This is a post-peer-review, version of the paper: Trkman, M.^a, Mendling, J.^b, Trkman, P.^c & Krisper, M.^d Impact of the conceptual model's representation format on identifying and understanding user stories. Published in Information and Software Technology (Volume 116, December 2019). The final version is available on ScienceDirect at: <https://doi.org/10.1016/j.infsof.2019.08.001>. This version of the paper is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs Licence (CC BY-NC-ND)

^a Jozef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia

^b Wirtschaftsuniversität Wien - WU Vienna, Institute for Information Business, Building D2, Entrance C, 3rd Floor, Welthandelsplatz 1, 1020 Vienna, Austria

^c University of Ljubljana, School of economics and business, Kardeljeva ploscad 17, 1000 Ljubljana, Slovenia

^d University of Ljubljana, Faculty of computer and information science, Vecna pot 113, 1000 Ljubljana, Slovenia

Impact of the conceptual model's representation format on identifying and understanding user stories

Abstract

Context. Eliciting user stories is a major challenge for agile development approaches. Conceptual models are used to support the identification of user stories and increase their understanding. In many companies, existing model documentation stored as either use cases or BPMN models is available. However, these two types of business process models might not be equally effective for elicitation tasks due to their formats.

Objective. We address the effectiveness of different elicitation tasks when supported either with visual or textual conceptual model. Since the agile literature shows little attention to reusing existing BPMN documentation, we propose several hypotheses to compare it to the use of textual use case models.

Method. We conducted an experiment to compare the effectiveness of the two business process formats: textual use cases and visual BPMN models. We studied their effects on three elicitation tasks: identifying user stories and understanding their execution-order and integration dependencies.

Results. The subjects better understood execution-order dependencies when visual input in the form of BPMN models was provided. The performance of the other two tasks showed no statistical differences.

Conclusion. We addressed an important problem of user story elicitation: which informationally equivalent model (visual BPMN or textual use case) is more effective when identifying and understanding user stories.

1 Introduction

The effective elicitation of user stories is a major challenge for agile software development projects [1-3]. Project stakeholders have to identify and understand user stories. A key challenge of understanding user stories relates to the interdependencies between them [4]. Execution order and integration dependencies are particularly important in this respect [2, 3, 5].

Conceptual models are fundamental to system analysis [6] and can help with identifying [7] and understanding [2, 8] user stories. Business process models are conceptual models that are often used to improve the understanding of a business domain and, consequently, the context of requirements in general [9-11]. The actual usage of business process models as conceptual models remains poorly understood, particularly in terms of how and why development project's stakeholders would use multiple models in combination when doing their work [6]. Representation formats of business process models, namely textual (e.g. textual use case models) and visual (e.g. BPMN models), have an impact on task performance [11].

The usage of use case models is well established in the agile development literature. These types of model can be used as a source for identifying user stories [7] and for understanding the dependencies [2]. It has recently been shown that BPMN models can also be beneficial to agile development with user stories. Textual use case models and visual BPMN models are both used to present an organization's business processes [11]. So far, there is limited evidence on the mutual benefits of these two different representation formats of business processes, namely visual and textual, in the context of identifying and understanding user stories. Trkman et al. [8] showed that BPMN business process models help in understanding user stories. However, they did not study the identification task and did not compare BPMN models to use case models.

In this paper, we extend their research by investigating the research question: how effective are BPMN models compared to informationally equivalent use case models in identifying and understanding user stories? We use an experimental design for comparison of two business process representation formats, namely visual (BPMN models) and textual (use case models). Our hypotheses are grounded in cognitive fit theory, cognitive load theory, and cognitive theory of multimedia learning. Our main finding is that different representation formats are suitable for different purposes. More specifically, the BPMN model is better at presenting alternative paths of a business process and therefore enables a better understanding of a user story's execution order dependencies.

We proceed as follows. In Section 2, we discuss the approaches to user story identification and which ones are better for understanding the dependencies. Furthermore, we provide a theoretical justification for the hypotheses. In Section 3, we describe the methodology and results of the experiment. In Section 4, we discuss the findings of our research.

2 Background

2.1 Conceptual modelling

Conceptual modelling is one of the cornerstones of information systems engineering. It helps capturing the general knowledge of the system or the business process in a conceptual schema [12]. Conceptual modelling is often used in requirements engineering as a way to improve the quality of the software production process, in particular when software requirements identified [13]. A major task in this context is to document the common understanding that stakeholders have about a real-world domain that is intended to be supported by an information system [14]. Modelling as a discipline has been around for several decades, with a huge body of knowledge developed in academic and practitioner literature. The benefits of modelling and its contributions to organizational tasks are well known [15].

For these reasons, the use of conceptual models in engineering and research is highly recommended. However, conceptual models must be of high quality, which applies to both conceptual models and the modelling language used to build them [16]. This brings about several challenges and increases the importance of quality evaluation [16]. So far numerous methods have been proposed for requirements modelling, but few have seen widespread uptake in practice [17].

Modelling languages have proved to be an effective tool to specify and analyse various perspectives of enterprises and information systems [18]. In this respect, business process models are key artefacts in the development of information systems [19]. A process model can provide a comprehensive understanding of a process, if it is captured in a correct and complete way. Using the right model involves taking into account the purpose of the analysis [20]. Potential misalignments between purpose and model in academia and industry are discussed in [18]. The understanding of process models by end-users is not only an important concern in industry, but also field of ongoing academic investigation [19]. The understandability is a specific quality aspect that is connected with personal-, model-, and content-related factors [21].

One weakness of many requirements modelling methods is lack of empirical evaluation of their benefits. This means that there is a need for evaluation that considers both the theoretical and practical aspects [17]. In this respect, several studies have focused on identifying and understanding user stories [22] by the help of experimental design involving students are proxies for junior analysts [23], which is also an approach used in our paper. Insights from an empirical evaluation also offers the opportunity to enhance method and models [24]. In this way, model-driven system design is informed [25].

It is therefore important to collect empirical evidence concerning the conditions under which the use of a certain modelling approach makes a practical difference [26]. Typical empirical evidence in this area includes a paper on the impact colours in tailoring process models to enhance understandability [27] or that offering more material to participants to read and browse increases that effort but also improves comprehension [28].

2.2 Identification of user stories

Identifying user stories is an important task within the overall elicitation process. In essence, identification refers to writing user stories by the help of a particular user story template, such as “I as a <role> can <activity>” with an optional addition of [so that <benefit>] [29, 30]. The identification task is typically assigned to multiple development projects’ stakeholders. The users are often the ones who need to be trained on how to identify user stories [3, 31-33]. The adoption of user stories to specify user requirements has grown substantially in recent years [1, 22, 34, 35]. However, user stories can fail in providing an overview of the development project [36]. Thus, empirical studies on the effectiveness of user stories are needed [34].

One way of addressing potential quality issues is to utilize modelling techniques in order to support user story identification. These techniques include interviews [3, 37, 38], which are often advocated as an effective method [38-40], as well as surveys [3], focus group sessions [2], and story-writing workshops [3]. Sometimes, these approaches are complemented by an organization’s existing documentation, which is often available as business process models [7, 41, 42].

Business processes can be modelled in different representation formats, namely visual and textual. An organization’s existing business process models can be available as textual use case or visual models [7]. The decision which software development approach to choose is often made once the scope of the effort is understood and project participants have been identified [7]. It is important to choose the approach that best aligns with the nature of the project, team, and organizational culture. Therefore, if the decision has not yet been made to use Extreme programming, then the persons who are gathering the initial requirements might choose to create something other than user stories, for example, use cases. In this situation, they would need to convert use cases into user stories.

Figure 1 shows how several pieces of information that is needed for identifying user stories can be obtained from a textual use case model. The actor captures the information needed to identify the user story’s role, and the event describes the user story’s activity.

Figure 1: Identifying user stories from a textual use case model of a bank (from the educational material in Appendix E)

In many companies, documentation in terms of business process models are readily available as BPMN models [43-45]. Companies maintain repositories of several hundred to several thousand process models [46] initially made for diverse purposes. Once business process models from the repository are selected, project stakeholders need to pin down each model's

activity elements which will be affected by new software and identify user stories as suggested in Figure 2. Depending on the granularities of the BPMN model's activity elements, the derived user stories can have different granularity levels (size, complexity). Consequently, some user stories derived from the same business process model can have the size of an epic, while others can be smaller and have the size of a feature or user story. Nevertheless, they are all backlog items, and in our paper, we refer to them as “user stories”.

Figure 2 presents an informationally equivalent BPMN process model to that of Figure 1. Figure 2 shows how user stories can be identified by using BPMN models. The swimlane captures the information needed to identify the user story's role, and the activity element describes the user story's activity.

Figure 2: Identifying user stories from a BPMN model of a bank (from the educational material in Appendix E)

The business process models in Figures 1 and 2 are based on different traditions in software development. BPMN has its roots in Web service composition [47], while use cases are used for agile development [7, 48]. Despite these differences, a substantial conceptual overlap exists, which means that business processes can be captured by both notations in an informationally equivalent way. We have chosen to compare these two specific notations because the use case models have well-established use in agile development, while BPMN models do not. Our findings could promote the use of BPMN models in agile development. Up until now, there has been no research on whether a textual or visual business process model is more effective for identifying and understanding user stories.

2.3 Understanding user stories

A key challenge for eliciting user stories is not only their identification but also ensuring understanding of their context by all development project stakeholders. One of the problems that can arise from using a list of user stories is missing the big picture [2]. Having a conceptual model of the development project is important because user stories are not necessarily isolated from one another. They can overlap or be related in terms of temporal and logical order. Prior research distinguishes two categories of user story dependencies: execution order and integration dependencies [2, 3, 5]. Understanding these dependencies can increase the understanding of user stories [8].

Execution order dependencies refer to the fact that the completion of a specific user story is the trigger for starting another one. Various studies emphasize the need to make execution order dependencies explicit. Different types of conceptual models are proposed as a means of doing so in this context, including scenarios [49, 50], requirement networks [4], level-0 data-flow diagrams [7], BPMN models [8], and events of use case models [2].

To support understanding of the execution order dependencies, textual and visual formats of business process models can be treated as alternatives. The textual use case model was previously discussed by Leffingwell [2] as a means with potential to support understanding of execution order dependencies. We decided to empirically evaluate his suggestion and compare its effectiveness with the effectiveness of BPMN models. The effectiveness of BPMN models as a means to support understanding of user stories has been evaluated before. Trkman et al. [8] showed that user stories' execution order dependencies are better understood when the list of user stories is supported by visual BPMN models, versus not being supported by any conceptual model at all. We believe BPMN models have unexplored potential, so we continued the work of Trkman et al. by comparing the effectiveness of use case and BPMN models on understanding execution order dependencies. In Section 3.3, we explain how to understand execution order dependencies by using BPMN and use case models.

Integration dependencies refer to situations in which a user story describes a matter that is part of another more general user story [2, 3]. In such a case, one user story can be regarded as a refinement of an aspect identified in another one [5]. Several approaches have been proposed to better understand integration dependencies. These include visualization by placing cards with more fine-granular user stories below those of the more general user story [50] or placing user story association codes next to BPMN activity elements [8].

The understanding of the integration dependencies of user stories can be supported by both BPMN models [8] and use case models. Trkman et al. [8] showed that a user story's integration dependencies are better understood when the list of user stories is supported by visual BPMN models, versus not being supported by any conceptual model at all. In this study, we go further: we propose a simple technique for organizing knowledge on integration dependencies by using a textual use case model. For this purpose, we took Patton's [50] technique of placing user story cards of finer granularity under those of a more abstract nature and applied it to a textual use case model with a table view. In Section 3.3, we demonstrate how to understand integration dependencies by using BPMN and use case models.

2.4 Hypotheses

The impact of the representation format on task performance has been demonstrated in different application areas of information systems research and computer science. In our research, we build on cognitive fit theory, cognitive load theory, cognitive theory of multimedia, and the requirements engineering field.

Cognitive load theory was developed to provide guidelines intended to assist in the presentation of information in a manner that encourages learner activities that optimize mental performance. It was developed by John Sweller [51] based on a study of problem-solving. In the 1990s, cognitive load theory was extended in different directions, among others emphasizing the so-called *split attention effect*, which was used to support our hypotheses. The split attention effect is the decrease of cognitive performance due to information being scattered, such that it cannot be perceived at a single glance.

The initial argument for the *cognitive fit theory* was provided by Larkin and Simon [52] and was refined in 1991 by Vessey [53, 54]. The key proposition of the theory is that a task's performance depends upon the learner's mental model of both the task at hand (problem-solving task) and the input material's representation format. Problem-solving with cognitive fit between task and representation format leads to effective and efficient problem-solving performance [53].

Agarwal et al. [55] emphasized the importance of the cognitive fit theory in conducting experiments on the usefulness of various methodologies and models for requirements engineering. In the context of user story elicitation, we focus on testing the performance of three interrelated but different tasks: identification of user stories, understanding of execution order, and understanding of integration dependencies. According to the cognitive fit theory and cognitive theory of multimedia learning, the effect of using different representation formats is different for these three tasks.

Mayer [56] developed the *cognitive theory of multimedia learning*, which focuses specifically on how to design effective multimedia material (using words and pictures) to foster learning. He developed and tested design principles to reduce extraneous processing (e.g. signalling principle), manage essential processing, and foster generative processing (e.g. multimedia principle). The cognitive theory of multimedia learning has three assumptions about how humans learn:

- Humans possess separate information-processing channels for represented material. One way of conceptualizing the differences is with the presentation-mode approach. The presentation-mode approach focuses on whether the presented stimulus is verbal (e.g. spoken or printed words) or nonverbal (e.g. illustrations). In our experiment, we address the verbal channel when using use case models and the nonverbal channel when using BPMN models.
- Humans are limited in the amount of information that can be processed in each channel at one time. Due to this limitation, we used only a small part of the company's specific business process model in our materials. For example, in the BPMN model, which addresses the opening of the account (Case OPEN), we focused on nine activity elements and one decision point. Similarly, for closing the account (Case CLOSE), we used eight activity elements and one decision point.
- Humans actively engage in cognitive processing to construct a coherent mental model (knowledge structure) of their experiences. One of the basic mental models is a

process represented as a flow chart. In our BPMN treatment, the BPMN model was a stimulus for the learner to better create his or her mental model.

To identify user stories, the required information about the user story's role and activity can be obtained from both the BPMN and use case models. However, they differ in how closely relevant information items are positioned relative to one another. While textual use case models show both the user story code and event in the same table row, BPMN models enable activity elements to be scattered around the model and grouped in swimlanes. Because the activity elements and swimlanes can have quite some distance, we can assume that a *split attention effect* might be observed [51]. Therefore, the identification task might show better cognitive fit with a textual use case model. This leads to hypothesis H1.

H1: The identification of user stories is more accurate when supported by a textual use case model than by a BPMN model.

To understand execution order dependencies, visual representation facilitates quick spatial inspection of the preceding and succeeding activities and their associated user stories. Mayer's *signalling principle* states that individuals learn better when cues that highlight the organization of the essential material are added. BPMN models offer a huge variety of elements, which makes it possible to visually explain complex alternative flows of a business process [57], while use case models do not. Consequently, this might represent an advantage for understanding execution order dependencies supported by BPMN models. This leads to hypothesis H2.

H2: The understanding of the execution order dependencies of a user story is greater when supported by a BPMN model than by a textual use case model.

To understand the integration dependencies, the pointers to user stories are placed approximately equally close to the activity/event they refer to in both representations. This proximity is in line with the *spatial contiguity principle* formulated by Mayer [56], which states that learners learn better when corresponding words and pictures are presented near rather than far from each other on the page. Also, the occurrence of a *split attention effect* is avoided. However, Mayer's *multimedia principle* says that persons learn better from words and pictures than from words alone. When words and pictures are both presented, learners have an opportunity to construct verbal and visual mental models and to build connections between them. Also, pictures offer the option to inspect the material in different directions from a distance and to group visual elements into logical units [52]. Consequently, BPMN material combines a visual BPMN model and textual user stories to foster generative processing, while use case material involves textual representation only. For this reason, a BPMN model should support the understanding of integration dependencies better than a use case model. This leads to hypothesis H3.

H3: The understanding of the integration dependencies of a user story is greater when supported by a BPMN model than by a textual use case model.

There are differences to take into account in terms of the granularity of user stories associated with corresponding business process activities or events. Therefore, for each hypothesis, we investigated two types of association granularity, namely when a user story is a) approximately equally as big as or b) more abstract than its corresponding activity element

(in the BPMN model) or event (in the use case model). Table 1 lists the respective levels and their characteristics.

Granularity level of user story association	Characteristics
More abstract	<p>The user story is associated with multiple activity elements (BPMN model)/events (use case model).</p> <p>AND</p> <p>The associated activity elements (BPMN model) are between the same “OR decision” and “OR merge” gateway.</p> <p>AND</p> <p>The user story’s activity is more abstract than associated activity elements (BPMN model)/events (use case model).</p>
Equally detailed	<p>The user story is associated with one activity element (BPMN model)/event (use case model).</p> <p>AND</p> <p>The user story’s activity is equally detailed as the activity element (BPMN model)/event (use case model).</p>

Table 1: Definitions of “more abstract” and “equally detailed” adapted from Trkman et al. [8]

3 Research method

3.1 Research design

To contribute to middle-range theory, we used a generalization strategy from lab to field in a sample-based way [58]. We took the guidelines by Juristo and Moreno [59] into account, most importantly in terms of why and how to experiment, the design of the experiment, the analysis of the data, assuring the replications, and precautions to be taken into account with this research approach. To test our hypotheses, we prepared a crossover experiment design [60]: each subject is exposed to two treatments; however, different subjects receive the treatments in a different order. We use variants of two business cases for each treatment, as well as the order of the treatments.

We argue that the two treatments have a different effect in a population of business users of an information system who are involved in identifying (Task 1) and understanding (Tasks 2-3) user stories. The mechanism that produces the causal influence is the representation format of a conceptual model. The conceptual model in our experiment was the business process model. We compared subjects’ performance in solving Tasks 1-3 while being exposed to different representation formats of the business process model, namely visual (BPMN model) and textual (use case model). We took multiple and complex issues involved in the design and analysis of crossover experiments into account [60]. In Section 4.1, we discuss the experiment’s threats to validity as suggested by Wohlin et al. [61].

3.2 Material

We started designing the experimental material using real-world BPMN models from a European bank. We chose two models related to account management, namely “close a bank account” (Case CLOSE) and “open a bank account” (Case OPEN). We modified these two models to make them compliant with the quality criteria of the seven process modelling guidelines [10]. This led to a consistent usage of message flow arrows, the addition of merge gateways for decision gateways, the translation of labels into English, and the usage of verb-object activity labels. In both models, we used the same set of BPMN elements (activity,

pool, swimlane, OR gateway, and flow arc). We used the definitions from Table 1 to systematically create user stories that were either equally as detailed as or more abstract than the associated activity description.

From the BPMN models, we created informationally equivalent textual use case models. We used the mapping rules summarized in Table 2 to ensure that both the BPMN and use case models captured the same information content. We adopted only the sections of the use case model as described in the Rational Unified Process (RUP) which were needed to capture the corresponding process information, namely the Basic flow of events, Alternative flows of events, and Additional information. We created two use case models for the two cases, CLOSE and OPEN.

Type of information	BPMN model	Use case model
User story's function	Activity description	Event description
User story's user role	Swimlane	Actor
Big picture	Business process name	Use case name
Associated user stories	User story references placed next to the activity descriptions	User story references placed in the same row as the event descriptions
Conditions in which a set of alternative flows can be started	OR split gateway element	Note in the "Additional information" section
Conditions in which business process activity can be started	OR merge gateway element	Note in the "Additional information" section
Situation "more abstract"	Yellow text box with a user story reference within it	Yellow text colouring of a column in tables
Situation "equally detailed"	Blue text box with a user story reference within it	Blue text colouring of a column in tables

Table 2: Ensuring the informational equivalence of the BPMN and use case models

To create material for Task 1 (the identification of user stories), we enriched the model in Appendix A with text boxes (user story codes) placed next to the activity elements (Figure 3). On the other hand, the informationally equivalent use case model in Appendix B was enriched by placing events in tables and adding an additional column with user story references (Figure 4).

The material for Tasks 2 and 3 (the understanding of user stories) was built from the material for Task 1. We additionally provided reference codes to user stories that were *more* abstract than the associated activity and a corresponding list of user stories (Figure 5). We created user stories by using the short version of the user story template since <benefit> is not crucial for understanding integration and execution order dependencies. In the informationally equivalent use case model, we added another column to the table which has references to user stories that were more abstract than the associated activity/event (Figure 6).

Figure 3: BPMN treatment: Case OPEN material for Task 1 (from Workbook A)

Use case name: OPEN A BANK ACCOUNT

Basic flow of events

Actor	Event	US
Clerk in the front office	Note 1 Decision #1: according to what the customer wants none, one (A or B) or two alternative flows (A and B) are triggered. See additional Note 2.	
Clerk in the front office	Send notification letter to DEPT A. Note 3	US_21

Alternative flows of events

Alt. flow	Actor	Event	US
A	Clerk in the front office	Fill in an application to order a new Maestro card.	US_24
	Clerk in DEPT B	Assess the proposal for the card.	US_26
B	Clerk in the front office	Fill in a form to open e-banking.	US_23
	Commission member	Evaluate the application for it.	US_27

Additional information:

Note 1: After the clerk in the front office changes the status of customer's account to "active" he/she can (if the customer wants to) order a Maestro card and open e-banking.

Note 2 on Decision #1: No matter how many alternative flows are triggered, clerk in the front office always fills in an application to order an identification card. He/she makes it on the spot and hands it to the customer.

Note 3: Clerk in DEPT A opens physical archive file for the account.

Figure 4: Use case treatment: Case OPEN material for Task 1 (from Workbook B)

Case 1 material (upgraded)

List of user stories for Case 1:

- US_20 I as a commission member can "evaluate an application for a MasterCard."
- US_21 I as a clerk in in the front office can "send notification letter to DEPT A".
- US_22 I as a clerk in DEPT B can "evaluate the application".
- US_23 I as a clerk in the front office can "fill in a form to open e-banking".
- US_24 I as a clerk in the front office can "fill in an application to order a new Maestro card".
- US_25 I as a commission member can "evaluate the application".
- US_26 I as a clerk in DEPT B can "assess the proposal for the card".
- US_27 I as a commission member can "evaluate the application for it".
- US_29 I as a clerk in the front office can "fill in a form".
- US_31 I as a clerk in in the front office can "send internal notification letter".

Figure 5: BPMN treatment: Case OPEN material for Tasks 2 and 3 (from Workbook A)

Case 2 material (upgraded)

List of user stories for Case 2:

- US_20 I as a commission member can "evaluate an application for a MasterCard."
- US_21 I as a clerk in the front office can "send notification letter to DEPT A".
- US_22 I as a clerk in DEPT B can "evaluate the application".
- US_23 I as a clerk in the front office can "fill in a form to open e-banking".
- US_24 I as a clerk in the front office can "fill in an application to order a new Maestro card".
- US_25 I as a commission member can "evaluate the application".
- US_26 I as a clerk in DEPT B can "assess the proposal for the card".
- US_27 I as a commission member can "evaluate the application for it".
- US_29 I as a clerk in the front office can "fill in a form".
- US_31 I as a clerk in the front office can "send internal notification letter".

Use case name: OPEN A BANK ACCOUNT

Basic flow of events

Actor	Event	US	US (+1)
Clerk in the front office	Note 1 Decision #1: according to what the customer wants none, one (A or B) or two alternative flows (A and B) are triggered. See additional Note 2.		
Clerk in the front office	Send notification letter to DEPT A. Note 3	US_21	US_31

Alternative flows of events

Alt. flow	Actor	Event	US	US (+1)
A	Clerk in the front office	Fill in an application to order a new Maestro card	US_24	US_29
	Clerk in DEPT B	Assess the proposal for the card	US_26	US_22
B	Clerk in the front office	Fill in a form to open e-banking	US_23	US_29
	Commission member	Evaluate the application for it	US_27	US_25

Additional information:

Note 1: After the clerk in the front office changes the status of customer's account to "active" he/she can (if the customer wants to) order a Maestro card and open e-banking.

Note 2 on Decision #1: No matter how many alternative flows are triggered, clerk in the front office always fills in an application to order an identification card. He/she makes it on the spot and hands it to the customer.

Note 3: Clerk in DEPT A opens physical archive file for the account.

Figure 6: Use case treatment: Case OPEN material for Tasks 2 and 3 (from Workbook B)

3.3 Tasks

Task 1 focused on writing a list of user stories. First, the subjects had to identify user stories that were equally detailed as the business activity with which they were associated (see the left-hand side of Figure 7). They had to systematically rewrite appropriate text from the input document, which was either a use case (e.g. Figure 1) or BPMN model (e.g. Figure 2). Second, the subjects had to identify user stories that were more abstract than the activity and fill in the gaps on the right-hand side of Figure 7. For this, the subjects needed to refer to the list of detailed user stories on the left page. The subjects were given instructions to generalize a verb or object of a detailed user story and to create a more general user story, which would ideally group several detailed user stories.

A correct user story was defined as a correct combination of a user story role and user story activity. More specifically, our assessment guidelines defined which parts of a user story role and activity had to be the same and the cases in which certain synonyms were acceptable. For control purposes, we also counted the wrong (wrong or missing user role, and wrong or missing activity) and empty user stories (that is, without both user role and activity entries).

WAIT FOR INSTRUCTIONS for basic elicitation:	WAIT FOR INSTRUCTIONS for advanced elicitation:
User story (US) template: I as a <user role> can "<activity>".	Here you need to make more general user stories out of the detailed ones. If possible, try to group at least two detailed user stories under the more general one that you define below. Remember: only one user role can be present in one user story.
Detailed user stories:	More general user stories:
US_21 I as a _____ can " _____ ".	US_a I as a _____ can " _____ ".
US_23 I as a _____ can " _____ ".	US_b I as a _____ can " _____ ".
US_24 I as a _____ can " _____ ".	US_c I as a _____ can " _____ ".
US_26 I as a _____ can " _____ ".	US_d I as a _____ can " _____ ".
US_27 I as a _____ can " _____ ".	

Figure 7: Task 1 for Case OPEN (from all workbooks)

Task 2 focused on the understanding of execution order dependencies. As suggested by Gemino and Wand [62] and Mayer [56], understanding of the given material was measured using problem-solving questions. We used a template question [8] to measure the understanding of execution order dependencies. The reason for using templates is that agile development stakeholders are often concerned with recurring types of questions [2, 7]. An example of Task 2 can be found in Figure 8. To find the execution order dependencies from the textual use case model in Figure 6, the subjects had to trace all preceding events from the event chronologically backward towards the beginning of the model. The informationally equivalent BPMN model in Figure 5 shows that the directly preceding activity elements, from the activity element associated with the user story of US_21, are associated with US_26 or US_27. In both materials, US_26 is directly preceded by US_24, while US_27 is directly

preceded by US_23. Therefore, the two execution order dependencies of US_21 are as follows:

- 1) US_24 → US_26 → US_21
- 2) US_23 → US_27 → US_21

About US_21

a) How to approach to the execution of US_21. Fill in the gaps with user story codes.

Sequence 1: US_ ____ -> US_ ____ -> US_21

Sequence 2: US_ ____ -> US_ ____ -> US_21

b) Between which two user stories in the sequences of a) does a change of user roles happen?

- ☐ I do not know if there are any pairs.
- ☐ I am certain that there are no pairs.
- ☐ There are the following pairs:
US_ ____ & US_ ____; _____

c) Are there any decision points on the way to US_21?

- ☐ I do not know.
- ☐ No, I do not believe there are any decision points.
- ☐ Yes.

If Yes: We want to start the execution of US_21. We are waiting for (correct) input information. What could have gone wrong at previous decision point(s) (please tick)?

☐ Maybe a request for a new Maestro card should have been made but it was not.

☐ Maybe a request for opening e-banking should have been made but it was not.

☐ Maybe a request for new identification should have been made but it was not.

Figure 8: Task 2 for Case OPEN

Task 3 focuses on the understanding of integration dependencies. From Trkman et al. [8], we adopted a template question for measuring the understanding of integration dependencies. An example of Task 3 can be found in Figure 9. To find the integration dependencies from a textual use case model, the subjects needed to trace user stories within the same row (event). For example, the textual use case model in Figure 6 has a column labelled “US” for *equally detailed* user stories and a column labelled “US (+1)” for *more abstract* user stories. Similarly, the BPMN model in Figure 5 shows the user stories’ subordinates by visually placing a box with a more detailed user story (e.g. US_21) within a box containing a more general user story (e.g. US_31). The annotation and visual highlighting in both Figures 5 and 6 show that, for example, US_31 is more abstract than US_21.

Sometimes the activity “send notification letter to DEPT A (activity from US_21)” can be executed in different variations. Which user stories (codes) present those variations?

- ☐ There are no user stories presenting those variations.
- ☐ The following user stories are presenting those variations:

Figure 9: Task 3 for Case OPEN

3.4 Treatments

We created one treatment for one representation format. Consequently, our experiment has two treatments, namely, BPMN and use case. The structure of each treatment is presented in Table 3.

Hypo.	Treatment component	The task is assigned to a user story which has the following association granularity
/	Comprehension test (control purpose)	/
H1	Task 1: Identifying user stories	Equally detailed
		More abstract
H2	Task 2: Understanding execution order dependencies	Equally detailed
		More abstract
H3	Task 3: Understanding integration dependencies	Equally detailed
		More abstract
/	Recall test (control purpose)	/
/	Questions about perceived usability (informative purpose)	/

Table 3: Structure of one treatment

Each workbook had three sequential parts:

- 1) Background questions (control purpose)
- 2) First treatment
- 3) Second treatment

We created four workbooks, namely, A, B, C, and D. As shown in Table 4, Workbooks A (in Appendix C) and B (in Appendix D) started with the BPMN treatment, while Workbooks C and D started with the Use case treatment.

Workbook	Sequence of treatments	BPMN treatment case	Use case treatment case
A (see Appendix C)	First BPMN treatment, then Use case treatment	Case OPEN	Case CLOSE
B (see Appendix D)		Case CLOSE	Case OPEN
C	First use case treatment, then BPMN treatment	Case CLOSE	Case OPEN
D		Case OPEN	Case CLOSE

Table 4: Composition of the workbooks

3.5 Subjects

We used a strategy of subject recruitment similar to many related studies [63]. The subjects were undergraduate students enrolled in the Business Information Systems II course at Vienna University of Economics and Business in Austria. They attended one of six classes

for this course, managed by five lecturers. The subjects were 22.14 years old on average. Of the subjects, 63.9% were male and 36.1% were female. Table 5 provides additional information on the subjects' background. Their previous knowledge about BPMN and use case modelling was approximately equally low.

Variable	Scale	Mean
English	1-7	4.84
Familiarity with BPM	1-7	1.30
Familiarity with the BPMN model	1-7	1.21
Familiarity with the use case model	1-7	1.36
Familiarity with the process of opening a bank account	1-7	1.54
Familiarity with the process of closing a bank account	1-7	1.55

Table 5: Background of the subjects

3.6 Operation

We conducted three pilot experiments between October and November 2015 to test the understanding of the material. We also fine-tuned the material based on the insights from the pilots to make sure it could be completed in one hour.

The workbooks were printed in landscape orientation on both sides. In this way, two pages were seen at the same time (e.g. pages 2 and 3, pages 4 and 5). The left page presented the material, and the answers were filled out on the right-hand side. The workbooks were given to the subjects in a way that ensured that neighbouring subjects had different versions. We instructed them not to flip forward unless they had been told to do so. They were not allowed to flip the page back. We encouraged the subjects to work diligently by offering them bonus points and candy.

All executions followed the same experimental protocol: The lecturer introduced the two researchers who were conducting the experiment. One researcher led the execution, and the other was present to provide any needed help. One researcher introduced the experiment, during which she explained its purpose and motivated the students to participate (average time spent: 5 min). Next, the researcher handed out the workbooks. Then, the subjects answered questions about their background (average time spent: 2 min), particularly their age, gender, and familiarity with different subjects (English language, business process management, BPMN and use case models, specific bank processes).

The researcher used an educational material to explain the business process modelling notation of the *first treatment* (average time spent: 2 min). Next, the subjects solved the comprehension test (maximum time spent: 2 min), which measured the subjects' modelling knowledge. Then, the researcher used a demonstration to show how to identify user stories using the available material (average time spent: 6 min). Afterwards, the students were asked to perform Task 1 (maximum time spent: 8 min). The researcher then used a demonstration example to show how to understand the execution order dependencies of user stories (average time spent: 5 min). Afterwards, the students were asked to perform Task 2 (maximum time spent: 7 min). Next, the researcher used a demonstration to show how to understand the integration dependencies of user stories (average time spent: 1 min). After this demonstration,

the subjects were asked to perform Task 3 (maximum time spent: 2 min). Finally, the subjects were asked to solve a recall test, which was about the context of Tasks 2 and 3, and to evaluate the perceived usability of the material given in Tasks 2 and 3 (maximum time spent: 2 min).

When the subjects completed the first treatment, the researcher briefly introduced the business process modelling notation of the second treatment. The same process was repeated for the second treatment. The experiments lasted between 67 and 74 min.

The educational material used for workbooks in which the first treatment is the BPMN treatment (applied in Workbooks A and B) is available in Appendix E. Figure 10 presents the experiment's execution flow when using Workbook A, in which the first treatment is the BPMN treatment and the treatment case is Case OPEN. Note: in Workbook B, the BPMN treatment case is Case CLOSE.

Figure 10: Execution flow from the subject's point of view for Workbook A

3.7 Experimental results

3.7.1 Descriptive statistics

We collected 130 completed workbooks and applied various data cleaning operations. We searched for outliers based on the subjects' scores in the comprehension and recall tests. The subjects needed to have at least 9 points in both comprehension tests. We eliminated 13 workbooks because of students bad performance on the comprehension test with the use case model and one additional workbook due to bad performance on the comprehension test with the BPMN model. We conducted no additional filtering based on the results of the recall test. Thus, the statistical analysis was based on 116 workbooks.

To verify H1, we summed up the scores from the variables Identification_EQUALus and Identification_ABSTRus into the variable Identification_sum (see Table 6). We calculated measures for H2 and H3 in a similar way. Table 6 also shows the maximum scores of the two cases and the calculation rules. Table 7 shows descriptive statistics for the experimental variables.

Task	Main variable (sum of scoring variables)	Scoring variable	Case CLOSE max score	Case OPEN max score	Scoring rules
/	Comprehension_ sum	/	15	15	One point for every correctly marked checkbox. One point for correct number insertions.
Task 1	Identification_ sum	Identification_ EQUALus	5	5	One point for every correctly defined user story (both <user role> and <function>).
		Identification_ ABSTRus	4	4	
Task 2	ExecOrder_ sum	ExecOrder_ EQUALus	13	13	One point for every correct number insertion. One point for every correct pair of user stories and for choosing the correct radio button. One point for choosing the correct radio button and for every correctly marked checkbox.
		ExecOrder_ ABSTRus	7	7	
Task 3	Integration_ sum	Integration_ EQUALus	1	1	One point for marking the correct radio button.
		Integration_A BSTRus	2	2	One point for marking the correct radio button and for correct insertion of a user story code.
/	Recall_sum	/	6	6	One point for correct insertion in the blank gap.

Table 6: Variables and scoring rules for different parts of the treatment

Variable	BPMN treatment		Use case treatment	
	Mean	Std. dev.	Mean	Std. dev.
Identification_sum	5.95	2.20	5.76	2.04
ExecOrder_sum	11.75	4.83	7.67	5.69
Integration_sum	1.58	1.11	1.62	1.11

Table 7: Descriptive statistics for experiment's main variables

Table 8 shows descriptive statistics of the informative variables addressing perceived usability of the material provided in the treatments.

Informative variable	BPMN treatment		Use case treatment	
	Mean	Std. dev.	Mean	Std. dev.
EaseOfUnderstanding (1-7)	3.874	1.571	2.840	1.396
Frustrating (1-7)	3.756	1.573	4.538	1.770
EaseOfUse (1-7)	3.924	1.445	2.823	1.363
EaseOfLearning (1-7)	4.244	1.501	3.151	1.527

Table 8: Descriptive statistics for the measures of usability

3.7.2 Hypothesis testing

We used the Kolmogorov–Smirnov test to compare the scores of our variables to a normally distributed set of scores with the same mean and standard deviation. The test in Table 9 is significant ($p < .05$) for all variables, which means that the distribution is significantly different from a normal distribution. Because one of the assumptions of parametric tests is normally distributed data, we had to use a non-parametric test.

Non-parametric tests are sometimes known as assumption-free tests. Most of the tests work on the principle of ranking the data: that is, finding the lowest score and giving it a rank of 1, then finding the next score and giving it a rank of 2, and so on. This process results in high scores being represented by large ranks and low scores being represented by small ranks. The analysis is then carried out on the ranks rather than the actual data. The ranks for the variables are presented in Table 10.

There are many non-parametric tests. The Mann-Whitney test is one of the most commonly used non-parametric tests [64]. We chose it because we wanted to test differences between two treatments, and different subjects were used in each condition. Table 4 shows that our subjects were presented with both the BPMN and use case treatment; however, each of the treatments had its own case to work with. The Mann-Whitney test results are presented in Table 11.

The scores for the variable Identification_sum within the BPMN treatment (Mdn = 6) were not significantly different from those within the use case treatment (Mdn = 6), with $U = 6139.00$, $z = -1.17$, $p > 0.05$, $r = -0.076$. These results do not support H1.

The scores for the variable ExecOrder_sum within the BPMN treatment (Mdn = 12) were significantly higher than those within the use case treatment (Mdn = 7), with $U = 3924.00$, $z = -5.49$, $p = <0.05$, $r = -0.360$. The effect accounted for 12.99% of the variance, which points to a medium effect size. These results support H2.

The scores for the variable Integration_sum within the BPMN treatment (Mdn = 1.5) were not significantly different from those within the use case treatment (Mdn = 2), with $U = 6582.00$, $z = -0.30$, $p > 0.05$, $r = -0.020$. These results do not support H3.

Variable	Treatment	Kolmogorov–Smirnov test result	Sig.
Identification_sum	BPMN	0.156	0.000
	Use case	0.216	0.000
ExecOrder_sum	BPMN	0.066	0.046
	Use case	0.112	0.000
Integration_sum	BPMN	0.198	0.000
	Use case	0.164	0.000

Table 9: Kolmogorov–Smirnov test results

Variable	Treatment	N	Mean rank	Sum of ranks
Identification_sum	BPMN	116	121.58	14,103
	Use case	116	111.42	12,925
	Total	232		
ExecOrder_sum	BPMN	116	140.67	16,318
	Use case	116	92.33	10,710
	Total	232		
Integration_EQUALus	BPMN	116	115.24	13,368
	Use case	116	117.76	13,660
	Total	232		

Table 10: Ranks

Hypotheses	H1	H2	H3
Variable	Identification_sum	ExecOrder_sum	Integration_sum
Mann–Whitney U	6139.00	3924.00	6582.00
Z	-1.17	-5.49	-0.30
R	-0.076	-0.360	-0.020
Asymp.Sig (2-tailed)	0.242	0.000	0.767

Table 11: Mann–Whitney test results

4 Discussion

4.1 Validity of the results

External validity refers to the potential of an experiment to be generalized to industrial practice [61]. We discuss three types of interactions with the treatment that can threaten external validity. Firstly, the interaction of treatment and setting relates to the extent to which

the used material and tasks are representative of real-world material and tasks. We took BPMN models that were part of a bank's business process repository to create the BPMN treatment material. In a similar way, any other structured business processes from other business domains (e.g., healthcare) could be used. The textual use case models (for the Use case treatment) were derived from BPMN models. Our list of user stories is fictional, which reduces the external validity. Our experiment's tasks, namely, identification and understanding of user stories, were studied in isolation from other tasks in the software development project. This context can affect the way in which either business users or software developers identify and understand user stories. Also, there are other types of dependencies between user stories aside from execution order and integration dependencies that we did not include in our study.

Secondly, an interaction between treatment and history was related to the extent to which the time of the experiment's operation affected the results. All of the experimental sessions were held on the same week in different classes of the same course. Thirdly, the interaction of treatment and selection is related to the extent to which a subject population is not representative of the population of generalization. We perceived our subjects as novices in both their modelling and prior domain knowledge. Similar to Trkman et al. [8] and Batra et al. [65], we generalized our subjects to the population of typical business users of an information system who are involved in identifying and understanding user stories. Given that we generalized them to the general non-developer population involved in agile development projects, the use of the student sample is justifiable [66], even more so because many previous studies have shown that the decision-making outcomes are similar between business users and students [67, 68]. When the tasks do not require a high level of industrial experience, the use of students should not be a major problem [28].

Internal validity demands that the treatment be isolatable as the single cause of the effect [61]. One concern here is maturation, which causes the subjects to respond differently as time passes [61] due to, for example, learning effect (see [60]), tiredness, or boredom. Because the subjects were exposed to both treatments in sequence, we had to study potential order effects. In Workbooks A and B, the treatment with the BPMN material was first, and the use case model treatment was second. In Workbooks C and D, the order was swapped. This randomization helped to neutralize the potential maturation effects. Furthermore, we ensured that the maximum time for each treatment was limited and that all subjects started both the first and second cases together. They were encouraged to work through to the end of the workbook and were additionally motivated by either bonus points or treats.

Another aspect of internal validity is related to resentful demoralization [61]. This means that a subject who is exposed to a less desirable treatment might give up on finishing the workbooks or not perform as well as he or she could have. The subjects were encouraged to work from the beginning until the end. If they did not solve both treatments, their workbook was discarded. We also controlled their performance with control tests, namely comprehension and recall tests. The workbooks of the subjects who did not pass the comprehension tests for both treatments were discarded. For the Levene test of homogeneity of variance, we summed up multiple scores from the treatment (namely for the comprehension test, identification test, understanding test, and recall test) in the variable SumOfScores. We calculated the mean value of SumOfScores for the following groups: BPMN treatment first (N = 56, Mean = 36.07), BPMN treatment second (N = 60, Mean =

38.73), Use case treatment first (N = 60, Mean = 33.05), and Use case treatment second (N = 56, Mean = 32.82). The Levene test results show no significant differences based on a treatment being the first or second treatment in a workbook. For the SumOfScores, the variances were equal for the BPMN treatment first and BPMN treatment second groups, $F = 0.33$, $p = 0.56$. They were also equal for the Use case treatment first and Use case treatment second groups, $F = 2.85$, $p = 0.09$. Therefore, the treatment results are reliable in that respect.

Conclusion validity is concerned with the relationship between the treatment and the results, as well as the conclusions drawn from the results [61]. One of its aspects concerns the reliability of the measures. One risk is that our directly measured variables (Identification_EQUALus and Identification_ABSTRus, ExecOrder_EQUALus and ExecOrder_ABSTRus, Integration_EQUALus and Integration_ABSTRus) show scattered results compared to their sum variables, namely Identification_sum, ExecOrder_sum, and Integration_sum. We ran a Mann–Whitney test for all nine calculated and directly measured variables mentioned above and focused on the significance of their differences and effect sizes. Three variables had similar significant differences: ExecOrder_sum ($r = 0.36$), ExecOrder_EQUALus ($r = 0.39$), and ExecOrder_ABSTRus ($r = 0.21$). Scores for the other six variables show no significant differences. These findings are in line with our main results in Table 11.

Another aspect concerns the reliability of the experimental replications [61]. The experiment was conducted in six classes, each managed by one of five lecturers. We calculated SumOfScores for the six different classes and conducted the Levene test of homogeneity of variance. The variances for the six classes were not significantly different for SumOfScores in either the BPMN treatment ($F = 1.01$, $p = 0.41$) or Use case treatment ($F = 1.49$, $p = 0.20$). This indicates that replications in our setting were reliable; executions in various settings would be needed to ensure replicability.

Construct validity is related to a threat in which the subject is involved in more than one study, because treatments from the different studies could interact [61]. In our research design, all subjects were exposed to both treatments. However, to strengthen the construct validity of our study, we created four workbooks in which we varied the first treatment taken and the case with which the subjects worked within the treatment. Another threat is related to the experimenters' expectations [61]. Experimenters can potentially influence the results of a study, both consciously and unconsciously, based on what they expect from the experiment [69]. Therefore, the workbooks were graded by two independent evaluators who were not aware of the hypotheses. Their scoring was based on a previously prepared list of correct answers for the two cases.

4.2 Implications for research

Our paper contributes to two middle-range theories used in the requirements engineering field, namely the cognitive theory of multimedia learning and the cognitive fit theory.

We contribute to the cognitive theory of multimedia learning with a discussion on why we could not support H1 and H3. In H1, we predicted that the textual representation format should perform better because the visual one should generate extraneous cognitive load. Our statistical analysis showed that the identification of user stories from either of the formats is approximately equally effective. This might be because the models used in our experiment

were just selected chunks from real BPMN models, which lowered the cognitive effort of working with them. This poses a limitation to our research and requires investigation in future work. Next, in H3, we predicted that the understanding of integration dependencies is better with a visual representation format rather than a textual format because the visual one uses graphics that more readily reveal containment, while the textual one just uses text. The visual cues have to be structured in a way that is adjusted for the chosen problem-solving task. Our study showed that the process structure might not be the most effective way to present integration dependencies of user stories.

Our findings emphasize an important insight from the cognitive fit theory [53]: if tasks differ in their characteristics, we observe differences in benefits of representation formats. This observation, however, comes with a caveat: many tasks do not show pure affinity with visual or textual representations, as we have seen regarding H1 and H3. Future research can build on our theoretical argument and experimental design to study the suitability of visual versus textual representation formats [56, 70]. Such studies should carefully analyse the differences of tasks and their characteristics in the context of user story elicitation to identify which of them might benefit best from which representations. There is no silver bullet representation that always supports the full spectrum of diverse tasks. Business process models can be helpful regarding the understanding of execution order and integration dependencies. They may, however, not be sufficient for deriving good user stories especially since the benefit of execution of each activity is not explicitly stated in the process models.

4.3 Implications for practice

The implications for research can be readily translated into implications for practice. In line with Dumas et al. [42], it appears desirable to make more extensive use of existing business process models in different activities in user story elicitation, namely identification and understanding of user stories. Textual use case models can be readily available to support this task before a development approach is selected [7]. Also, BPMN models can be readily available because companies tend to model their business processes for different reasons [71].

While textual use case models are widely used in agile development, BPMN models are not. Our findings could promote a broader uptake of BPMN models to support user story elicitation. BPMN process models support understanding of execution order dependencies better than their informationally equivalent textual use case models. We generalized this finding to visual and textual representation formats of business process models: visual business process models offer better support for understanding execution order dependencies. This finding applies to BPMN notation because it offers a variety of elements which help to describe the process's alternative paths better. Therefore, when a development team needs to develop an application which is process oriented, they should be able to understand execution order dependencies between user stories better when they use available BPMN models. By understanding the dependencies better, they should understand the customer's business better.

For the identification of user stories and understanding of integration dependencies, the visual format apparently provides no disadvantages as compared to more established techniques that build on textual use case descriptions, as proposed by Ambler [7] and Leffingwell [2].

However, if BPMN models are available in a customer's repository, they can be useful for the identification and understanding of integration dependencies.

Both BPMN and use case models have their specific strengths. BPMN models offer an orientation beyond software development projects. They can be a valuable source of information about where in the process the new software will be used. Depending on the extent of previous BPMN modelling projects, the information about how existing applications interact with the activities and among each other can also be available. Textual use case models do not provide direct support for such a larger overview.

5 Conclusion

In this paper, we investigated the research question how effective visual BPMN models are for agile development with user stories in comparison to well-established textual use case models. We found that different representation formats are suitable for different tasks of the elicitation process. Our subjects achieved better results in problem-solving tasks about execution order dependencies when the visual BPMN model was provided. On the other hand, our subjects performed about equally well in the problem-solving task on integration dependencies and in identifying user stories whether there was a visual BPMN or textual use case model.

Our findings highlight the potential to better integrate existing documentation of BPMN models into an agile development process, as well as the importance of a visual format for managing execution order dependencies. With these findings, we contribute towards a better understanding of how the elicitation process can be supported and how the existing BPMN documentation can be reused for agile development. The findings are relevant to project stakeholders who want to identify user stories and understand their context by using business process models.

Further research is needed to better understand the advantages and limitations of the reuse of existing documentation in the context of user story elicitation. A good starting point for future work is to develop methods and tools to support or even automate the identification of user stories from existing business process models in both visual and textual formats. Another research question in this regard is how to organize user story elicitation with BPMN models for large projects. Consequently, empirical research is needed that evaluates the impact of those methods on the elicitation of user stories. In this respect, special attention should be paid to study not just the identification and understanding of the user stories but also the business benefits acquired from the whole process. This aspect was recently highlighted as a limitation of user stories and suggested job stories as an alternative that embeds goal-oriented principles by emphasizing customer situations [72]. On the other hand, an analysis of the BPM field [73] emphasizes that BPM should move beyond investigating how to model the processes and improve their understandability towards also more explicitly study their business benefits. Such synergy of those two fields could help to get closer to the ultimate goal of both fields: to improve processes and support them with the proper software application.

Finally, a limitation of our research design is that it does not offer generalization to a population of specialist requirements engineers. We suggest future research to improve understanding of how different conceptual models can influence the performance of different user story elicitation tasks, not just the identification and understanding of user stories. For example: possibly the validation of user stories, identified from business process models, is more difficult because of the absence of the <benefit> part of the user story template. The impact of this aspect needs to be considered in future research.

6 References

- [1] Lucassen, G., Dalpiaz, F., Martijn EM van der Werf, J., and Brinkkemper, S.: Improving agile requirements: the quality user story framework and tool. *Requirements engineering*, 21 (3), 2016, pp. 383-403. DOI:10.1007/s00766-016-0250-x.
- [2] Leffingwell, D.: *Agile software requirements: lean requirements practices for teams, programs, and the enterprise*. Addison-Wesley, Boston, USA, 2011. ISBN: 9780321635846.
- [3] Cohn, M.: *User stories applied for agile software development*. Addison-Wesley, Boston, USA, 2004. ISBN: 978-0321205681.
- [4] Kulshreshtha, V., Boardman, J., and Verma, D.: The emergence of requirements networks: the case for requirements inter-dependencies. *International journal of computer applications in technology*, 45 (1), 2012, pp. 28-41. DOI:10.1504/IJCAT.2012.050130.
- [5] Martakis, A. and Daneva, M.: Handling requirements dependencies in agile projects: a focus group with agile software development practitioners. *International conference on research challenges in information science*, Paris, 2013. DOI:10.1109/RCIS.2013.6577679.
- [6] Ali Jabbari Sabegh, M. and Recker, J.: Combined use of conceptual models in practice: an exploratory study. *Journal of database management*, 28 (2), 2017, pp. 33. DOI:10.4018/JDM.2017040103.
- [7] Ambler, S.: *Agile modeling: effective practices for extreme programming and the Unified Process*. Wiley computer publishing, New York, 2002. ISBN: 0-471-20282-7.
- [8] Trkman, M., Mendling, J., and Krisper, M.: The use of business process models for better understanding of integration and execution order dependencies of user stories. *Information and software technology*, 71 (March2016), 2016, pp. 58-76. DOI:10.1016/j.infsof.2015.10.006.
- [9] Saghafi, A. and Wand, Y.: Do ontological guidelines improve understandability of conceptual models? A meta-analysis of empirical work. *Hawaii international conference on system sciences*, Waikoloa, HI 2014. DOI:10.1109/HICSS.2014.567
- [10] Mendling, J., Strembeck, M., and Recker, J.: Factors of process model comprehension - findings from a series of experiments. *Decision support systems*, 53 (1), 2012, pp. 195-206. DOI:10.1016/j.dss.2011.12.013.
- [11] Ottensooser, A., Fekete, A., Reijers, H.A., Mendling, J., and Menictas, C.: Making sense of business process descriptions: an experimental comparison of graphical and textual notations. *Journal of systems and software*, 85 (3), 2012, pp. 596-606. DOI:10.1016/j.jss.2011.09.023.
- [12] Krogstie, J., Lothe, A., and Brinkkemper, O.S.: *Conceptual modelling in information systems engineering*. Springer, 2007. ISBN: 978-3-540-72676-0.

- [13] Insfrán, E., Pastor, O., and Wieringa, R.: Requirements engineering-based conceptual modelling. *Requirements engineering*, 7 (2), 2002, pp. 61-72. DOI:10.1007/s007660200005.
- [14] Recker, J., Rosemann, M., Green, P., and Indulska, M.: Do ontological deficiencies in modeling grammars matter? *MIS Quarterly*, 35 (1), 2011, pp. 57-79. DOI:10.2307/23043489
- [15] Sandkuhl, K., Fill, H.G., Hoppenbrouwers, S., Krogstie, J., Leue, A., Matthes, F., Andreas, L., OpdahlGerhard, S., Ömer, U., and Winter, R.: Enterprise modelling for the masses—from elitist discipline to common practice. *IFIP Working Conference on The Practice of Enterprise Modeling* Cham, 2016. DOI:10.1007/978-3-319-48393-1_16.
- [16] Giraldo, F.D., España, S., Giraldo, W.J., Pastor, Ó., and Krogstie, J.: A method to evaluate quality of modelling languages based on the Zachman reference taxonomy. *Software quality journal*, (In press), 2019, pp. 1-31. DOI:10.1007/s11219-018-9434-6.
- [17] Abrahão, S., Insfran, E., Carsí, J.A., and Genero, M.: Evaluating requirements modeling methods based on user perceptions: a family of experiments. *Information sciences*, 181 (16), 2011, pp. 3356-3378. DOI:10.1016/j.ins.2011.04.005.
- [18] Giraldo, F.D., España, S., Giraldo, W.J., and Pastor, Ó.: Evaluating the quality of a set of modelling languages used in combination: A method and a tool. *Information systems*, 77 (September 2018), 2018, pp. 48-70. DOI:10.1016/j.is.2018.06.002.
- [19] Reijers, H.A. and Mendling, J.: A study into the factors that influence the understandability of business process models. *IEEE transactions on systems, man and cybernetics*, 41 (3), 2011, pp. 449-462. DOI:10.1109/tsmca.2010.2087017.
- [20] Aguilar-Savén, R.S.: Business process modelling: review and framework. *International journal of production economics*, 90 (2), 2004, pp. 129-149. DOI:10.1016/S0925-5273(03)00102-6.
- [21] Mendling, J. and Strembeck, M.: Influence factors of understanding business process models. *Business information systems*, Berlin, 2008. DOI:10.1007/978-3-540-79396-0_13.
- [22] Rocha Silva, T., Winckler, M., and Bach, C.: Evaluating the usage of predefined interactive behaviors for writing user stories: an empirical study with potential product owners. *Cognition, technology & work*, (In press.), 2019, pp. 1-22. DOI:10.1007/s10111-019-00566-3.
- [23] Dumas, M., La Rosa, M., Mendling, J., Mäesalu, R., Reijers, H.A., and Semenenko, N.: Understanding business process models: the costs and benefits of structuredness. *Advanced information systems engineering* 2012. DOI:10.1007/978-3-642-31095-9_3.
- [24] Bernárdez, B., Durán, A., and Genero, M.: Empirical evaluation and review of a metrics-based approach for use case verification. *Journal of research and practice in information technology* 36 (4), 2004, pp. 246-258. DOI:10.1206/22613961049.
- [25] Anaya, V., Berio, G., Harzallah, M., Heymans, P., Matulevičius, R., Opdahl, A.L., Panetto, H., and JoseVerdecho, M.: The unified enterprise modelling language—overview and further work. *Computers in industry*, 61 (10), 2010, pp. 99-111. DOI:10.1016/j.compind.2009.10.013.
- [26] Fernández-Sáez, A.M., Genero, M., Caivano, D., and Chaudron, M.R.V.: Does the level of detail of UML diagrams affect the maintainability of source code?: a family of experiments. *Empirical software engineering*, 21 (1), 2016, pp. 212-259. DOI:10.1007/s10664-014-9354-4.

- [27] Kummer, T.F., Recker, J., and Mendling, J.: Enhancing understandability of process models through cultural-dependent color adjustments. *Decision support systems*, 87 (July 2016), 2016, pp. 1-12. DOI:10.1016/j.dss.2016.04.004.
- [28] Scanniello, G., Gravino, C., Genero, M., Cruz-Lemus, J.A., Tortora, G., Risi, M., and Doderio, G.: Do software models based on the UML aid in source-code comprehensibility? Aggregating evidence from 12 controlled experiments. *Empirical software engineering*, 23 (5), 2018, pp. 2695–2733. DOI:10.1007/s10664-017-9591-4.
- [29] Lin, J., Yu, H., Shen, Z., and Miao, C.: Using goal net to model user stories in agile software development. *International conference on software engineering, artificial intelligence, networking and parallel/distributed computing*, Las Vegas, NV 2014. DOI:10.1109/SNPD.2014.6888731
- [30] Wautelet, Y., Heng, S., Kolp, M., Mirbel, I., and Poelmans, S.: Building a rationale diagram for evaluating user story sets *International conference on research challenges in information science* Grenoble, 2016. DOI:10.1109/RCIS.2016.7549299
- [31] Mishra, D., Mishra, A., and Yazici, A.: Successful requirement elicitation by combining requirement engineering techniques. *First international conference on the applications of digital information and web technologies* Ostrava, Czech Republic 2008. DOI:10.1109/icadiwt.2008.4664355.
- [32] Kassel, N. and Malloy, B.A.: An approach to automate requirements elicitation and specification. *International conference on software engineering and application*, Dubai, UAE, 2003.
- [33] Nuseibeh, B. and Easterbrook, S.: Requirements engineering: a roadmap. *Conference on the future of software engineering* Limerick, Ireland, 2000. DOI:10.1145/336512.336523.
- [34] Lucassen, G., Dalpiaz, F., Martijn EM van der Werf, J., and Brinkkemper, S.: The use and effectiveness of user stories in practice. *Requirements engineering: foundation for software quality*, Cham, 2016. DOI:10.1007/978-3-319-30282-9_14.
- [35] Yves, W., Heng, S., Kolp, M., and Mirbel, I.: Unifying and extending user story models. *International conference on advanced information systems engineering*, Cham, 2014. DOI:10.1007/978-3-319-07881-6_15.
- [36] Yves, W., Heng, S., Hintea, D., Kolp, M., and Poelmans, S.: Bridging user story sets with the use case model. *International conference on conceptual modeling*, Cham, 2016. DOI:10.1007/978-3-319-47717-6_11.
- [37] Garg, N., Agarwal, P., and Khan, S.: Recent advancements in requirement elicitation and prioritization techniques. *International conference on advances in computer engineering and applications*, Ghaziabad, India, 2015. DOI:10.1109/icacea.2015.7164702.
- [38] Hadar, I., Soffer, P., and Kenzi, K.: The role of domain knowledge in requirements elicitation via interviews: an exploratory study *Requirements engineering*, 19 (2), 2014, pp. 143-159. DOI:10.1007/s00766-012-0163-2.
- [39] Davis, A., Dieste, O., Hickey, A., Juristo, N., and Moreno, A.M.: Effectiveness of requirements elicitation techniques: empirical results derived from a systematic review. *14th IEEE International requirements engineering conference*, Washington, 2006. DOI:10.1109/RE.2006.17
- [40] Maiden, N. and Rugg, G.: ACRE: selecting methods for requirements acquisition. *Software engineering journal*, 11 (3), 1996, pp. 183-192. DOI:10.1049/sej.1996.0024.
- [41] Nolte, A., Bernhard, E., Recker, J., Pittke, F., and Mendling, J.: Repeated use of process models: the impact of artifact, technological, and individual factors. *Decision support systems*, 88 (2016), 2016, pp. 98-111. DOI:10.1016/j.dss.2016.06.002.

- [42] Dumas, M., La Rosa, M., Mendling, J., and Reijers, H.: Fundamentals of business process management. Springer-Verlag Berlin Heidelberg, Berlin, Germany, 2013. ISBN: 978-3-642-33142-8.
- [43] vom Brocke, J. and Mendling, J.: Business process management cases. Digital innovation and business transformation in practice. Springer, Cham, 2017. ISBN: 978-3-319-58307-5.
- [44] Yan, Z., Dijkman, R., and Grefen, P.: Business process model repositories – framework and survey. *Information and software technology*, 54 (4), 2013, pp. 380-395. DOI:10.1016/j.infsof.2011.11.005.
- [45] Weber, B., Reichert, M., Mendling, J., and Reijers, H.: Refactoring large process model repositories. *Computers in industry*, 62 (5), 2011, pp. 467-486. DOI:10.1016/j.compind.2010.12.012.
- [46] La Rosa, M., Dumas, M., Uba, R., and Dijkman, R.: Business process model merging: an approach to business process consolidation. *ACM transactions on software engineering and methodology*, 22 (2), 2013, pp. 1-42. DOI:10.1145/2430545.2430547.
- [47] Jurič, M.B. and Pant, K.: Business process driven SOA using BPMN and BPEL. Packt Publishing, Birmingham, 2008. ISBN: 978-184719-146-5.
- [48] Cockburn, A.: Writing effective use cases. Addison-Wesley, Upper Saddle River, NJ, 2001. ISBN: 978-0201702255.
- [49] Milicic, A., Perdikakis, A., El Kadiri, S., Kiritsis, D., and Ivanov, P.: Toward the definition of domain concepts and knowledge through the application of the user story mapping method *International journal of product lifecycle management*, 7 (1), 2014, pp. 3-16. DOI:10.1504/IJPLM.2014.065456.
- [50] Patton, J.: User story mapping: discover the whole story, build the right product. O'Reilly, 2014. ISBN: 978-1-491-90490-9.
- [51] Sweller, J.: Cognitive technology: some procedures for facilitating learning and problem solving in mathematics and science. *Journal of educational psychology*, 81 (4), 1989, pp. 457-466. DOI:10.1037/0022-0663.81.4.457.
- [52] Larkin, J.H. and Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, 11 (1), 1987, pp. 65-100. DOI:10.1016/S0364-0213(87)80026-5.
- [53] Vessey, I.: Cognitive fit: a theory-based analysis of the graphs versus tables literature. *Decision sciences*, 22 (2), 1991, pp. 219-240. DOI:10.1111/j.1540-5915.1991.tb00344.x.
- [54] Vessey, I. and Galletta, D.: Cognitive fit: an empirical study of information acquisition. *Information systems research*, 2 (1), 1991, pp. 63-84. DOI:10.1287/isre.2.1.63.
- [55] Agarwal, R., Sinha, A.P., and Tanniru, M.: Cognitive fit in requirements modeling: a study of object and process methodologies *Journal of management information systems* 13 (2), 1996, pp. 137-162. DOI:40398219
- [56] Mayer, R.E.: Multi-media learning, Second edition. Cambridge university press, USA, 2009. ISBN: 9780521735353.
- [57] Leopold, H., Mendling, J., and Gunther, O.: Learning from quality issues of BPMN models from industry *IEEE software*, 33 (4), 2016, pp. 1-9. DOI:10.1109/MS.2015.81
- [58] Wieringa, R.J.: Design science methodology for information systems and software engineering. Springer, Heidelberg, 2014. ISBN: 9783662438398.
- [59] Juristo, N. and Moreno, A.M.: Basics of software engineering experimentation. Springer, 2001. ISBN: 978-0-7923-7990-4.

- [60] Vegas, S., Apa, C., and Juristo, N.: Crossover designs in software engineering experiments: benefits and perils. *IEEE transactions on software engineering*, 42 (2), 2016, pp. 120-135. DOI:10.1109/TSE.2015.2467378.
- [61] Wohlin, C., Runeson, P., Host, M., Ohlsson, M.C., Regnell, B., and Wesslen, A.: *Experimentation in software engineering, an introduction*. Springer, Boston, 2000. ISBN: 978-3-642-29043-5.
- [62] Gemino, A. and Wand, Y.: A framework for empirical evaluation of conceptual modeling techniques. *Requirements engineering*, 9 (4), 2004, pp. 248-260. DOI:10.1007/s00766-004-0204-6.
- [63] Mendling, J., Recker, J., Reijers, H.A., and Leopold, H.: An empirical review of the connection between model viewer characteristics and the comprehension of conceptual process models. *Information systems frontiers*, (In press.), 2018, pp. 1-25. DOI:10.1007/s10796-017-9823-6.
- [64] Field, A.: *Discovering statistics using SPSS*. SAGE publications, Canada, 2009. ISBN: 987-1-4462-4917-8.
- [65] Batra, D., Hoffler, J.A., and Bostrom, R.P.: Comparing representations with relational and EER models. *Communications of the ACM*, 33 (2), 1990, pp. 126-139. DOI:10.1145/75577.75579
- [66] Compeau, D., Marcolin, B., Kelley, H., and Higgings, C.: Research commentary—generalizability of information systems research using student subjects—a reflection on our practices and recommendations for future research. *Information systems research*, 23 (4), 2012, pp. 1093-1109. DOI:10.1287/isre.1120.0423
- [67] Narayanan, A. and Moritz, B.B.: Decision making and cognition in multi-echelon supply chains: an experimental study. *Production and operations management*, 24 (8), 2015, pp. 1216-1234. DOI:10.1111/poms.12343.
- [68] Croson, R. and Donohue, K.: Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management science*, 52 (3), 2006, pp. 323-336. DOI:10.1287/mnsc.1050.0436.
- [69] Rosenthal, R.: *Experimenter effects in behavioral research*. Appleton-Century-Crofts, East Norwalk, CT, US, 1966. ISBN: 978-0829008760.
- [70] Gregor, S.: The nature of theory in information systems. *MIS Quarterly*, 30 (3), 2006, pp. 491-506. DOI:10.2307/25148742
- [71] Kovačič, A. and Bosilj Vukšič, V.: *Management poslovnih procesov: Prenova in informatizacija poslovanja*. GV Založba, Ljubljana, 2005. ISBN: 86-7061-390-5.
- [72] Lucassen, G., van de Keuken, M., Dalpiaz, F., Brinkkemper, S., Sloof, G.W., and Schlingmann, J.: Jobs-to-be-done oriented requirements engineering: a method for defining job stories (a chapter in book *Requirements engineering: foundation for software quality*). 2018. DOI:10.1007/978-3-319-77243-1_14
- [73] Klun, M. and Trkman, P.: Business process management – at the crossroads. *Business process management journal*, 24 (3), 2018, pp. 786-813. DOI:10.1108/BPMJ-11-2016-0226

Appendices

Appendix A

Part of the original BPMN model taken from bank's repository.

Appendix B

UC model which is informationally equivalent to BPMN model in Appendix A

<p>Use case name: CLOSE A BANK ACCOUNT</p> <p>Basic flow of events</p> <ol style="list-style-type: none">1. ... Refer to the following alternative flows: A, B.2. Clerk in the front office writes a notification letter about closure of the account to the customer. <p>Alternative flows of events</p> <p>Alt.flow A:</p> <ol style="list-style-type: none">1. Clerk in the front office fills in a request for cancelling authorizations.2. Clerk in DEPT 2 deletes all the permissions. <p>Alt.flow B:</p> <ol style="list-style-type: none">1. Clerk in the front office fills in a form to close e-banking2. Clerk in DEPT 3 closes it. <p>Additional information:</p> <p>After the clerk in DEPT 1 changes the status of customer's account to "passive" the clerk in the front office needs to close all open businesses, such as e-banking and authorizations. According to the characteristics of the account none, one (A or B) or two alternative flows (A and B) are triggered. No matter how many alternative flows are triggered, clerk in the front office always writes sends an email about the closure to the DEPT 4 who needs to send him/her all of the account's physical documents. The event (step 2 of the basic flow) can start after opened authorizations are closed, opened e-banking is closed and all the account's physical documents are handed to the clerk in the front office.</p>
--

Appendix C

Workbook A is available at:

<https://www.dropbox.com/s/8u1mwc14m99jxtl/Appendix%20C%20-%20Workbook%20A%20-%20no%20identifying%20data.pdf?dl=0>

Appendix D

Workbook B is available at:

<https://www.dropbox.com/s/zzkn79a83qp5txc/Appendix%20D%20-%20Workbook%20B%20-%20no%20identifying%20data.pdf?dl=0>

Appendix E

<https://www.dropbox.com/s/528azr8gkzvcjhg/experiments%20educational%20material%20-%20BPMN%20then%20UC%20-%20no%20identifying%20data.pptx?dl=0>