla La Mancha, Paseo de la Universidad 4, 13071 Ciudad Real, Spain

1 Introduction

Business process (BP) modeling is used principally for organizational redesign and for system development [1]. It is well known that a key activity in the BP lifecycle is the design [2,3], and it consists mainly of BP modeling, blueprints of organizational issues, which are called process models or conceptual models come about as a result. These can be used to make decisions about where, how and why changes to the processes should be enacted to guarantee improved operational efficiency [4].

Improvement of process models constitutes the main focus of the research presented in this paper. In this context, it is also important to consider that BP models are significant artifacts in the BP lifecycle, and sometimes a considerable number of casual modelers, who have diverse background, are involved in the creation of them. This being so, maintenance and quality assurance of these process models are a critical challenge [5], since a poor-quality conceptual model may increase the development effort or result in a process or a system which does not satisfy users (as a consequence of not detecting or not correcting defects) [6].

The quality of process models has been a widely tackled topic in related literature during recent years. There is, however, a lack of consensus in the research community about quality characteristics of business process models [7], which means that we are not sure about the factors that contribute to building models with an acceptable level of quality. A recent systematic literature review about the modeling quality field presented in [8] confirms this limitation and promotes the need to gain in maturity in the field.

On the other hand, there has been a significant effort in evaluating some factors which affect quality characteristics of process models, such as understandability and modifiability. The main trend has been the evaluation of BP model quality by means of suitable measures to discover to what extent the model satisfies a specific quality attribute. The tendency has also been to state empirical connections between internal measures about model complexity and external measures such as error probability or understandability, by means of correlations or regression models [9]. But while these models can be useful and provide a first insight about the quality of the model, they do not directly help in actual decision making, given that design decisions typically require a "yes" versus "no" assessment as to whether a certain change should be made. Taking measurements, and the results of these measures (e.g., number of activities = 25), are not in themselves enough to make design decisions; they do not enable us to give a qualitative assessment (e.g., complexity level = 'high'). The evaluation of measurement results therefore requires the determination of critical values, which indicate when the quality begins to decline. These limit values are called "thresholds," and they can be treated as alarms that give alerts whenever there is low quality. The measurement terminology which will be followed through this paper will be based on the SMO (software measurement ontology) [10], according to which thresholds are typically used for the definition of *decision criteria*. Indicators are the union of *measures* and *analysis models* (which are composed of *decision criteria*). Let us consider, for example, an indicator about correctness of the model: *indicator of structuredness* (CIStruc), which is based on the *structuredness* measure, and whose analysis model is:

- If $Struct. \leq 0.79$, then *CIStruc* = "it is likely that the model does not have errors"
- If $Struct. > 0.79$, then *CIStruc* = "it is likely that the model has errors"

This indicator denotes that the model is considered to be an *error model* when structuredness is higher than 0.79. Finally, measurement results allow designers to detect the need to carry out improvement on the models; at this stage, it is fundamental to have modeling guidelines available to help to improve such models. This means that some redesign initiatives should be applied on models for quality improvement purposes (e.g., if the complexity level is high, then 'activities should be grouped in subprocesses').

In summary, improvement of process models requires: firstly, a good characterization of the quality characteristic to be focused on; secondly, its evaluation by means of useful indicators; and finally, the realization of improvement actions by applying guidelines according to the results of the indicators. These three key pillars have, however, been treated in isolation in related research.

Bearing in mind the ideas mentioned above, our main contribution has been the development of BPMIMA (model improvement based on measurement activities), which attempts to put these three pillars together by describing a group of characteristics for the quality assessment of business process models, indicators which include decision criteria to evaluate such quality characteristics, as well as guidelines associated with indicator results. We focus on three quality characteristics which have higher maturity in the research field and which can be the most relevant in practice: understandability [11–17], modifiability [12,15,18] and correctness [11,12,14,19]. In addition, to validate the potential effectiveness of the BPMIMA framework in practice, it has been applied in a representative case study from the health sector; that study will be the main focus of this paper. Therefore, the principal purpose of this work is to illustrate by means of a case study how to guide the improvement of process models based on measurement results.

The remainder of the paper proceeds as follows. In Sect. 2, we analyze the related work. The background to the proposal is provided in Sect. 3, which presents an overview of the BPMIMA framework. The case study in a hospital, in which the practical utility of the application of modeling guidelines associated with indicators are tested, is presented in Sect. 4. Discussion of the results is included in Sect. 5. Finally, in Sect. 6, the implications and conclusions of this research are presented and future work is outlined.

2 Related work

In this section, we synthesize the related work about quality of BP models. First of all, an overview of quality reference models is provided, to then focus on one of the most relevant and widely researched quality characteristics: understandability of BP models, along with its empirical body of knowledge. Then, research about process modeling guidelines is summarized.

2.1 Process model quality

Due to the lack of consensus in the research community, there is no reference document which describes how to determine the level of quality of business process models. There is some related work which deserves some attention. Mohagheghi et al. [12] state six quality characteristics (correctness, completeness, consistency, comprehensibility, confinement and changeability), extracted from forty studies. Their aim is to evaluate quality of process models for software development.

Other authors, such as Guceglioglu and Demirors [13] or Yudong et al. [20], describe a group of quality characteristics based on international standards (ISO 9126 [21] and ISO 25010 [22]). They relate them to a group of measures, in order to evaluate empirically the level of quality of conceptual models. In a previous work [7], we proposed a quality model for BP models, composed of a set of quality characteristics and related measures. The proposed quality characteristics were selected and/or adapted from a systematic literature review and an exhaustive analysis of international standards (for example, ISO 25010 [22]). The quality model thus includes these characteristics and sub-characteristics: usability (understandability, learnability, user interface esthetics); maintainability (modularity, modifiability); adaptability; correctness; and completeness.

2.2 Process model understandability and its empirical validation

In [9], there is a description of a systematic literature review (SLR) of measures for business processes which covers modeling and execution stages. The measures obtained in the SLR were related to quality characteristics. One of the most significant conclusions of this SLR is that understandability is among the most widely researched quality characteristics in the community. In accord with the main topic of this paper, this subsection will focus on this quality characteristic, with emphasis on empirical support.

In addition to the results provided by the SLR about understandability measurement, some related work deserves some attention. Figl et al. [23] focus on graphical elements, as they consider how routing symbol designs affect an individual's ability to comprehend process models. The authors found that notational characteristics, such as perceptual discriminability and pop out, are significantly associated with perceived cognitive load and model comprehension accuracy, but not with comprehension efficiency. Understandability has also been analyzed from other perspectives, as is the case in Recker et al. [24] who, based on the ideas of Sweller et al. [25], classify understanding of BP models as "surface understanding," which is the understanding of the elements of a business domain, "deep understanding," which is the understanding of the actual and possible relationships among elements in a business domain, and "effort of understanding," which consists in the resource investment required to understand the domain. In Recker's work, an experiment is done to check whether two specific factors can influence the understanding of a business which developers obtain from a process model. These are the content presentation form chosen to articulate the business domain, and the user characteristics of the developers working with the model. The results show that process modelers with training in a given process modeling grammar perform reasonably well in understanding process models depicted with another unfamiliar grammar. In the research of Reijers et al. [26], an experiment is presented that tests the hypothetical benefits of highlighting decision operators for understandability. Their findings about their modeling tool demonstrate the feasibility of automatic highlighting by providing both formalization and an implementation within that tool; their research noted an improvement in the quality of process models.

Mendling et al. [27] use theories of semiotics and cognitive load to theorize how model and personal factors influence process model understandability. They studied factors characterizing process models in terms of activity labels and modeling expertise. They go on to report on a four-part series of experiments, in which they examined these factors. Their results show that additional semantic information impedes syntax comprehension and that theoretical knowledge eases syntax comprehension. Moreover, complexity of activity labels also affects process understandability, as does the level of expertise of modeling analysts.

Since there are diverse pieces of research about process model understandability, Houy et al. [28] collected the most important work related to process understandability, along with related empirical work. They classified work into subcharacteristics of understandability, depending on the measurement mechanism: objective measurement (effectiveness and efficiency of understandability) and subjective measurement (effectiveness seen from the point of view of perceived understandability). They concluded that there is no consensus about how to measure understandability.

2.3 Guidelines of BP modeling

Guidelines for BP redesign have usually been captured by consulting experts in the field (good practices, patterns, etc.). These guidelines can help from different perspectives, as for instance in improving the interface esthetics. In this research stream, there is a variety of proposals, such as [29,30] which propose guidelines related to color, texture, shape of elements, layout, among other main aspects. In this paper, we focus on the guidelines for BP modeling, specifically on the structural perspective. There is some relevant work available in this line, such as "the Guidelines of Modeling" [19], which was conceived as a way of serving final users who may have different objectives, different modeling techniques and tools. The proposal reveals six general techniques to fit the models to different perspectives and objectives of the users. Other guidelines focus on operational aspects of the models, such as the proposal by Sharp and McDermott [31]. A more pragmatic set of guidelines called 7PMG (seven process modeling guidelines) is proposed in [32]. This proposal synthesizes the best knowledge about guidelines in the field; it defines simple rules; some empirical validation is also provided. Taking into consideration these key aspects, our approach considers

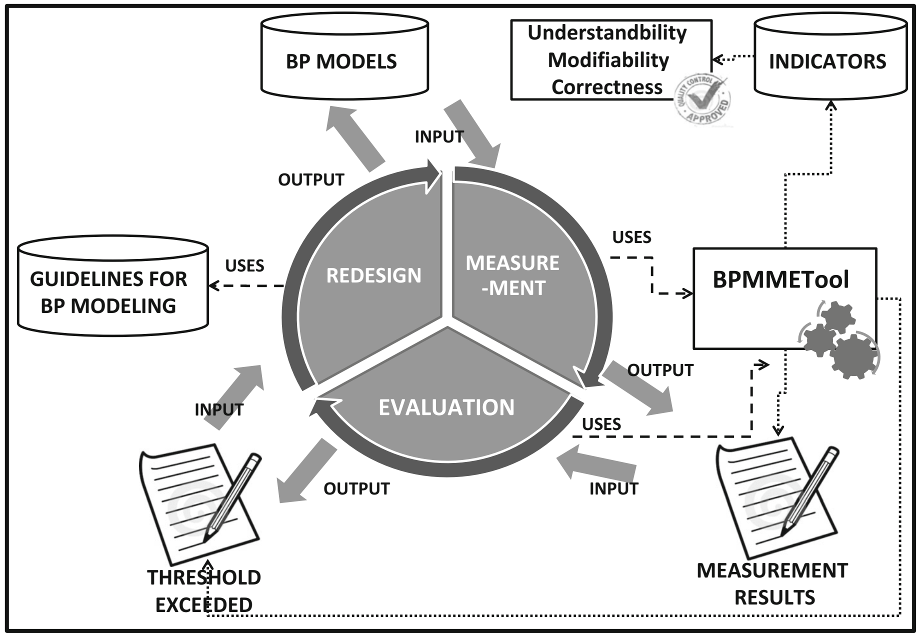


Fig. 1 Elements of BPMIMA framework

7PMG as a set of initial guidelines and attempts to contribute in this field by linking threshold values to these guidelines of modeling (see Sect. 3), aiming to drive the improvement of models based on measurement results. This pragmatic relationship between indicator values and guidelines could be established through experimentation.

3 Background: BPMIMA framework

The key idea of BPMIMA is to support a suitable application of measurement activities during the early stage of the lifecycle, the *design and analysis* stage, in order to obtain the feedback needed to create more high-quality BP models. Basically, this can be achieved from measurement results, by promoting the building of models with lower values of structural complexity. The activities proposed for BP model improvement are *measurement*, *evaluation* of measurement results and model*redesign*. The pragmatic idea underlying these activities is to discover unsafe design, as well as hazardous or unexpected structures, and to correct all of these. In Fig. 1, the main elements of BPMIMA may be observed.

The activities that make up the BPMIMA process are the following:

(a) *Measurement* This implies the application of measures (base, derived or indicators) to BP models. These measures typically check the quality level of a BP model from a structural point of view. A report with the measurement results is then generated. All the tasks related to measurement are performed by the responsible for measurement in the organization.

(b) *Evaluation* The evaluation of measurement results involves providing an objective assessment of these. Absolute numerical results offer only information that is useful for comparison between models, rather than delivering an independent interpretation. That makes it necessary to consider the thresholds or limit values, in order to indicate for what specific value the quality of the measure begins to decline. Extracting thresholds is no small task, and several techniques have been presented for this purpose [33–35].

(c) *Redesign* This focuses on modifying some parts of the model, in the quest to improve its general quality. Those parts which are candidates for alteration can be identified through the use of measures and their corresponding indicators. For example, after measuring a model which is candidate to be redesigned, we obtain no-suitable measurement values; for example, according to the control-flow complexity (CFC) measure [36], we know by means of an indicator that "if CFC is higher than 44, then the model is likely to be difficult to understand." This indicates that some redesign actions in the model could improve its understandability, for instance, by using suitable guidelines. The redesign activity thus implies the use of the threshold measures exceeded, along with the guidelines for BP modeling, in order to guide such redesign.

To support the former activities, the BPMMETool was developed, whose main functionalities are automatic model measurement (modeled with BPMN [37]), definition of new measures and indicators, varied representation of measurement results and advice support for BP model redesign (to see Sect. 3.3).

3.1 Evaluation of BP models quality

Evaluation of measurement results is a key goal of the BPMIMA framework, and it can be difficult to deal with. In the case study presented in this paper, we evaluate BP models described in BPMN [38], which is one of the most widely used BP model notations. The definition of indicators to support the evaluation of BPMN models was based on the extraction and validation of thresholds, as follows:

- *Thresholds and indicators for correctness* Correctness can be defined as the degree to which a model does not have workflow errors or faults, such as deadlocks [39]. We analyzed this quality characteristic previously, in [34]. In the case of that study, thresholds were determined by the application of ROC curves [40] for a specific group of structural measures. The end result was that one limit value or threshold was obtained per measure; this indicates when the conceptual model is *likely to have errors* (behavioral errors). The indicators for correctness can be seen in http://alarcos.esi.uclm.es/bpmima/indicators.htm.
- *Thresholds and indicators for understandability and modifiability* Understandability can be defined as attributes of models that have a bearing on the users' effort to recognize the logical concept and its applicability, while modifiability is seen as the degree to which a model can be effectively and efficiently modified without introducing defects or degrading existing product quality [21]. We analyzed these characteristics [34] previously too, as a result of which some thresholds were obtained by using the Bender method [41]. In contrast to the ROC curves, which obtained one threshold value, the Bender method obtained four limit values which divide the measure domain in five different levels. We used these thresholds for indicator definition. The information of these indicators is organized similarly to that of correctness indicators. The list of indicators is included in http://alarcos.esi.uclm.es/bpmima/indicators.htm.

3.2 Indicators associated with Guidelines for BP modeling

The derived thresholds have the potential to define modeling guidelines in a more precise manner; their values can therefore guide modelers as to what kind of structures they have to avoid. For example, if the indicator IUCFC is based on the CFC measure, and the CFC measure value is 30, the model is considered as *very difficult to understand*. In such a case, the modeler is encouraged to reduce the number of gateway splits, especially the OR-splits. Therefore, the indicators must be applied to drive the application of redesign guidelines. In particular, we associated six out of the 7PMG guidelines [32] with indicators, as follows:

G1. Size reduction guideline

- Recommendation 1. Modularize the model through the use of subprocesses.
- Recommendation 2. Eliminate obvious activities, or merge activities that have a low level of granularity.
- Recommendation 3. Relocate activities from the main process to the subprocess, or vice versa.

This guideline is about reducing the size of the model (for example, reducing the number of nodes). The measures involved in this guideline are the number of nodes [39], *TNA* (total number of activities) [42], *diameter* [39] and *TNSF* (total number of sequence flows) [42]. When some of these measures exceed thresholds, some redesign initiatives should be applied, for example modularizing. When a model has a huge number of nodes, it is in our interest to group some of them and to create a subprocess. This solution was introduced first in [32]. Moreover, some novice modelers design models with a very low level of granularity and a number of activities are very simple or obvious, so it can therefore be eliminated without any loss of information. Finally, some subprocesses have various activities in common, which means that they have to be relocated in the main process.

G2. Events reduction guideline

- Recommendation 1. Include only one start event and one end event per participant.

This guideline is about reducing the number of events in the model. The measures related to this guideline are *TNE* (total number of events) [42] and *NSFE* (number of sequence flows from events) [42]. The number of events (start, intermediate and end events) in a model directly affects the sum of flows from events. The proposed solutions for non-adequate measurement results are based on [32], as regards attempting to use just one start and end event per participant.

G3. Participants reduction guideline

- Recommendation 1. Eliminate the participants that are represented as a black box when they do not include relevant information.

This guideline is about eliminating the participants of the model when we consider they do not include important information. The measures related to this guideline are *NP* (number of pools) [42] and *CLP* (connectivity level among pools) [42], because both concern the element of participant in the model. A possible solution to this problem is to eliminate the participants that are represented as black boxes in the model. The specification of this kind of participants sometimes implies redundant information. For example, when an activity sends a message to a black-box participant, the information about who received the message can be specified in the activity itself, rather than by defining a black box.

G4. Gateway output/input complexity reduction

– Recommendation 1. Divide a gateway with a high number of outputs into various nested gateways whenever possible.

This guideline is about reducing the number of input/output arcs from decision nodes. The measures related to this guideline are *AGD* (average gateway degree) [39], *MGD* (mean gateway degree) [39] and *CNC* (connectivity coefficient) [39], which are related to the number of input/output of nodes, especially decision nodes. The solution is to separate a decision node with a high number of input/outputs into various decision nodes, in order to analyze a question in different steps. This can make the analysis tasks easier to perform, but it may also increase the number of decision nodes, thus leading to a higher CFC measure. The use of this guideline is restricted to the value of the related measures.

G5. Gateway complexity reduction

– Recommendation 1. Merge various gateways when the decisions specified in the gateways are related. Avoid an OR-split whenever possible.

The measures related to this guideline are *TNG* (Total no of gateways) [42], *CFC* and *GH* (gateway heterogeneity) [39]. All of the measures are related to decision nodes; for example, an increase in TNG may increase CFC and GH. The solution is based on the idea of reducing the number of decision nodes. The problem is that it is not possible to eliminate a decision in a model without any loss of information, so it is better to merge some of them when they are related. In that way, various questions are evaluated at the same time. It is also important to reduce, as far as possible, the number of OR-split nodes, because they increase the general complexity of the model. This guideline is not compatible with the G4 one, because they imply contrary operations. The modeler can therefore prioritize one particular guide or another, depending on the model's needs. For example, if the model has a node with a high number of outputs and the modeler considers that the quality level of the model is at risk, G4 guideline will be applied, instead of G5.

G6. Gateway mismatch reduction

– Recommendation 1. Use design patterns to avoid mismatch.

This guideline emphasizes the importance of structured modeling, as regards using gateways. Patterns concern making a join gateway accompany a split gateway [43]. This guideline assists mainly in synchronization tasks.

These guidelines (G1–G6) should be used when indicator values are *moderately understandable/modifiable* or *likely to have errors*. Threshold values differ, depending on the quality characteristic (understandability, modifiability or correctness), so the use of guidelines for improving the understandability (UND), modifiability (MOD) or correctness (COR) of models is described in Table 1. For instance, in order to establish the quantitative support for the application of guideline G1 to improve understandability (first row), we can consider several measures, such as Nodes, TNA, TNSF and diameter. If we focus on the "Nodes" and "Total Number of Activities (TNA)" measures, according to their thresholds (http://alarcos.esi.uclm.es/bpmima/indicators.htm), we know that when the value of "Nodes" exceeds 58.1 or the value of "TNA" exceeds 31.3, the model is difficult to understand. This recommendation is therefore included to support the application of G1. The recommendation is completed with the threshold values of the rest of the related measures (TNSF and diameter) which make the model difficult to understand when they are exceeded. The same criteria are applied to build the remaining explanations for each guideline (see Table 1).

3.3 BPMMETool

BPMMETool (Fig. 2) has been developed in order to support the definition and calculation of proposed indicators and to guide modelers as to which parts of a BP model should be redesigned to avoid endangering its understandability, modifiability of correctness. This prototype tool is able to apply the group of significant measures to BPMN models and to compare these results with threshold values in order to detect undesirable measure results.

When a group of measures with bad values is detected, the tool gives some advice on how to resolve this issue. This is supported by the definition of modeling tips and their application to measured models. As can be observed in Fig. 3, with respect to the modeling tips defined for the CFC indicator (Fig. 3, left), the tool, among other suggestions it offers, gives advice on the need to minimize the number of OR-splits, according to the particular measurement results of a model (Fig. 3, right).

Table 1 Guidelines for BP modeling

Characteristic	Measures	Explanation	Recommendation
UND	Nodes, TNA, TNSF and diameter	Do not use more than 58 nodes in general—no more than 31 activities. The longest path between a start node and an end node should not be higher than 16 nodes. Do not use more than 50 sequence flows	G1
COR	Nodes	Do not use more than 31 nodes	
UND	TNE, NSFE	Do not use more than 11 events and no more than nine sequence flows from an event	G2
UND	NP, CLP	Do not use more than four participants, and CLP should not be higher than 3.79	G3
UND	AGD, MGD, CNC	Do not use more than four input/output sequence flows from each gateway and two per node, up to a maximum value of 7	G4
MOD	AGD, MGD	Do not use more than five input/output sequence flows from each gateway, with a maximum value of 6	
COR	CNC, TS, AGD, MGD	No more than three input/outputs per connector or node, up to a maximum value of 4. To be specific, a seven output flow is acceptable for AND and OR joins	
UND	CFC, TNG, GH	Do not use more than 18 gateways, with a heterogeneity of no more than 0.71; the CFC should not be higher than 37	G5
MOD	TNG, CFC, GH	Do not use more than 16 gateways, with a heterogeneity of no more than 0.81; the CFC should not be higher than 31	
COR	GH	GH should not be higher than 0.4	
UND	GM	The GM should not be higher than 23	G6
MOD	GM	The GM should not be higher than 42	
COR	GM	GM should not be higher than 4.5	

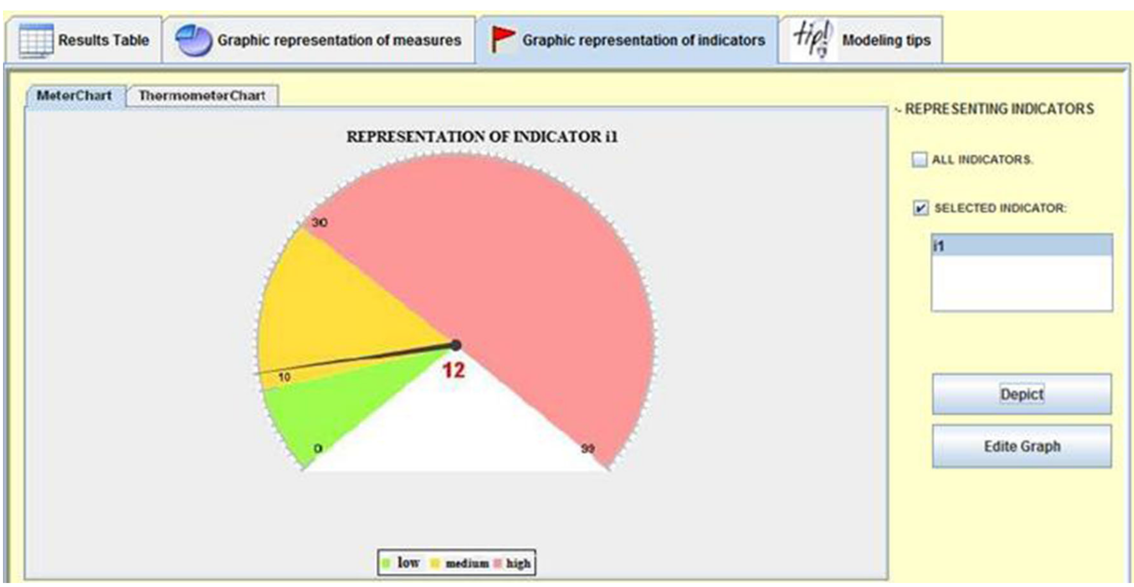


Fig. 2 BPMMETool

BPMMETool model improvement support can be enhanced by users by including new indicators (indicator and thresholds definition module) and by including new modeling guidelines and tips. Some demos which illustrate the usage of BPMMETool can be found in: http://alarcos.esi.uclm.es/bpmima/bpmmetool.htm.

4 Case study

In this section, we present a case study carried out with the purpose of checking the effectiveness of the BPMIMA framework in an organization. The protocol to conduct the case study complied with Runeson and Höst's guidelines [44],

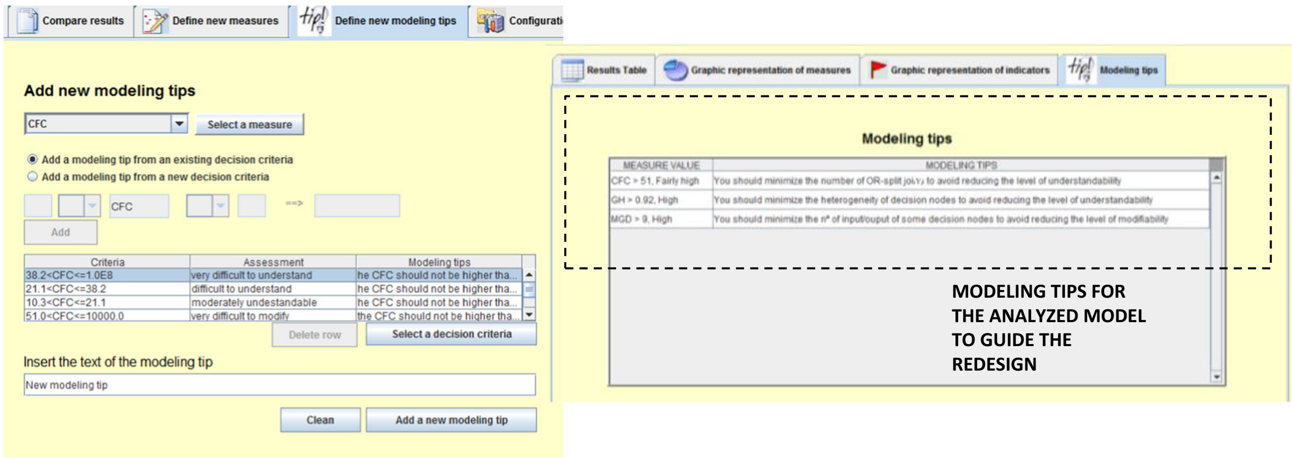


Fig. 3 Definition and application of BPMMETool "Modeling Tips"

which are compiled from different methodology handbooks in several research domains. These are mainly from social science and information systems, such as the proposals by Yin [45], Robson [46] and Benbasat et al. [47].

4.1 Case study design

The research conducted was a holistic single-case study following the classification by [45].

The general research objective was stated as follows: *To discover whether the application of guidelines for BP modeling driven by indicators is effective in improving the general quality of the BP model.*

4.1.1 Context

The organization was the "General Hospital of Ciudad Real" (HGCR), in which some BPs had previously been modeled. These models were used as candidates for improvement, according to the measurement results and guidelines developed in BPMIMA. The HGCR is part of the public health service of the region of Castilla La Mancha (Spain). Patients of this hospital are from 42 surrounding municipalities, which have a population of over 300.000 inhabitants between them, and the hospital has a staff of 2600 employees. Furthermore, this hospital is an official center of reference in Information Technology Development as applied to the health sector, and various national and international awards have been received in this respect. The hospital embraces continuous and ongoing education, with teaching and research as engines for evolution and improvement.

Prior to this work, a mixed work group consisting of professionals in health and expert in IT field modeled some BPs using BPMN [38] because the hospital had previously concluded that some processes were in need of improvement. To facilitate the modeling tasks, health experts were trained in modeling BPs using BPMN notation.

4.1.2 Case study phases and research questions

Along with a previous phase which served as a starting point (preparation, work group composition, selection of processes), this case study was carried out in two main stages: "1. Measurement and Application of Guidelines" and "2. Assessment of Improvement by Health and IT Experts." The stated research questions for each phase were the following:

Phase 1:

– *RQ1* Are BP modeling guidelines effective in improving internal quality of BP models?

Phase 2:

– *RQ2* Are BP modeling guidelines effective in improving the external quality of BP models?
– *RQ3* Do the subjects' personal opinions about the changes proposed by the BP modeling guidelines depend on the type of expertise (experts in the domain or in modeling)?

By means of RQ1 and RQ2, it is intended to assess whether the application of the guidelines driven by indicator results is effective in improving both internal and external quality of BP models. In addition, RQ3 was considered, because the term 'quality' has a relative meaning. Quality of a product, service or artifact can be expressed as "the totality of features and characteristics that bear on its ability to satisfy stated or implied needs"; these needs can be different for each kind of stakeholder. In this research, the focus is on the quality of BP models expressed in BPMN notation; its primary goal is "to provide a notation that is readily understandable by all business users, from the business analysts that create the initial drafts of the processes, to the technical developers responsible for implementing the technology that will perform those processes, and finally, to the business people who will manage and monitor those processes" [37]. The case study provided a good context in which to evaluate RQ3, as we can distinguish clearly two types of people: on the one hand, the health staff, who are the domain experts (business analysts), as well as being the process owners; on the other the IT staff, (in this case, they mainly play the role of modeling experts). Health staff may see a BP model through different eyes (due to their background, experience and knowledge) and from a different point of view (concerns and issues) than IT staff.

4.2 Case study execution

4.2.1 Phase 0: preparation

This phase was carried out prior to this study as part of related research to model processes of the General Hospital of Ciudad Real with BPMN, as well as to validate a set of measures of BPMN models (details can be found in [38]). The work group was composed of 11 professionals, 7 from the health sector, including doctors and nurses, along with quality and management staff, and 4 researchers from the University of Castilla-La Mancha. Three processes were modeled: one with low complexity, another medium-sized one and finally, a large BP, all of which supposedly had improvement needs. Two of these processes constitute the units of analysis of this case study, named "*Incorporation of a new employee"* and "*Appointments"*. All the health experts also had knowledge of those processes. The selection took place in several different meetings, including a brain-storming session. Health experts were trained in the modeling notation selected, i.e., BPMN. The creation of the models was collaborative and iterative. To achieve this goal, there were necessary work meetings with the responsible of the process for the refinement of the model to be carried out. Interviews also took place to obtain information from some of the participants in the processes. After gathering relevant information about these processes, the work group modeled the first version of them. The summarized description of the selected processes is as follows:

- *"Incorporation of a new Employee" (INE)* This was considered to be a low-complexity key process in the whole organization. It includes the plan for training and adaptation, as well as provision of relevant information for all those involved in the hospital, in an attempt to make their integration into the new job smoother. This is a purely administrative process, which is, however, related to a huge amount of users with different roles: doctors, pharmacists, nurses, psychologists and administrative technical staff, among others. The process is represented in Fig. 4 and described in Table 2.
- *"Appointments" (APP)* This process collects the tasks related to patients' appointments with doctors. When a patient asks the hospital about an appointment with a specific doctor, the hospital should check whether the patient can be dealt with in that hospital or not. Other tasks are then required so that an appointment for a given patient with a particular doctor can finally be arranged. There are different ways in which to deal with appointments in the hospital. In this case, health specialists decided to model two different processes in relation to hospital appointments. The processes specifically related to hospital appointments are *first consultation appointment* (FCA), represented in Fig. 5, and *appointment for examination with tests* (AEWT), represented in Fig. 6 and described in Table 3.

4.2.2 First Phase: measurement and application of guidelines

The first phase of the case study concerned improving three BP models, (INE, FCA and AEWT), by using the guidelines. This means improving the internal quality, through the measurement, evaluation and redesign of BP models. For the BP model measurement, we used BPMMETool, so as to detect values that exceed the thresholds. Moreover, this tool assigned a *modeling recommendation* to each of the values exceeded; this was extracted from the guidelines for BP modeling defined in this work. These *recommendations* helped us to redesign the model. After redesign, the *new* version was measured again, in order to check whether the measurement results had improved. The criteria to trigger the "alarms" coming from an indicator value were as follows: When an indicator result changed (for example, from very difficult to understand to moderately understandable), we considered that the model had been significantly improved. Sometimes the measurement result changed, although the indicator result did not (for example, the number of nodes measure, which is used to calculate the UIN indicator, decreases from 50 to 48, but the model is still considered to be moderately understandable). This case is considered to be a small improvement. If the measurement results indicate that the model has been improved, we consider that the guidelines are useful in improving the internal quality of BP models. The steps followed in this stage for each model were as follows:

Incorporation of a new Employee (INE)

- *Assessment of the initial model* (Fig. 4), whose results are shown in Table 4.

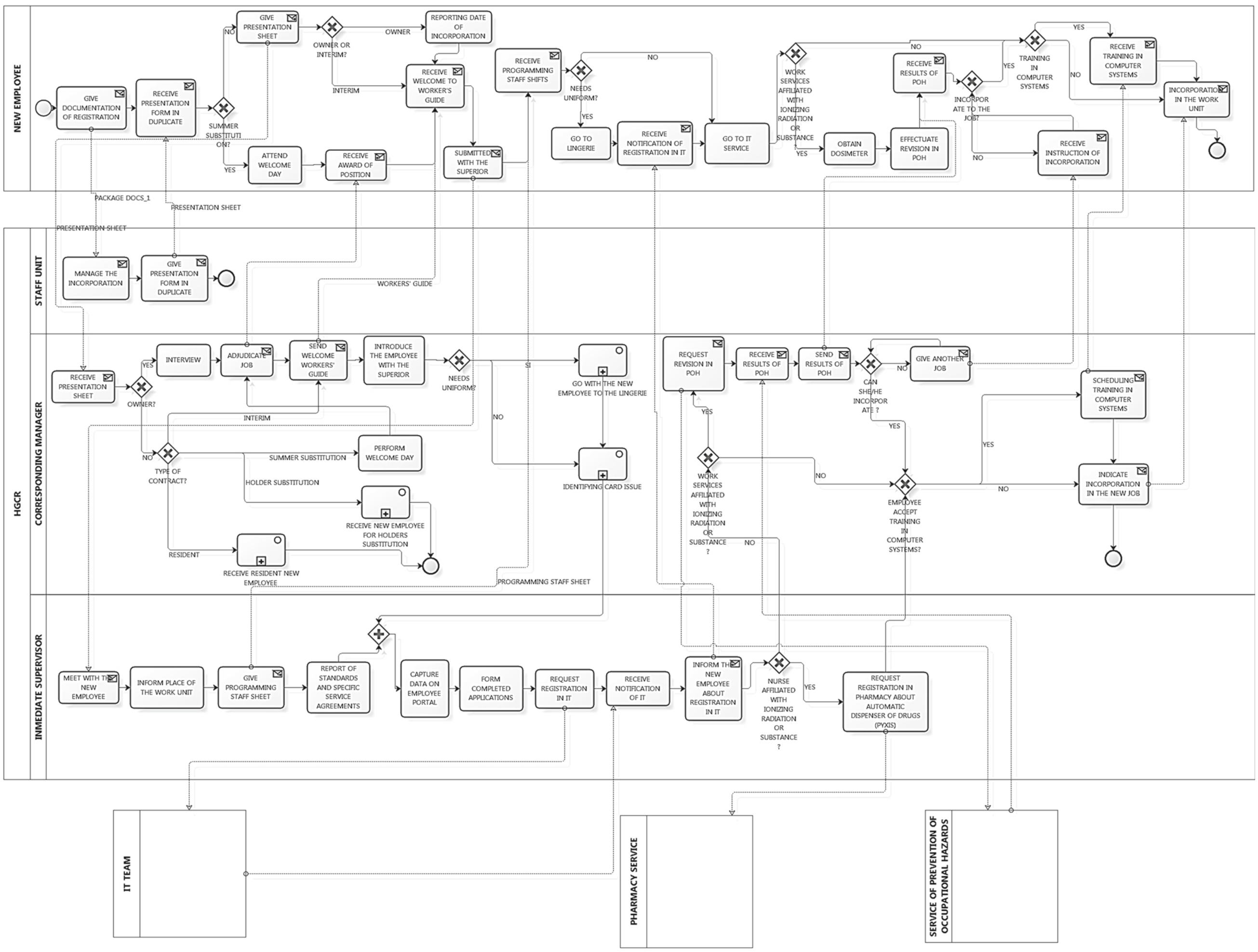


Fig. 4 Original version of incorporation of a new employee process (INE)

Table 2 Information about the INE process

Goal	To promote the organization of the process. This includes a plan for training and information, as well as for adaptation of the people involved, so they fit in with the hospital requirements, thus making their integration into the new job easier
Limits	This process starts when the professional comes to the hospital and finishes when he/she is incorporated into the new job
Customers	New professionals
Responsible	Those responsible for nursing, medical aspects and management
Participants	New professionals in the hospital, human resources, computer services, lingerie, pharmacy, prevention services, nursing and management service
Suppliers	Human resources, provisions, maintenance, training and information systems

- *Redesign* Based on the "alarms" triggered by the indicators evaluation detected in the previous step, this model was redesigned by applying the following guidelines.

Some of the changes applied, which are highlighted in the improved model were as follows:

- We reduced the number of participants (G3), eliminating the *black box* ones, because we did not believe that they had contributed any relevant information to the model semantics. This was the case of the IT team, pharmacy service and prevention of occupational hazards.
- We created a new subprocess (G1) which included activities related to IT management because we believed that these activities are very specific to a health field (they would be relevant if the process were related to IT). This change reduced some measures, which were mainly nodes, TNA and diameter. This new subprocess is shown in Fig. 7.
- Some activities which are considered to be very basic were resumed in only one activity (G1). This is the case of "effectuate revision in POH" and "receive results of POH", which are summed up in "manage revision in POH".

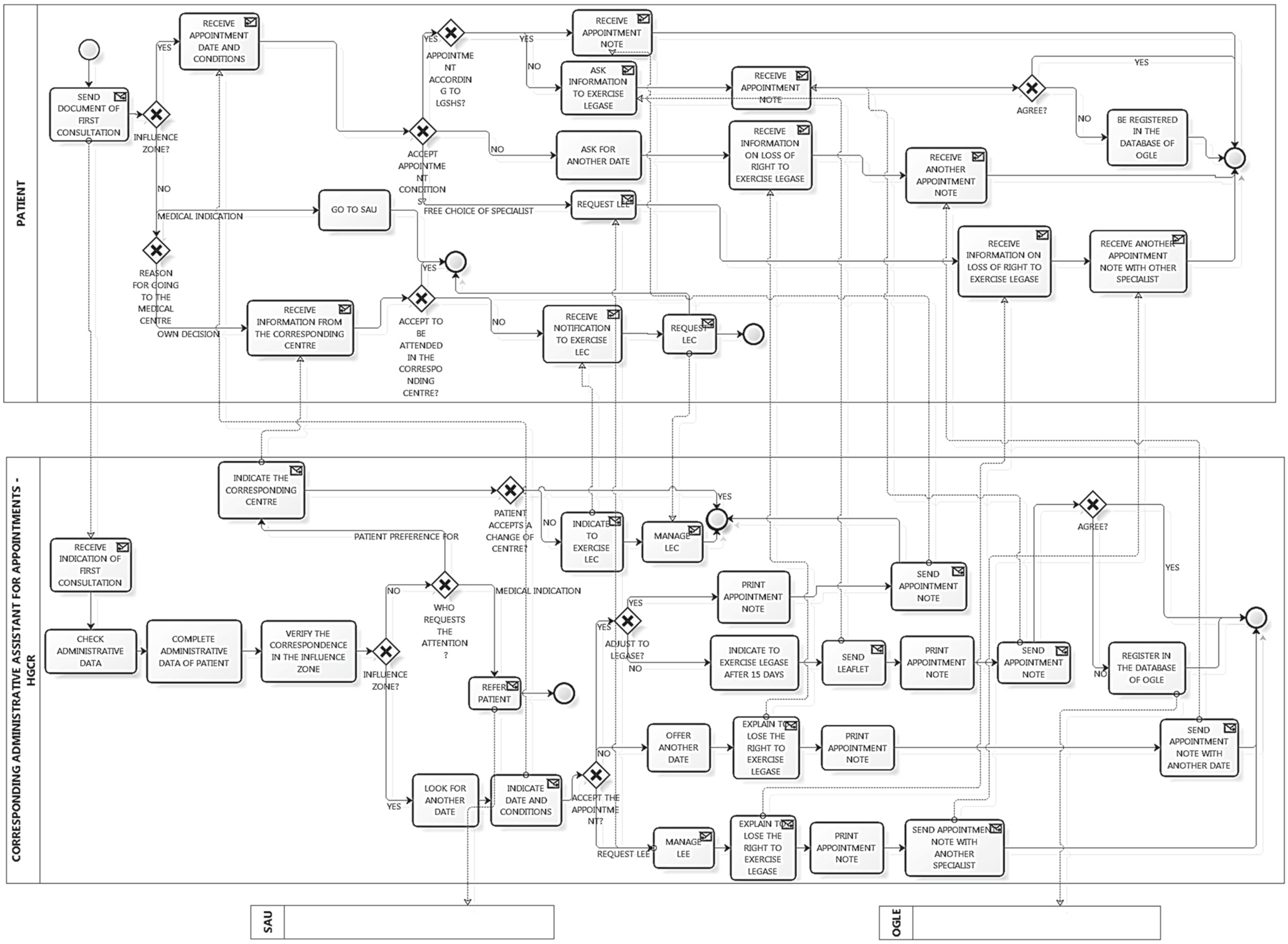


Fig. 5 Original version of first consultation appointment process (FCA)

- We unified (G1) two activities into a single one which includes both aspects. The activities in question were related to informing the employee about some kinds of aspects related to their incorporation ("Meet with the new employee" implies "Meet with the new employee" and "inform place of work unit").
- Some merge gateways (G5) were added according to the design patterns (all splits should be accompanied by a join). Some gateways (*owner?* and *type of contract?*) were unified in one gateway, which asks two questions at the same time, such as *type of contract?*

- *Assessment of redesigned model,* The redesigned model was assessed in order to check whether the guidelines applied improved the indicator values (see analysis subsection) (Fig. 8).

Appointment Processes (FCA and AEWT)

- *Assessment of initial process model* The assessment results of the process models FCA and AEWT are set out in Tables 5 and 6.

- *Redesign* According to the main "alarms" detected, it was noticed that both BP models share a group of tasks and that modularization strategies included in guidelines would be useful. That might concern creating subprocesses from a general process, or creating a general process from subprocesses. Some changes were thus made to both models in conjunction (FCA and AEWT):

  - First of all, those activities which are related to the detection of the particular center where the patient is going to be treated are in both models, and they should be moved to a *parent* model (G1). This change implies unifying some activity labels. For example, the activity "send a referral note of first consultation" in FCA, and the activity "send the referral note" in AEWT can be unified to "send a referral note," which is located in the *parent* process. Something similar to this occurs with the "verify the patient's address" activity, which is renamed as "verify and complete administrative information about the patient." Another group of activities which both models share is related to send-

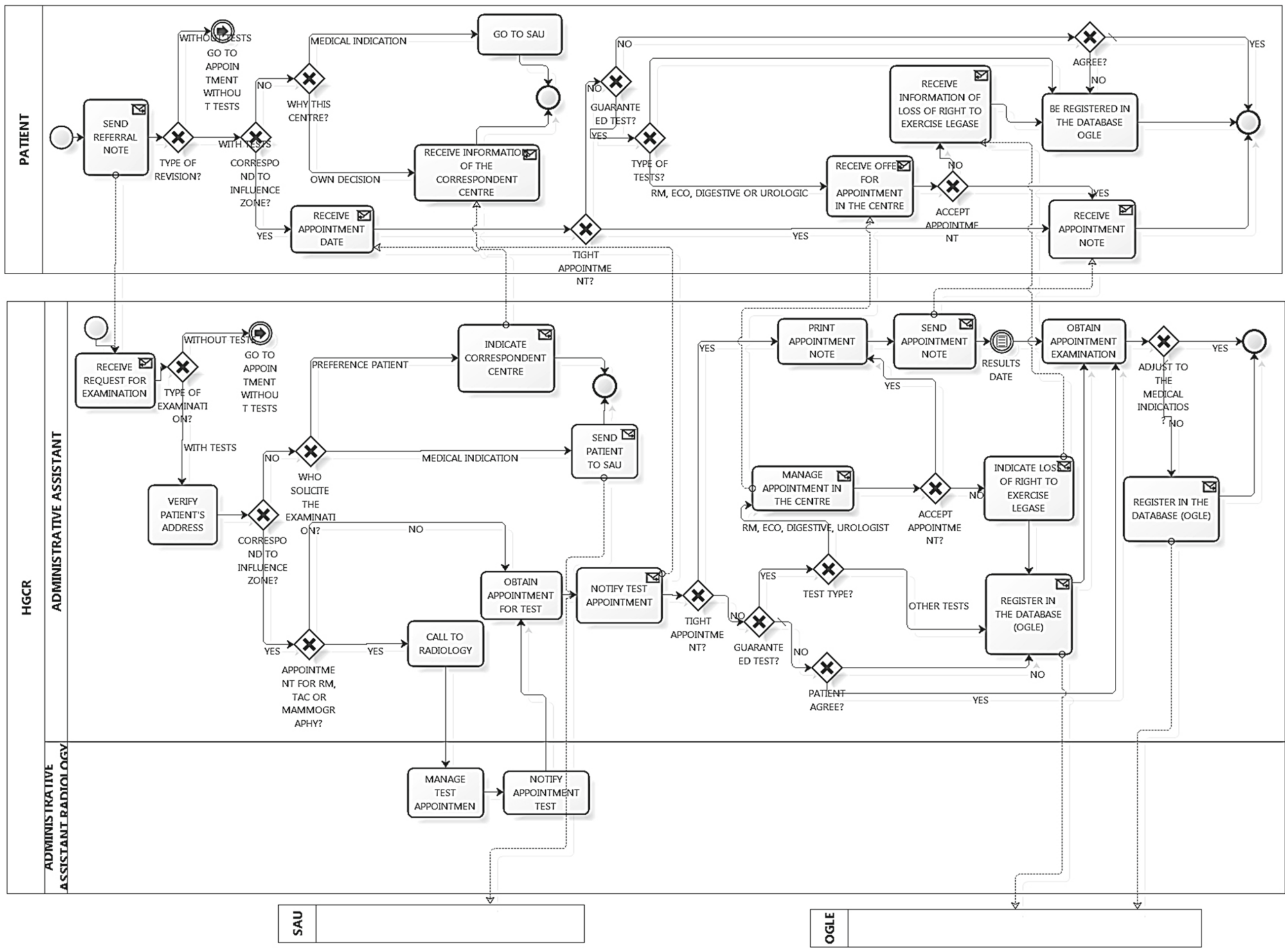


Fig. 6 Original version of appointment for examination with test process (AEWT)

ing the appointment note, as well as to including the patient request in the OGLE database. This change appears in Fig. 9, in which it can be observed that we extract activities shared by both models and we move them to a *parent* model. The specific activities in both models are therefore included as subprocesses.

– Other changes were also made to both models, such as the following: We merged some gateways (G5). For example, the gateway "influence zone?" and "reasons for going to the medical centre" were merged to become only one: "influence zone and reason for attending?"; we used patterns in order to avoid mismatch (G6).

Some changes were specifically for the FCA model. In this case, we eliminated some simple activities (G1), which we considered obvious. This was the case of "print referral note." We also eliminated some repeated activities, specifically "explain that patient loses the right to exercise LEGASE."

Figure 9 illustrates the modularization applied to create the process model "Appointments (APP)," which includes

Table 3 Information about the appointments processes (FCA and AEWT)

Goals	To offer appointments for consultation or/and medical test
Limits	The input of the process is the indication of a new appointment, and the output is the specific date/place of the consultation
Customers	Patients within the health area of the General Hospital of Ciudad Real
Responsible	Outpatient admission service, dependent on the general management
Participants	Administrative staff in the admission service, primary attention, hospitals, medical management, nursery management, person responsible for the service, user attention service (SAU) and waiting list management office (OGLE)
Suppliers	Information systems, informatics, maintenance, cleaning, admission, lingerie, prevention of occupational hazards (POH)

the shared tasks. The specific tasks related to each model (FCA and AEWT) were included in that different model as subprocesses (Figs. 10, 11,12, 13, 14). The complexity of both models was thereby considerably reduced (see analysis subsection).

4.2.3 Second phase: assessment of improvements by experts

The second phase consisted in checking the process experts' personal satisfaction with the improved BP models. This was achieved by asking them whether the new version of the BP model was better, from different perspectives.

The group of *subjects* was composed of health experts and IT experts:

1. The health expert subgroup was composed of four health professionals from the work group involved in phase 0 of this study. They had the same background in modeling, as they all had received the same previous training in BPMN. Since they had been working with that modeling notation for about 4 years (we started to work on this topic in 2008 with the publication [42]), we considered that they have a good level of knowledge of BPMN.
2. The IT expert subgroup was composed of seven Master Degree students in Computer Science who were also doing their PhD about process research topics. They all had similar knowledge of BPMN and were very familiar with other software modeling notations such as UML. Since all of them had the same level of qualification (PhD students), we considered that they had the same background.

The case study *material* was composed of a questionnaire for each improved model, whose aim was to evaluate the quality characteristics. The questionnaire focused on understandability and modifiability (to see Appendix 1). Although correctness was also used to determine the threshold, we considered that behavioral errors are not easy to detect by users and that this task requires a high expertise in modeling.

The analysis of results obtained in each phase of the case study is described in the following section.

Table 4 Assessment of the process of incorporation of a new employee (initial)

Measure result	Indicator result	Modeling recommendation
Nodes = 61	UIN = 4*; CIN = 1	G1
TNA = 46	UITNA = 4	
Separab. = 0.66	UISEP = 3; MISEP = 3	No recommendation
Diameter = 26	UIDIAM = 4	
Sequent. = 0.4	UISEQ = 4	
NSFG = 49	UINSFG = 5; INSFG = 5	
NEDDB = 13	UINEDDB = 4	
TNSF = 76	UITNSF = 4	
NMF = 17	UINMF = 4	
CNC = 1.52	UICNC = 4; CICNC = 1	G4
AGD = 3.5	UIAGD = 3; MIAGD = 3; CIAGD = 1	
MGD = 5	UIMGD = 3; IMGD = 3; CIMGD = 1	
GM = 19	UIGM = 3; MIGM = 3; CIGM = 1	G6
CFC = 20	UICFC = 3; MICFC = 3	G5
TNG = 14	UITNG = 4; MITNG = 4	
NSFE = 6	UINSFE = 3	G2
NP = 5	UINP = 4	G3
CLP = 2.83	UICLP = 3	

1 very easy to understand/modify/likely to have errors, 2 easy to understand/modify, 3 moderately understandable/modifiable, 4 difficult to understand/modify and 5 very difficult to understand/modify
* 0 not likely to have errors

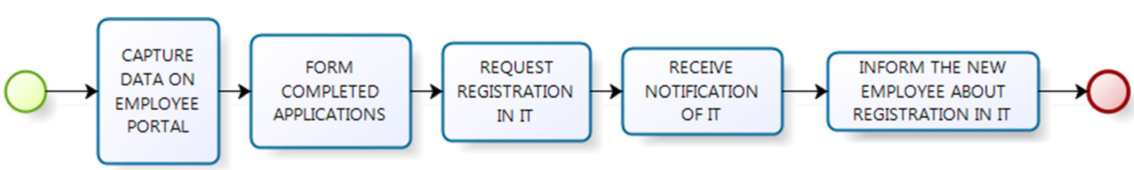


Fig. 7 Subprocess "IT management" in the INE process

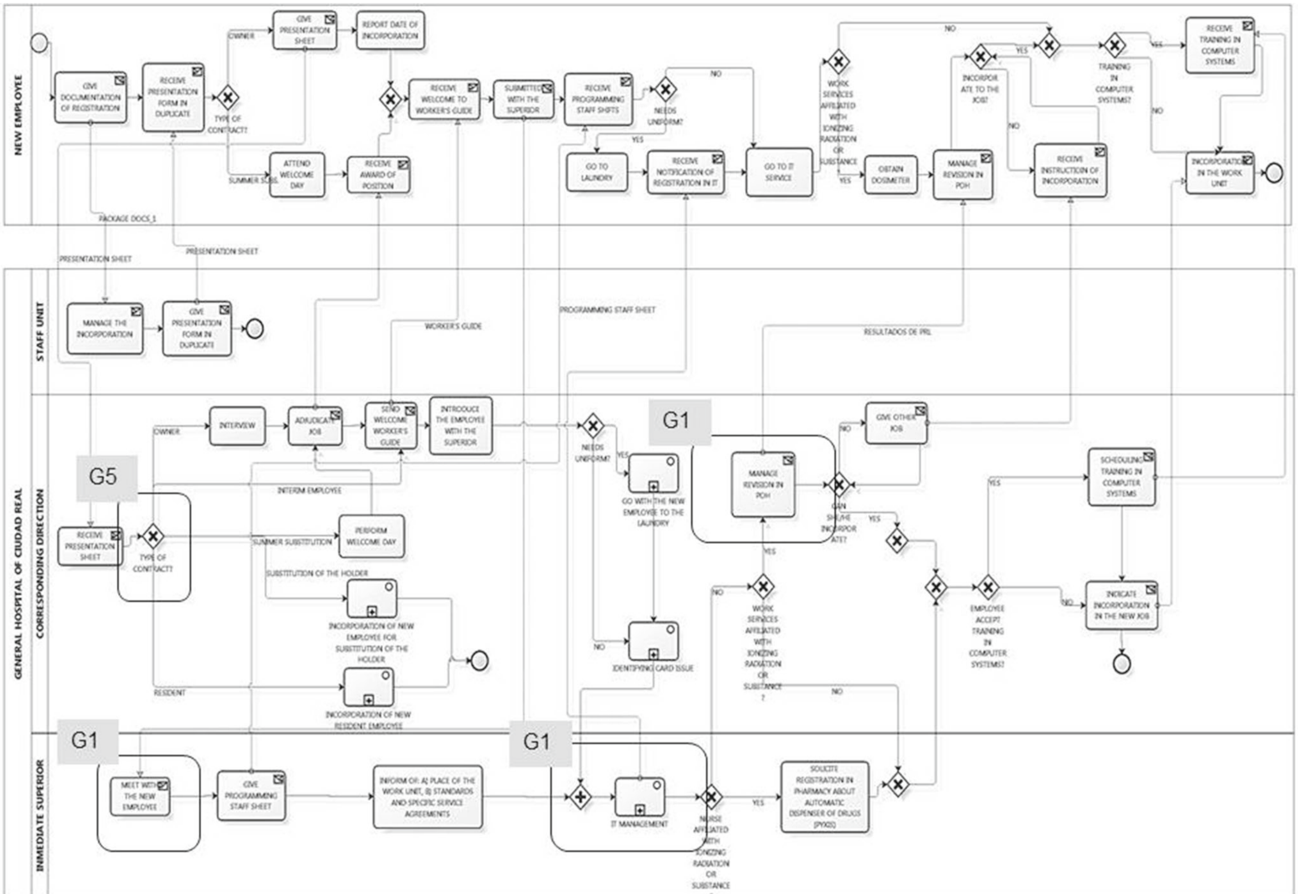


Fig. 8 Improved version of incorporation of a new employee (INE)

Table 5 Assessment of first consultation appointment (initial)

Measure result	Indicator result	Modeling recommendation
Nodes = 60	UIN = 4, CIN = 1	G1
Diameter = 16	UIDiam = 4*	
TNSF = 68	UITNSF = 4	
TNA = 41	UITNA = 4	
Separab. = 0.56	UISep = 3, MISep = 3	No recommendation
NEDDB = 12	UINEDDB = 5	
NSFG = 38	UISNFG = 4, MINSFG = 5	
Sequent. = 0.26	UISeq = 4	
NMF = 15	UINMF = 4	
NSFE = 17	UINSFE = 5	G2
CFC = 26	UICFC = 4, MICFC = 3	G5
TNG = 12	UITNG = 4, MITNG = 4	
GM = 26	UIGM = 4, MIGM = 4, CIGM = 1	G6
NP = 4	UINP = 3	G3
CLP = 3	UICLP = 3	
AGD = 3.16	UIAGD = 3, MIAGD = 3, CIAGD = 1	G4
MGD = 4	UIMGD = 3	

1 very easy to understand/modify/likely to have errors, 2 easy to understand/modify, 3 moderately understandable/modifiable, 4 difficult to understand/modify and 5 very difficult to understand/modify
* 0 not likely to have errors

Table 6 Assessment of appointment for examination with tests (initial)

Measures Result	Indicator result	Modeling recommendation
Nodes = 51	UIN = 3*, CIN = 1	G1
Diameter = 23	UIDiam = 4	
TNSF = 63	UITNSF = 4	
TNA = 24	UITNA = 3	
Separab. = 0.48	UISep = 3, MISep = 3	No recommendation
NEDDB = 18	UINEDDB = 5	
NSFG = 54	UINSFG = 5, MINSFG = 5	
Sequent. = 0.45	UISeq = 4	
NMF = 10	UINMF = 3	
GM = 36	UIGM = 4, MIGM = 4, CIGM = 1	G6
NP = 4	UINP = 3	G3
CNC = 1.43	UICNC = 4, MICNC = 1	G4
CFC = 36	UICFC = 4, MICFC = 4	G5
TNG = 18	UITNG = 5, MITNG = 5	
NSFE = 15	UINSFE = 4	G2
TNE = 9	UITNE = 3	

1 very easy to understand/modify/likely to have errors, 2 easy to understand/modify, 3 moderately understandable/modifiable, 4 difficult to understand/modify and 5 very difficult to understand/modify
* 0 not likely to have errors

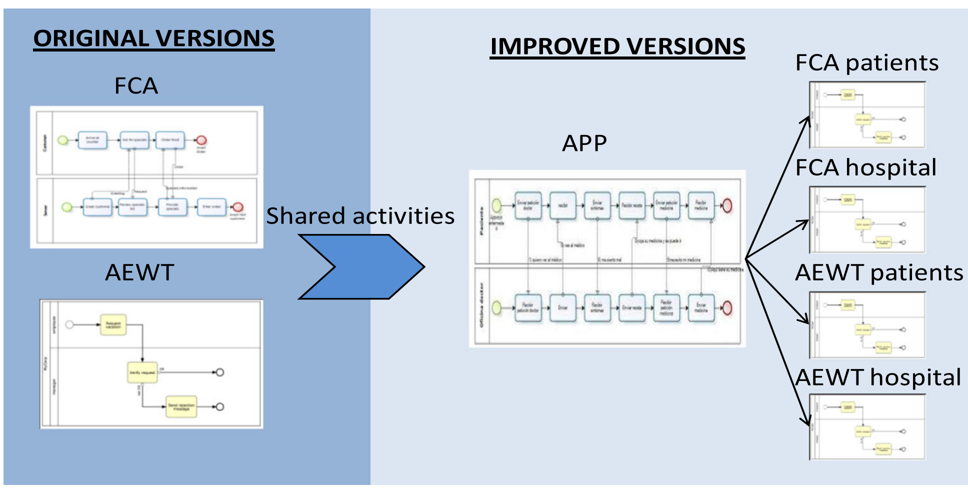


Fig. 9 Modularizing first consultation appointments and appointments for examination with tests

4.3 Analysis of results

This subsection analyzes the data collected to provide an answer to the stated research questions, as follows:
1st phase:

RQ1: Are BP modeling guidelines effective in improving internal quality of BP models?

To address this research question, the improved versions of BP models were analyzed through the application of indicators by using the BPMMETool.

With regard to *INE process*, the measurement results of its improved version (Fig. 8) are shown in Table 7, where the most important improvements in comparison with the original model are highlighted in Bold. The criteria to “trigger the alarms,” which have been explained in the former subsection, were applied. In addition, when the result of the measure is improved, but not the indicator result (i.e., we change the measure results but not the assessment), then this is highlighted in italic. The other results did not undergo any improvement. In specific terms, 52 % of the measures which obtained non-suitable results were improved by using guidelines. The other 48 % did not undergo any improvement

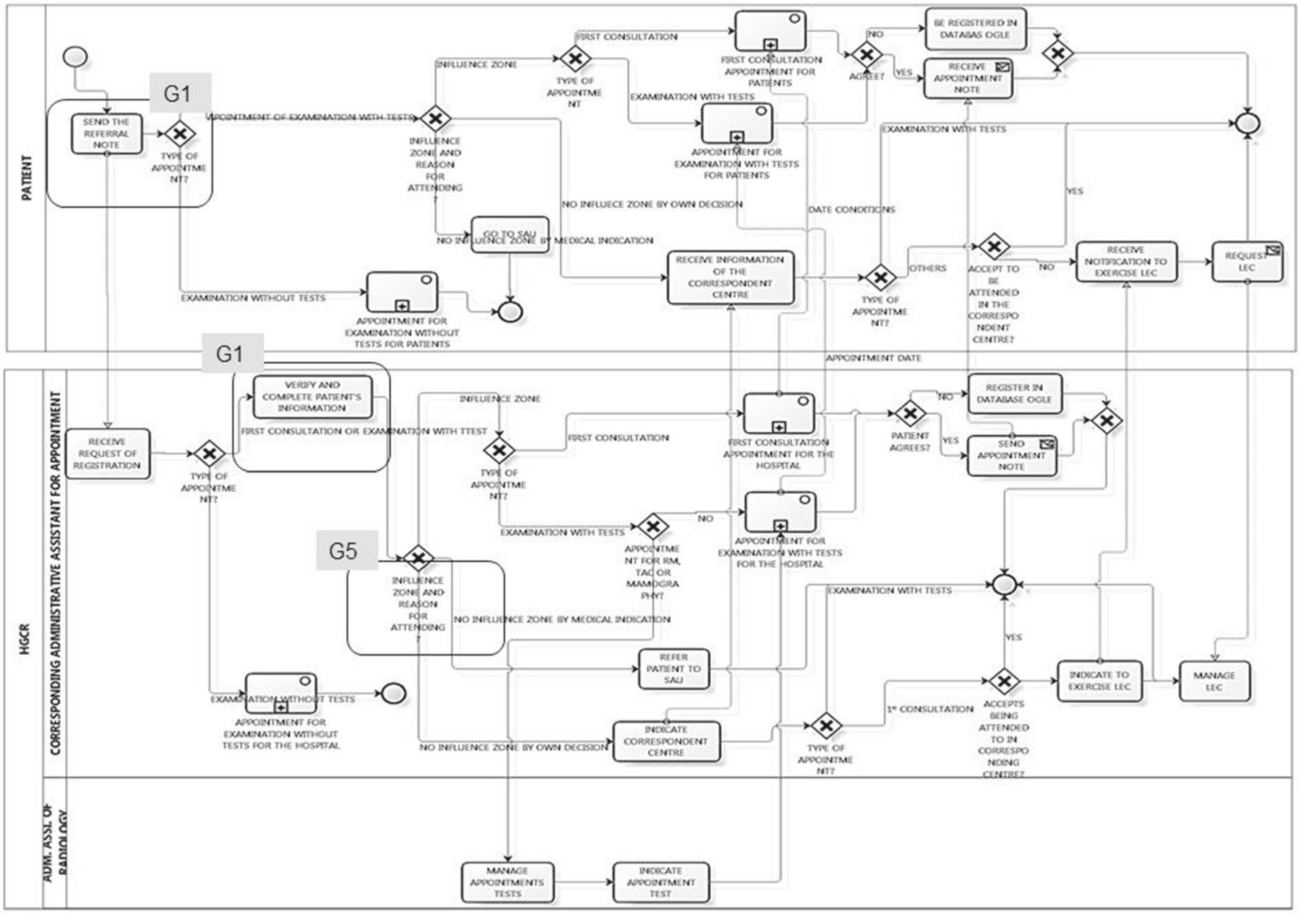


Fig. 10 Appointments process model (obtained from sharing tasks of FCA and AEWT)

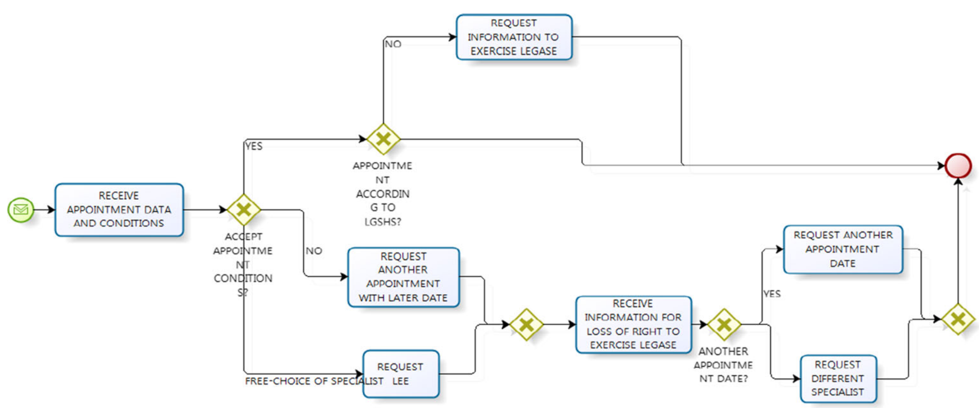


Fig. 11 Subprocess model of first consultation appointments for patients

As far as *FCA* process is concerned, the comparison between the indicator results from the original and improved models is represented in Table 8. To be specific, we compare the indicator results of the original version to the indicator results of the different models obtained by modularization. The results that have improved when the indicator results change the evaluation are shown in bold (for example, NP changes from *moderately understandable* to *easy to understand*). On the other hand, the fact that the improvement is only reflected in the subprocesses (FCA for hospital or FCA for patient) rather than in the main process (the process which contains the tasks shared between FCA and AEWT) is represented in italic. Finally, if the indicator results are worse, they are not highlighted in any color. Table 8 indicates that 72 % of the indicators analyzed obtained very good results or good results in comparison with the original version.

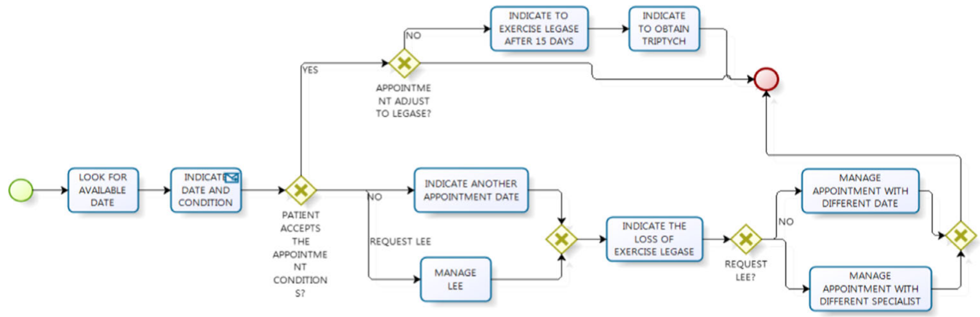


Fig. 12 Subprocess model of first consultation appointments for the hospital

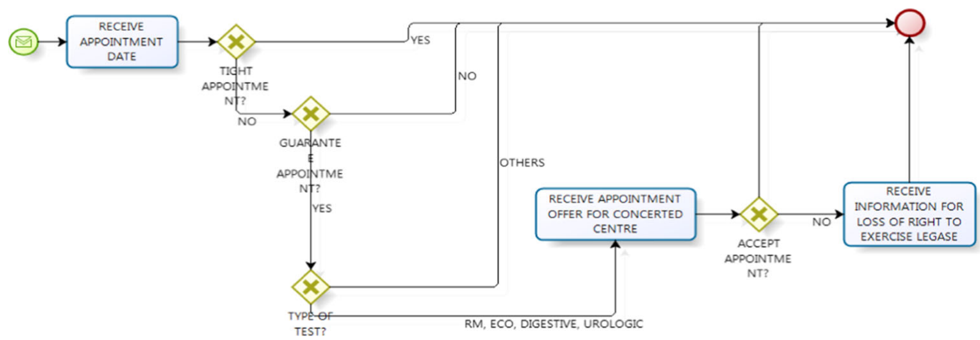


Fig. 13 Subprocess model of appointment for examination with tests for patients

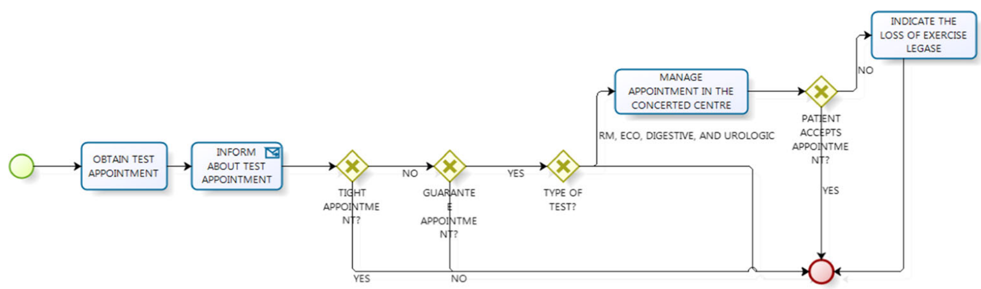


Fig. 14 Subprocess model of appointment for examination with tests for the hospital

A similar comparison was made for the AEWT model. Table 9 shows the indicator results of the original and the improved versions. The same highlighting used for the analysis of FCA is used here. The comparison enables us to observe that in all cases the indicators improve their results. The specific figures are that 43 % of the indicators obtained significant improvements, compared to the 57 % which also obtained good results, but to a lesser extent. In general terms, the modifications made to the AEWT model have improved its internal quality.

As a result of the previous analysis, it has been verified that the BP modeling guidelines are effective in improving internal quality of BP models. In this case, each BP model has been improved: In the first model, INE, 52 % of the measures which obtained non-suitable results were improved, and more specifically, 21 % of them obtained significant improvements. In this case, the indicators which underwent the biggest improvements were related to the size of the model (which includes NP, NMF, diameter and measures of the number of nodes). With regard to the FCA and AEWT models, the main improvements were based on modularizing strategies, which implied creating some process models that were linked by subprocesses. In our case, the general model (APP) included the tasks shared by both models (FCA and AEWT), and the specific functionality of each of them was hidden in the subprocesses. This meant that we changed two high-complexity models for a medium-complexity one and four fairly simple-complexity models. These changes are considered to represent an important reduction in the general complexity of the original versions. Figures also show that modifications made to FCA obtained 72 % of indicator improvements, while all the indicators with non-adequate results were improved in the AEWT. We may conclude that the modularizing of techniques implies more significant

Table 7 Comparison of measurement and indicator results of original and improved version of incorporation of a new employee

Measure	Result in original version	Result in improved version	Indicator	Result in original version	Result in improved version
NP	**5**	**2**	**UINP**	**4**	**2**
NMF	**17**	**11**	**UINMF**	**4**	**3**
Diameter	**26**	**23**	**UIDIAM**	**5**	**4**
Nodes	**61**	**55**	**UIN**	**4**	**3**
TNSF	*76*	*70*	*UITNSF*	*4*	*4*
AGD	*3.5*	*3.23*	*UIAGD*	*3*	*3*
			MIAGD	*3*	*3*
			CIAGD	*1*	*1*
GM	*19*	*15*	*UIGM*	*3*	*3*
			MIGM	*3*	*3*
			CIGM	*1*	*1*
GH	*0.23*	*0.20*	UIGH	2	2
			MIGH	*2*	*2*
TNA	*46*	*38*	*UITNA*	*4*	*4*
CNC	*1.52*	*1.47*	*UICNC*	*4*	*4*
			MICNC	*1*	*1*
CLP	2.83	4	UICLP	3	4
Seq	0.4	0.38	UISeq	4	4
NEDDB	13	16	UINEDDB	4	5
NSFG	49	55	UINSFG	5	5
			MINSFG	5	5
Sep	0.66	0.60	UISep	3	3
			MISep	3	3
MGD	5	6	UIMGD	3	4
			MIMGD	3	3
			CIMGD	1	1
CFC	20	21	UICFC	3	3
			MICFC	3	3
TNG	14	17	UITNG	4	4
			MITNG	4	5
NSFE	6	6	UINSFE	3	3

improvements. These results indicate that several improvements were obtained in all of the cases; the practical utility of guidelines driven by indicator results in improving the internal quality of BP models was thus demonstrated.

2nd phase:

The research questions addressed in this phase were as follows:

RQ2: Are BP modeling guidelines effective in improving the external quality of BP models?
RQ3: Do the subjects' personal opinions about the changes proposed by the BP modeling guidelines depend on the type of expertise (experts in the domain or in modeling)?

With regard to RQ2, we analyzed the experts' opinions about each of the guidelines applied to each BP model. As previously mentioned, the health experts group was composed of 4 hospital staff who participated in the phase 0 and who had previous experience in BPMN (from 2008). Each health expert answered only questions related to the process in which he/she participates; as a result, we collected three questionnaires per model. The IT experts group was composed of seven people with previous knowledge about BPMN and software modeling notations (UML, etc.).

The material used to collect the data in this phase is included in Appendix 1. Each expert was asked to judge whether the application of a specific guideline improved the model from the perspective of understandability and modifiability. The options were "the model is more understandable/modifiable," "the model is less understandable/modifiable"

Table 8 Comparison of indicator results for first consultation appointment

M	Original		Improved					
	FCA		APP		FCA patient		FCA hospital	
	Res	Res.Ind.	Res	Res.Ind.	Res	Res.Ind.	Res	Res.Ind.
NP	**4**	**UINP = 3**	**2**	**UINP = 2**	**1**	**UINP = 1**	1	**UINP = 1**
TNSF	*68*	*UITNSF = 4*	*54*	*UITNSF = 4*	*17*	*UITNSF = 1*	*19*	*UITNSF = 1*
NMF	**15**	**UINMF = 4**	**7**	**UINMF = 2**	**0**	**UINMF = 1**	**0**	**UINMF = 1**
DIAMETER	*16*	*UIDiam = 3*	*13*	*UIDiam = 3*	*10*	*UIDiam = 2*	*11*	*UIDiam = 2*
GM	**26**	**UIGM = 4, MIGM = 4, CIGM = 1**	**20**	**UIGM = 3, MIGM = 3, CIGM = 1**	**3**	**UIGM = 1, MIGM = 1**	**3**	**UIGM = 1, MIGM = 1**
CFC	*26*	*UICFC = 4, MICFC = 3*	*24*	*UICFC = 4, MICFC = 3*	*7*	*UICFC = 1, MICFC = 2*	*7*	*UICFC = 1, MICFC = 2*
NSFE	**17**	**UINSFE = 5**	**13**	**UINSFE = 4**	**4**	**UINSFE = 2**	**4**	**UINSFE = 2**
TNA	**41**	**UITNA = 4**	**23**	**UITNA = 3**	**7**	**UITNA = 1**	**9**	**UITNA = 1**
Nodes	**60**	**UIN = 4, CIN = 1**	**37**	**UIN = 2, CIN = 1**	**14**	**UIN = 1**	**16**	**UIN = 1**
TNE	**7**	**UITNE = 3**	**5**	**UITNE = 2**	**2**	**UITNE = 1**	**2**	**UITNE = 1**
CNC	1.38	UICNC = 3	1.64	UICNC = 4, CICNC = 1	1.21	UICNC = 2, CICNC = 1	1.18	UICNC = 3, CICNC = 1
AGD	3.16	UIAGD = 3, MIAGD = 3, CIAGD = 1	3.26	UIAGD = 3, MIAGD = 3, CIAGD = 1	3.2	UIAGD = 3, MIAGD = 3, CIAGD = 1	3.2	UIAGD = 3, MIAGD = 3, CIAGD = 1
MGD	4	MIMGD = 3	4	MIMGD = 3	4	MIMGD = 3	4	MIMGD = 3
Separ.	0.56	UISep = 3, MISep = 3	0.53	UISep = 3, MISep = 3	0.41	UISep = 3, MISep = 3	0.57	UISep = 3, MISep = 3
Seq.	**0.26**	**UISeq = 4**	**0.30**	**UISeq = 4**	**0.71**	**UISeq = 2**	**0.68**	**UISeq = 3**
NEDDB	**12**	**UINEDDB = 5**	**15**	**UINEDDB = 5**	**5**	**UINEDDB = 2**	**5**	**UINEDDB = 4**
NSFG	38	UINSFG = 4, MINSFG = 4	49	UINSFG = 5, MINSFG = 5	16	UINSFG = 3, MINSFG = 3	16	UINSFG = 3, MINSFG = 3
TNG	**12**	**UITNG = 4**	**15**	**UITNG = 4, MITNG = 4**	**5**	**UITNG = 3, MITNG = 3**	**5**	**UITNG = 3, MITNG = 3**

Table 9 Comparison of indicator results for appointment for examination with tests

M	Original		Improved					
	AEWT		APP		AEWT patient		AEWT hospital	
	Res.	Res.Ind.	Res.	Res.Ind.	Res.	Res.Ind.	Res.	Res.Ind.
Nodes	**51**	**UIN = 3, CIN = 1**	**37**	**UIN = 2, CIN = 1**	**9**	**UIN = 5**	**10**	**UIN = 5**
DIAMETER	**23**	**UIDiam = 4**	**13**	**UIDiam = 3**	**9**	**UIDiam = 2**	**10**	**UIDiam = 2**
TNSF	*63*	*UITNSF = 4*	*54*	*UITNSF = 4*	*12*	*UITNSF = 5*	*13*	*UITNSF = 5*
TNA	*24*	*UITNA = 3*	*23*	*UITNA = 3*	*3*	*UITNA = 5*	*4*	*UITNA = 5*
GM	**36**	**UIGM = 4, MIGM = 4, CIGM = 1**	**20**	**UIGM = 3, MIGM = 3, CIGM = 1**	**8**	**UIGM = 5, MIGM = 2, CIGM = 1**	**8**	**UIGM = 5, MIGM = 2, CIGM = 1**
NP	**4**	**UINP = 3**	**2**	**UINP = 2**	**1**	**UINP = 5**	**1**	**UINP = 5**
CFC	*36*	*UICFC = 4, MICFC = 4*	*24*	*UICFC = 4, MICFC = 3*	*8*	*UICFC = 5, MICFC = 2*	*8*	*UICFC = 5, MICFC = 2*
TNG	**18**	**UITNG = 5, MITNG = 5**	**15**	**UITNG = 4, MITNG = 4**	**4**	**UITNG = 3, MITNG = 5**	**4**	**UITNG = 3, MITNG = 5**
NSFE	*15*	*UINSFE = 4*	*13*	*UINSFE = 4*	*6*	*UINSFE = 3*	*6*	*UINSFE = 3*
TNE	**9**	**UITNE = 3**	**5**	**UITNE = 2**	**2**	**UITNE = 5**	**2**	**UITNE = 5**
CNC	*1.43*	*UICNC = 4, MICNC = 1*	*1.64*	*UICNC = 4, MICNC = 1*	*1.33*	*UICNC = 3, MICNC = 1*	*1.33*	*UICNC = 3, MICNC = 1*
Separab.	*0.48*	*UISep = 3, MISep = 3*	*0.53*	*UISep = 3, MISep = 3*	*1*	*UISep = 2, MISep = 2*	*1*	*UISep = 2, MISep = 2*
NEDDB	*18*	*UINEDDB = 4*	*15*	*UINEDDB = 5*	*4*	*UINEDDB = 3*	*4*	*UINEDDB = 3*
NSFG	*54*	*UINSFG = 5*	*49*	*UINSFG = 5*	*12*	*UINSFG = 3*	*12*	*UINSFG = 3*
		MINSFG = 5		*MINSFG = 5*		*MINSFG = 3*		*MINSFG = 3*
Sequent.	*0.45*	*UISeq = 4*	*0.30*	*UISeq = 4*	*1*	*UISeq = 2*	*1*	*UISeq = 2*
NMF	**10**	**UINMF = 3**	**7**	**UINMF = 2**	**0**	**UINMF = 5**	**0**	**UINMF = 5**

Table 10 Opinions of application of guidelines for understandability/modifiability

Model	Guideline	Understandability/Modifiability	
		IT experts (seven participants) (%)	Health experts (three participants) (%)
Incorporation of a new employee	G1: Eliminate obvious activities	87	67
	G1: Relocate activities from the main process to the subprocess or vice versa	87	33
	G5: Try to merge various gateways when the decisions specified in the gateways are related	43	33
	G6: Use design patterns to avoid mismatch	100	33
	G1: Eliminate or merge activities with a low level of granularity	57	67
	G2: Try to include only one start event and one end event per participant	57	33
	G3: Eliminate the participants represented as a black box when they do not include relevant information	57	67
Appointments and subprocesses: FCA for patients and the hospital and AEWT for patients and the hospital	G1: Relocate activities from the subprocess to the main process	87	100
	G1: Eliminate or merge activities with a low level of granularity	87	100
	G1: Eliminate obvious activities	57	33
	G5: Try to merge various gateways when the decisions specified in the gateways are related.	100	33
	G6: Use design patterns to avoid mismatch	100	67
	G2: Try to include only one start event and one end event per participant	71	67

or "neither of those options." When the subject believed that the application of a guideline improved the model, then 1 was added to the result indicated in the "IT experts" or "health experts" column. For example, "6 out of 7" signifies that 6 of the 7 subjects answered that the application of a specific guideline improved the model.

Table 10 shows the percentages of experts who perceived improvements in the application of each guideline. In the case of the improvement related to the INE process, the health experts considered that the modifications applied did not imply great improvements. This means that in some cases they preferred the original version to the modified one.

However, the changes proposed for FCA and AEWT were more positively evaluated. This is either because the changes proposed for the INE process imply small modifications to the model, or because they are more specific modifications. When a process model is improved without taking some related models into account (which is the case of INE), modifications are less significant. However, in FCA and AEWT, changes are meaningful because they imply relocating all the activities that the different models have in common (linked by subprocess), and the new version is very different than the original one. The number of nodes in FCA and AEWT was reduced substantially because some parts in common were detected and moved to the parent process (APP). Unlike the IT experts, therefore, the health experts appreciate changes when models are significantly modified. We can therefore conclude that guidelines really are effective in improving the external quality of models and, in particular, as regards their understandability and modifiability.

As regards RQ3, we wanted to analyze whether or not the personal opinion about the guidelines depends on the type of expertise. Most of the IT experts preferred some specific guidelines, such as the use of design patterns to avoid mismatch (G6, G3, G4). This change is only a graphical modification, and it does not imply any change in the way in which things are done in the hospital; that is why the structural changes promoted by guidelines are more positively assessed by IT experts than by health experts. With regard to the guidelines concerning the elimination of obvious activities (G1), this was positively assessed by all the experts in the case of the INE model, but the opinion differed for the FCA and AEWT models. Again, the health experts were more reticent about using guidelines to redesign the models. In the case of guideline G5, most of the IT experts agreed on their use, in contrast to the opinion of the health experts. As regards the guideline concerning the elimination or merging of obvious activities (G1), all the experts agreed on using this, and there was no difference in opinion related to their particular expertise. The guidelines related to including only one start event and one end event per participant did not entirely convince all the participants, and neither did the advice concerning the elimination of the participants represented as a black box when they do not include relevant information (G3). Finally, the opinion about relocating activities was considered to be useful by all the experts, especially in the case of relocating the activities shared by some models to another one. However, in the case of creating subprocesses in an effort to include a group of related activities, the health experts did not agree on using this technique for model improvement. In this specific case, they preferred to have all the activities represented in the same process diagram, thus avoiding the need to hide information in subprocesses.

In conclusion, in most of the cases, the two groups of experts (health and IT) disagreed about the usefulness of the guidelines. Consequently, we obtain initial insight into the fact that the personal opinions of the changes proposed by the BP modeling guidelines depend on the type of expertise (experts in the domain or in modeling). This came about because the health experts were reluctant to change their normal way of conducting their business, as opposed to the IT experts, who did not mind changing the structure by eliminating, merging, modularizing, etc. Moreover, health experts tend not to appreciate obtained improvements because they can only see them when the process is under execution (and not in the design stage). IT experts have the specific background to appreciate those improvements, however, just by analyzing the process model. The results related to this research question can therefore be considered as inconclusive and this deserves further investigation in future, with more empirical support.

In summary, the guidelines were more positively assessed by IT experts, given their background in modeling techniques; results also suggest the importance of providing health professionals with a modeling background, in order to promote improvement of the business processes from initial stages.

4.4 Validity threats analysis

In this section, we analyze the different threats to the validity of this case study. In order to describe these threats more clearly, we have divided them into the categories suggested by Runeson and Höst [44]:

- *Construct validity* The research question is whether the use of guidelines driven by indicators for BP modeling helps in quality improvement. Since we consider that quality can be analyzed in two different dimensions (internal and external quality), we have different measures to resolve the research question. We have verified the improvement of internal quality by measuring the structure of the models and contrasting these results with threshold values. The measures and indicators applied have been theoretically validated, and they therefore satisfy the construct validity. As a result, the internal quality improvement is guided by valid measures with their corresponding thresholds. On the other hand, the external quality is a subjective score provided by subjects. External quality is related to the subjects' personal opinion of the models, such as whether the model is easy to understand or to modify. We consider that asking for personal opinions about an improved model is a suitable means to discover whether the external quality of the model has actually improved and it does not constitute a threat to construct validity.
- *Internal validity* We can affirm that the internal quality is improved by using the guidelines and there are no other influencing factors. However, with regard to the evaluation of the external quality of BP models, we identified some possible threats. For example, the subjects' motivation to give their opinions about the guidelines might have been influenced by the time that was available for them to participate in this study. In the case of the health experts, it was noticed that little free time was available, but this was mitigated by the fact that they had participated in previous studies, and they were motivated by this research since the research results might help to improve their business processes. This threat was less pronounced for the IT experts as they had more time available. We consider that there were no fatigue effects because of the average time obtained for the respondents (about 30 minutes). Moreover, only two processes were analyzed. Some learning effects can be considered because sometimes we applied similar guidelines to improve certain BP models. We mitigated this possible threat by giving out the different questionnaires about the BP models in random order to the different subjects.
- *External validity* The material used to analyze the utility of the guidelines is considered valid because it is composed of two BPs extracted from a Hospital. We shall then analyze whether the subjects constituted a threat to external validity. The subjects are experts in their fields (health and IT) so the nature of the subjects does not constitute a threat. However, the number of subjects is very small, especially in the health group, owing to the fact that it is limited by the experts who had participated in modeling of BP models. Although it would have been better to have had more health experts, we consider that these subjects can offer us significant evidence about the practical utility of the guidelines. In order to reduce the limitation related to the number of subjects, this study should be replicated in other hospitals. In addition, the domain of this case study is the health sector, which can limit the possibility of generalizing the results to other domains. We believe that the profile of each type of user (business and tech-

nical) is similar in other domains, although future work must focus on such other domains if the external validity is to be strengthened.

- *Reliability* The application of guidelines is driven by thresholds of measures. These thresholds were obtained by using large experimental data. The main point to discuss here is whether these thresholds are dependent on this research or not. Since the subjects of the experiments used for threshold determination were quasi-graduates or postgraduates, we considered that thresholds are valid for subjects with similar background. This means that the level of, for example, understandability of a BP model might have varied depending on the background of the user, so it is possible that the guidelines for BP modeling could be considered very strict or very weak by different users.

5 Discussion

The results obtained in this case study strongly suggest that the application of indicators could detect non-suitable models from an understandability/modifiability perspective. Moreover, when non-suitable models have been detected by the use of indicators, we link indicators results with a group of advices in order to guide the improvement of models. Indeed, the application of indicators and guidelines on hospital process models indicates strong and favorable perceptions of quality improvements.

The main purpose of these guidelines is to help modelers to create correct models based on empirical foundations. Anyway, these guidelines should be applied in a reasonable way, given that they should be interpreted as advices for model improvement and not as strict rules, because its applications depend on the judgment of the modeler and, also, on the context of the business process. As a matter of fact, two or more experts in modeling can disagree about the usefulness of a particular guideline. This fact can limit the automatic application of the guidelines, for instance, by using supporting software tools. Anyway, in this research, we highlight the importance of providing modelers with indicators which can trigger the application of improvement actions by associating them to suitable guidelines.

We further argue that the results from our empirical investigations have implications for both research and practice on the quest toward guiding process modelers toward the design of high-quality process models. In our specific case of the hospital area, the processes of "Incorporation of a new Employee" and "Appointments" have been redesigned based on the proposed indicators results and guidelines, and quality, understood as level of understandability and modifiability, has been improved.

In summary, some insights have been obtained from business modeling in real contexts, about the need to support the improvement of the quality of models by using quantitative information from which improvements based on good practices and guidelines can be suggested.

6 Conclusions and future work

In this paper, we have presented a case study to evaluate the BPMIMA framework for BP model improvement. The main contribution of the research presented is the validation of the support that indicators can provide to properly apply redesign in the context of the BPMIMA cycle: modeling, evaluation, redesign (guidelines). To accomplish this main goal, a case study was conducted in the General Hospital in Ciudad Real. The results demonstrated that the redesign by applying guidelines driven by the indicator results was successful, as the understandability and modifiability of the models were improved. To test the quantitative results obtained, a questionnaire was answered by health professionals who participated in the modeling of the processes and with IT experts. As a result of this survey, the changes in the models according to guidelines were perceived as acceptable by all experts, with better scores being obtained by IT experts with respect to health professionals. These qualitative results can constitute a starting point to focus future validation of guidelines as perceived by modelers.

In summary, with the BPMIMA framework, it is expected that the work of novice modelers will be made easier, which will make for the building of high-quality models. One key implication derived from this research is that BP modeling improvement implies a suitable integration of quantitative evaluation of the models supported with thresholds. These in turn will "trigger" alarms that point to improvement needs that will be met by redesign guidelines. In addition, the usefulness of the guidelines must be demonstrated by a quantitative perspective (indicators calculation on redesigned models) and by qualitative evaluations (perception or score given by the modelers about the effectiveness of the guidelines). In order to adopt the BPMIMA cycle, therefore, it is fundamental for organizations to build a suitable knowledge base about guidelines with associated indicators which drive improvements. The research presented illustrated this by applying the 7PMG, as our main focus is on BPMN models, but this research can be generalized and enhanced by considering additional guidelines.

As regards the limitations which should be addressed in future work, the following ones deserve special attention: firstly, the guidelines driven by indicators described consider the structural perspective of the models, given that most of the measures considered for building the indicators are structural ones, and it is important to recall that not all the important aspects of the quality of the model are related to the structure. For instance, activity labeling or layout aspects can affect the understandability, modifiability and correctness. This issue is currently being researched by the BP community, and results may be very useful in providing additional guidelines related to other aspects apart from structure. In addition, research about improvement guidelines should advance to provide a more complete empirical body of knowledge, given that doubts might appear about the suitability of applying some guidelines in a given context. For instance, more than one end event (G2 related) can improve understandability in some contexts. It should also be said that the preliminary results about IT and health assessment of guidelines suggest that empirical studies should be conducted to focus on analyzing how guidelines can be perceived by modelers with different profiles.

Acknowledgments The authors would like to thank Jan Mendling for his support in research related to this study. This work has been developed under the projects: GEODAS-BC (Min. de Economía y Competitividad and Fondo Europeo de Desarrollo Regional FEDER, TIN2012-37493-C03-01) and INGENIOSO (PEII-2014-050-P), Junta de Comunidades de Castilla-La Mancha and FEDER.

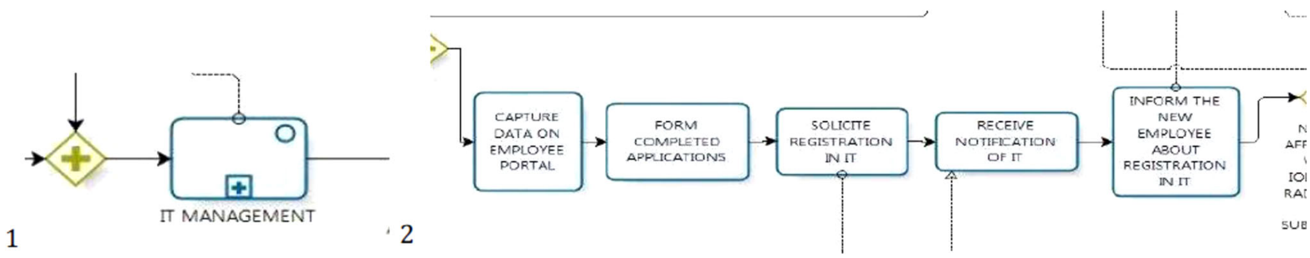
Appendix 1: Questionnaires of the case study

Questionnaire 1: Busines Process model "Incorporation of a new employee"

Let us imagine we want to analyze a BP model to understand and modify it. Indicate what proposed changes most affect the understandability and modifiability of the models. For example, if you think that the subprocess IT management, which is carried out by the immediate superior, has led the model to be more understandable/modifiable, choose this option.

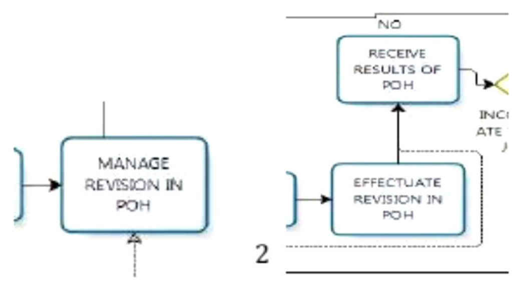
1. Using subprocesses such as "IT management"[1] which is composed of a group of related activities, instead of displaying all the activities[2] …makes the model

   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these



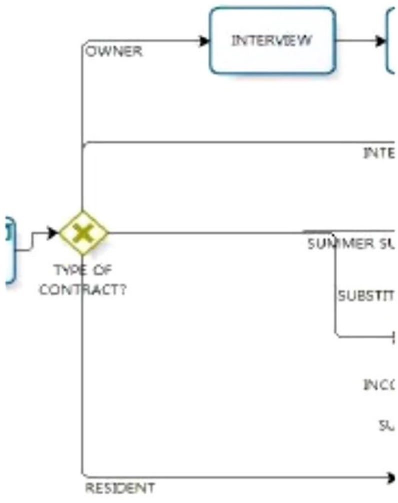
2. Using an activity called "Manage revision in POH"[1] which is composed of the activities "Effectuate revision in POH" and "Receive results of POH"[2] … makes the model

   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these

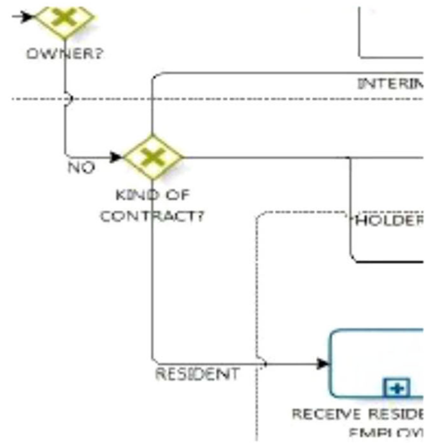


3. Using decision nodes with the highest number of output possibilities, for example, the gateway "type of contract?" in the participant "correspondent direction" has 5 possibilities[1] instead of 4[2] (in the original model)… this makes the model

   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these



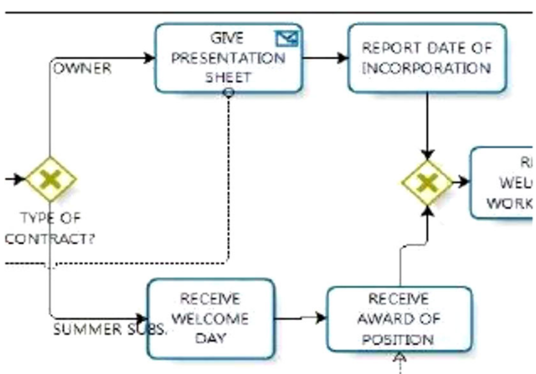
1



2

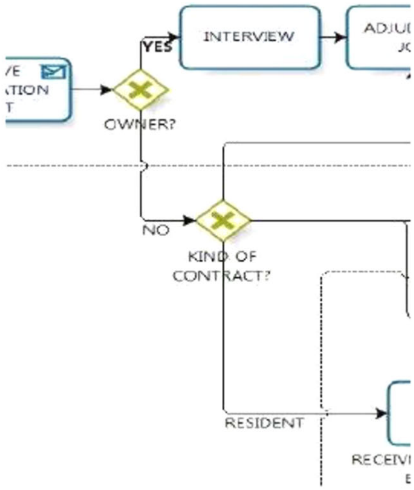
4. Using "join" gateways accompanying "split" ones, for example, the join gateway which accompanied the split "type of contract?" in "new employee" participant… makes the model

   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these

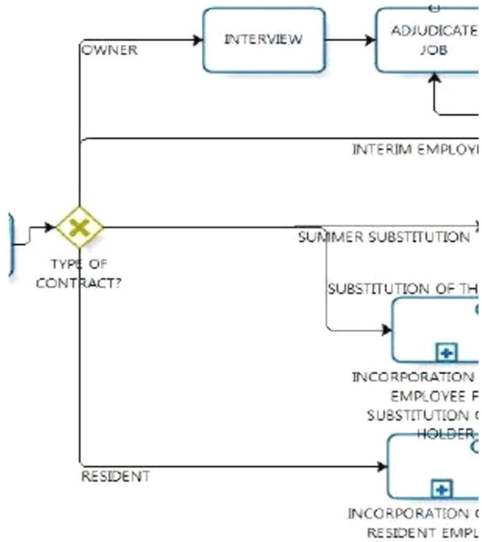


5. Using gateways which include two questions at the same time[1] (fusion of gateways). For example, the gateway "type of contract?"[2] implies taking a decision between owner or summer substitution, etc… this makes the model

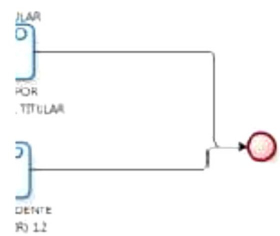
   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these



1

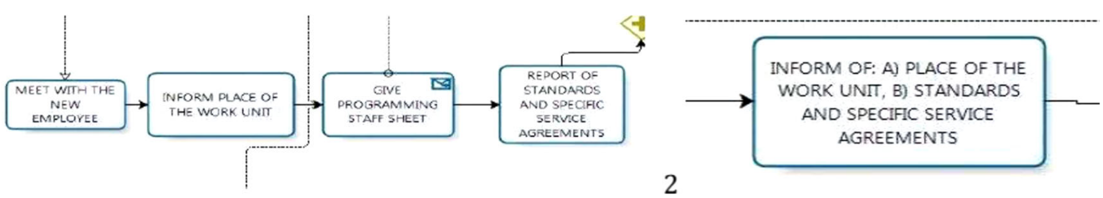


2



6. Joining two activities[1] in only one[2] . That is the case of two activities called “Inform place of the work unit” and “Report standards and specific service agreement”, to make only one which is called “Inform of a) place of the work unit and b) standards and specific service agreement”… this makes the model

   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these



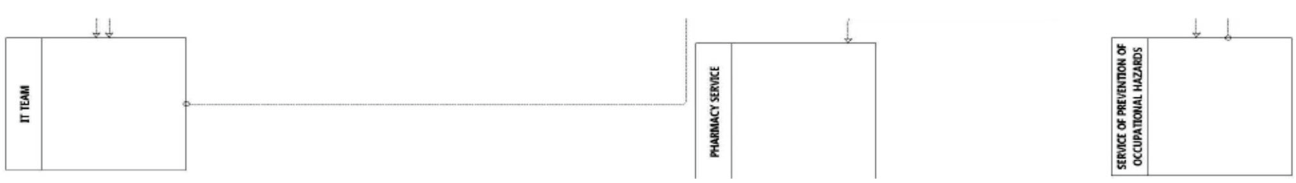
1 2

7. Using as many events as necessary, although there are more than the number of participants… makes the model

   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these

8. Eliminating participants of “IT team”, “pharmacy service” and “Service of prevention of occupational hazards”…makes the model

   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these



Questionnaire 2: Business Process models of “First consultation appointment” and “Appointment for examination with tests”

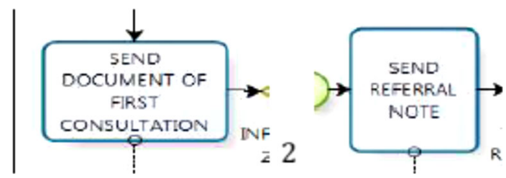
Let us imagine we want to analyze a BP model to understand and modify it. Indicate what proposed changes most affect the understandability or modifiability of models. For example, if you think that eliminating the participants “SAU” and “OGLE” makes the model easier to understand and modify, choose this option.

1. Selecting the common part of subprocess “Appointment for examination with tests” and “First consultation appointment”, for example the activities “Go to SAU” or “Receive correspondent centre” in the participant “new employee” and moving them to the parent model “Appointments”, in order to indicate the particular activities of each subprocess in each subprocess … makes the model

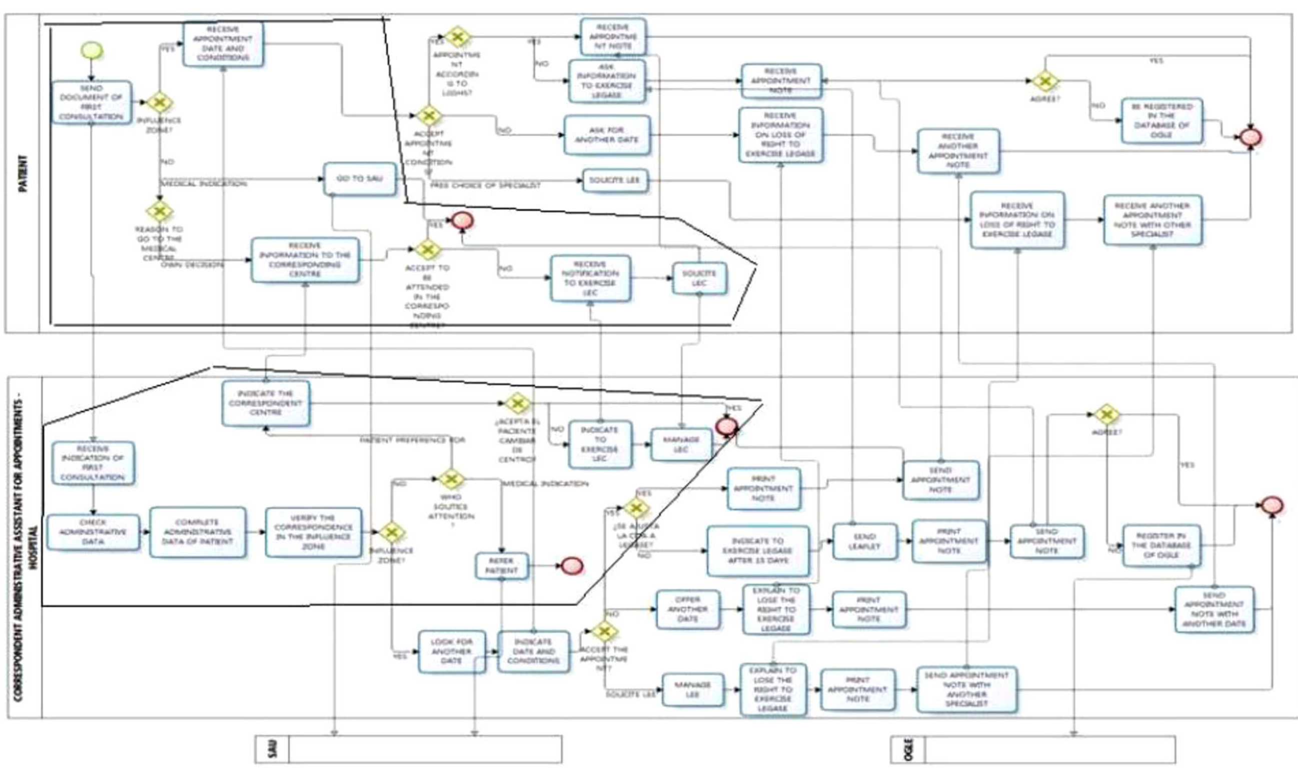
   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these

2. Unifying the activity “Send document of first consultation” [1] in the subprocess “First consultation appointment” and the activity “Send referral note” [2]in the subprocess “Appointment for examination with tests” in only one (in common for both) called “send referral note”, indicating it in the parent process “Appointments”... makes the model

   ☐ Easier to understand or modify
   ☐ Less easy to understand or modify
   ☐ Neither of these

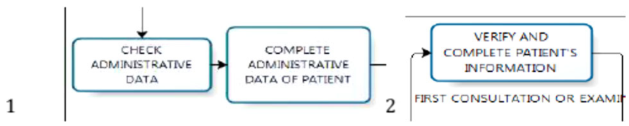


3. Unifying activities of “check patient address” in the participant “administrative assistant for appointments” in the subprocess “Appointment for examination with tests” and the activity “check administrative data” and “complete administrative data of patient” [1] in the process “First



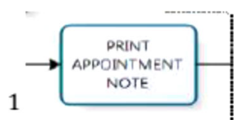
consultation appointment" as only one for both subprocesses, and moving it to the parent process "Appointments" ... makes the model

☐ Easier to understand or modify
☐ Less easy to understand or modify
☐ Neither of these



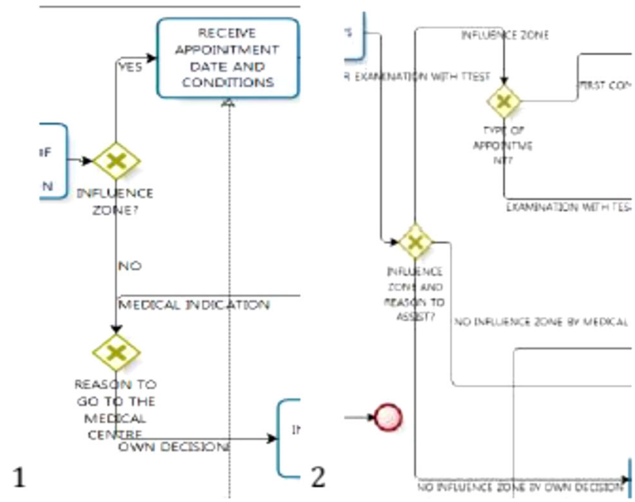
4. Using basic activities such as "print appointment note" [1] in the subprocess "First consultation appointment" in the participant "administrative assistant for appointments" instead of not including them...makes the model

☐ Easier to understand or modify
☐ Less easy to understand or modify
☐ Neither of these



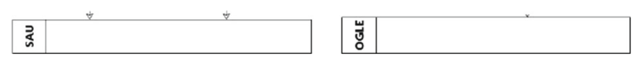
5. Using repeated activities after the satisfaction of two different conditions [1] instead of indicating it only once [2], for example in the subprocess "First consultation appointment"… makes the model

☐ Easier to understand or modify
☐ Less easy to understand or modify
☐ Neither of these



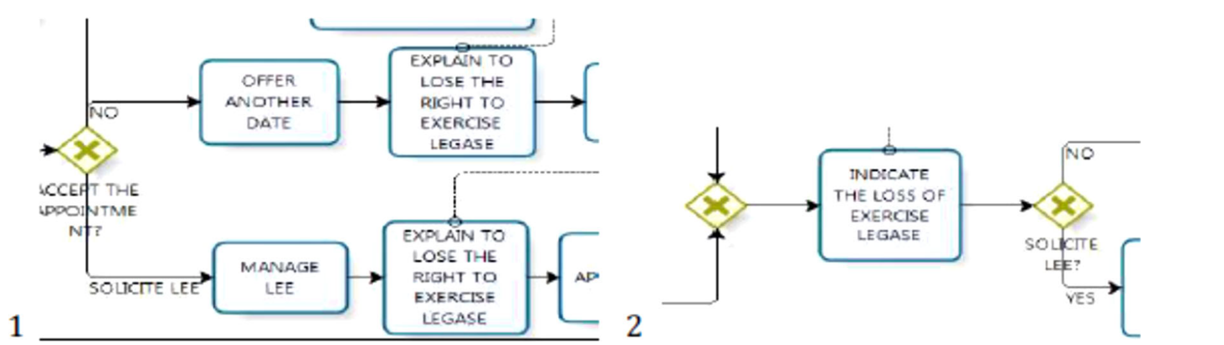
7. Eliminating the participants "SAU" and "OGLE" in both subprocesses...makes the model

☐ Easier to understand or modify
☐ Less easy to understand or modify
☐ Neither of these



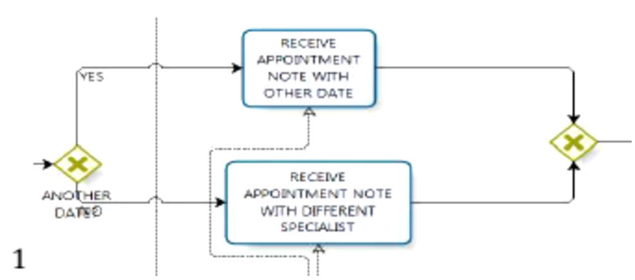
8. Using "join" gateways accompanying "Split" gateways, for example in "another date?" [1] in the subprocess "First consultation appointment...makes the model

☐ Easier to understand or modify
☐ Less easy to understand or modify
☐ Neither of these



6. Merging questions from two gateways into only one. The Gateway "Influence zone" and "Reason to go to medical centre" in the participant "Administrative assistant for appointments" in both subprocesses into only one gateway which includes both questions… makes the model

☐ Easier to understand or modify
☐ Less easy to understand or modify
☐ Neither of these



9. Using a start event in all the process and subprocesses in each participant …makes the model

☐ Easier to understand or modify
☐ Less easy to understand or modify
☐ Neither of these

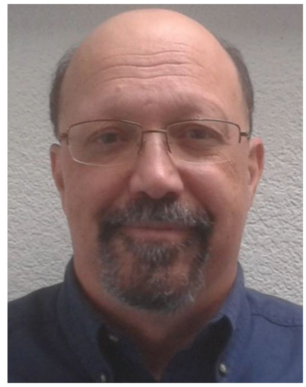
References

1. Kettinger, W.J., Teng, J.T.C., Guha, S.: Business process change: a study of methodologies, techniques and tools. MIS Q. **21**, 55–80
2. Weske, M.: Business Process Management: Concepts, Languages, Architectures, 1st edn. Springer, New York (2007)
3. Dumas, M., et al.: Fundamentals of Business Process Management. Springer, New York (2013)
4. Recker, J., Safrudin, N., Rosemann, M.: How novices model business processes. In: Proceedings of the 8th international conference on Business process management, pp. 29–44. Springer-Verlag, Hoboken, NJ, USA (2010)
5. Leopold, H., Smirnov, S., Mendling, J.: On the refactoring of activity labels in business process models. Inf. Syst. **37**(5), 443–459 (2012)
6. Moody, D.: Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. Data Knowl. Eng. **55**, 243–276 (2005)
7. Sánchez-González, L., et al.: Toward a quality framework for business process models. Int. J. Coop. Inf. Syst. **22**(01), 1350003 (2013)
8. de Oca, Moreno-Montes, et al.: A systematic literature review of studies on business process modeling quality. Inf. Softw. Technol. **58**, 187–205 (2015)
9. Sánchez-González, L., et al.: Measurement in business processes: a systematic review. Bus. Process Manag. J. **16**(1), 114–134 (2010)
10. García, F., et al.: Towards a consistent terminology for software measurement. Inf. Softw. Technol. **48**, 631–644 (2005)
11. Yu-dong, Q., Ning, Q., Xie, X.-F.: Towards a preliminary ontology for conceptual model quality evaluating. In: International conference on web information systems and mining, pp. 329–334 (2010)
12. Mohagheghi, P., Dehlen, V., Neple, T.: Definitions and approaches to model quality in model-based software development: a review of literature. Inf. Softw. Technol. **51**(12), 1646–1669 (2009)
13. Gucegliogllu, S., Demirors, O.: Using software quality characteristics to measure business process quality. In: van der Aalst, W.M.P., Benatallah, B., Casati, F., Curbera, F. (eds.) Business Process Management. Lecture Notes in Computer Science, vol. 3649, pp. 374–379. Springer, Berlin (2005)
14. Rittgen, P.: Quality and perceived usefulneses of process models. In: SAC 10, proceedings of the ACM symposium on Applied computing (2010)
15. Satpathy, M., et al.: A generic model for assessing process quality. In: Proceedings of the 10th international workshop on new approaches in software measurement, pp. 94–110. Springer-Verlag (2000)
16. Matook, S., Indulska, M.: Improving the quality of process reference models: a quality function deployment-based approach. Decis. Support Syst. **47**(1), 60–71 (2009)
17. Heravizadeh, M., Mendling, J., Rosemann, M.: Dimensions of business processes quality (QOBP). In: 6th internatioanl conference on business process management workshops (BPM workshops 2008) (2008)
18. Lindland, O.I., Sindre, G., Solvberg, A.: Understanding quality in conceptual modeling. IEEE Softw. **11**(2), 42–49 (1994)
19. Becker, J., Rosemann, M., von Uthmann, C.: Guidelines of Business Process Modeling, in Business Process Management. Springer, Berlin (2000)
20. Yu-dong, Q., et al.: Analysis of contribution of conceptual model quality to software reliability. In: 2010 International conference on computer application and system modeling (ICCASM) (2010)
21. ISO/IEC, 9126–1, Software Engineering—Product Quality—Part 1: Quality Model (2001)
22. ISO/IEC, 25010, Systems and Software Engineering—System and Software Product Quality Requirements and Evaluation (SQuaRE)—System and Software Quality Models (2011)
23. Figl, K., Recker, J., Mendling, J.: A study on the effects of routing symbol design on process model comprehension. Decis. Support Syst. **54**(2), 1104–1118 (2013)
24. Recker, J., Dreiling, A.: The effects of content presentation format and user characteristics on novice developers' understanding of process models. Commun. Assoc. Inf. Syst. **28**(6), 65–84 (2011)
25. Sweller, J.P., Chandler, P.: Why some material is difficult to learn. Cognit. Instr. **12**(3), 185–223 (1994)
26. Reijers, H.A., et al.: Syntax highlighting in business process models. Decis. Support Syst. **51**(3), 339–349 (2011)
27. Mendling, J., Strembeck, M., Recker, J.: Factors of process model comprehension—findings from a series of experiments. Decis. Support Syst. **53**(1), 195–206 (2012)
28. Houy, C., Fettke, P., Loos, P.: Understanding understandability of conceptual models—what are we actually talking about? In: Proceedings of the 31st international conference on conceptual Modeling, pp. 64–77. Springer-Verlag, Florence, Italy (2012)
29. Sutcliffe, A., Kurniawan, S., Shin, J.E.: A method and advisor tool for multimedia user interface design. Int. J. Hum. Comput. Stud. **64**, 375–392 (2005)
30. Kim, J., Lee, J., Choi, D.: Designing emotionally evocative homepages: an empirical study of the quantitative relations between design factors and emotional dimensions. Int. J. Hum. Comput. Stud. **59**(6), 899–940 (2003)
31. Sharp, A., McDermott, P.: Workflow Modeling: Tools for Process Improvement and Application Development. A.H. Publishers, London (2001)
32. Mendling, J., Reijers, H.A., van der Aalst, W.M.P.: Seven process modeling guidelines (7PMG). Inf. Softw. Technol. **52**(2), 127–136 (2010)
33. Sánchez-González, L., et al.: Quality Assessment of Business Process Models Based on Thresholds. In: Meersman, R., Dillon, T., Herrero, P. (eds.) On the Move to Meaningful Internet Systems: OTM 2010, pp. 78–95. Springer, Berlin (2010)
34. Mendling, J., et al.: Thresholds for error probability measures of business process models. J. Syst. Softw. **85**(5), 1188–1197 (2012)
35. Sanchez-Gonzalez, L., et al.: Quality indicators for business process models from a gateway complexity perspective. Inf. Softw. Technol. **54**(11), 1159–1174 (2012)
36. Cardoso, J.: Process control-flow complexity metric: an empirical validation. In: SCC'06: Proceedings of the IEEE international conference on services computing, pp. 167–173 (2006)
37. OMG: Business process model and notation (BPMN) 2. http://www.omg.org/bpm (2011)
38. Rolón, E., et al.: Process modeling of the health sector using BPMN: a case of study. Healthinf **2**, 173–178 (2008)
39. Mendling, J.: Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness. Springer, New York (2008)
40. Zweig, M., Campbell, G.: Receiver-operating characteristic (ROC) Plots: A fundamental evaluation tool in clinical medicine. Clin. Chem. **39**(4), 561–577 (1993)

41. Bender, R.: Quantitative risk assessment in epidemiological studies investigating threshold effects. Biom. J. **41**(3), 305–319 (1999)
42. Rolón, E., García, F., Ruiz, F.: Evaluation measures for business process models. In:Symposioum in applied computing SAC06 (2006)
43. van der Aalst, W.M.P., et al.: Workflow patterns. Distrib. Parallel Databases **14**(1), 5–51 (2003)
44. Runeson, P., Höst, M.: Guidelines for conducting and reporting case study research in software engineering. Empir. Softw. Eng. **14**, 131–164 (2009)
45. Yin, R.K.: Case Study Research, Design and Methods. Sage Publications, Beverly hills (2003)
46. Robson, C.: Real World Research: A Resource for Social Scientists and Practitioner-Researchers, 2nd edn. Blackwell Publishing, Malden (2002)
47. Benbasat, I., Goldstein, D.K., Mead, M.: The case research strategy in studies of information systems. MIS Q. **11**(3), 369–386 (1987)

Laura Sánchez-González finished her MSc in 2007 and PhD in 2012, both in Computer Science at the University of Castilla-La Mancha (UCLM), Spain. She is a member of Alarcos Research Group, Technologies and Information Systems Institute. Her activity is in the field of business processes, especially in process modeling and measurement topics.

Félix García received his MSc (2001) and PhD (2004) degrees in Computer Science from the University of Castilla-La Mancha (UCLM). He is currently an associate professor in the Department of Information Technologies and Systems at the UCLM. He is member of the Alarcos Research Group and his research interests include business process management, software processes, software measurement and agile methods. More information about publications in these topics can be found in http://alarcos.esi.uclm.es/staff/fgarcia.



Francisco Ruiz has a PhD in Computer Science for the University of Castilla-La Mancha (UCLM) and an MSc in Chemistry–Physics for the Complutense University of Madrid. Full professor of the Department of Information Technologies and Systems at UCLM in Ciudad Real (Spain). Between 1993 and 2000, he was Dean of the Computer Science Faculty. Previously, he was CIO (Chief Information Officer) in the mentioned university (1985–1989). Along last 30 years, he has also worked in private companies as analyst-programmer, project manager and IT consultant. His current research interests include enterprise architecture; integration of paradigms SOC (service-oriented computing), MDE (model-driven engineering) and BPM (business process management); method engineering and software processes. He has written 8 books and 27 chapters and he has published 26 articles in refereed international journals and more than one hundred papers in other journals, congresses, conferences and workshops. He belongs to several scientific and professional international associations (ACM, IEEE-CS, PMI, CIO Index, AEA, ISO JTC1/SC7, EASST).

Mario Piattini is a full professor at the UCLM. He holds the PhD degree in Computer Science from the Technical University of Madrid and leads the Alarcos Research Group. He is CISA, CISM, CGEIT and CRISC by ISACA. His research interests include software quality, metrics and maintenance. More information can be found in http://alarcos.esi.uclm.es/staff/mpiattini.