



HAL
open science

Evaluating the Usage of Predefined Interactive Behaviors for Writing User Stories: An Empirical Study with Potential Product Owners

Thiago Rocha Silva, Marco Winckler, Cédric Bach

► To cite this version:

Thiago Rocha Silva, Marco Winckler, Cédric Bach. Evaluating the Usage of Predefined Interactive Behaviors for Writing User Stories: An Empirical Study with Potential Product Owners. *Cognition, Technology and Work*, 2019, 10.1007/s10111-019-00566-3 . hal-02879291

HAL Id: hal-02879291

<https://hal.science/hal-02879291>

Submitted on 23 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluating the Usage of Predefined Interactive Behaviors for Writing User Stories: An Empirical Study with Potential Product Owners

Thiago Rocha Silva^{1,2}, Marco Winckler^{1,3}, Cédric Bach⁴

¹ *ICS-IRIT, Université Toulouse III - Paul Sabatier, Toulouse, France*

² *Department of Computer Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*

³ *SPARKS-i3S, Université Nice Sophia Antipolis (Polytech), Sophia Antipolis, France*

⁴ *Human Design Group, HDG, Toulouse, France*

NTNU, Department of Computer Science, NO-7491 Trondheim, Norway,

thiago.silva@ntnu.no

Evaluating the Usage of Predefined Interactive Behaviors for Writing User Stories: An Empirical Study with Potential Product Owners

Behavior-Driven Development (BDD) has gained popularity in agile software development as a means of specifying user requirements through User Stories and scenarios for interactive systems under construction. Templates for guiding the writing of such stories are widely employed and can be helpful to ensure that consistent information about the requirements is provided. With the aim of getting preliminary results about how Product Owners (POs) write their own User Stories under a predefined template, we conducted a case study with potential POs at the department in charge of business trips in a French research institute. The participants were invited to write their own User Stories to describe a feature they are used to perform. The resultant stories have been analyzed to check their adherence to a template including common interactive behaviors defined by an ontology for interactive systems. Although the participants have followed different specification strategies, we observed an overall high level of adherence to the proposed template (62,26%). The results also pointed out to a wide use of domain-dependent behaviors, with the interactive behaviors defined by the ontology being, to some extent, reproduced by the participants even without prior training in the adopted vocabulary.

Keywords: Behavior-Driven Development (BDD); User Stories; Agile Software Development; User Requirements.

1. Introduction

Understanding user requirements is critical to the success of interactive systems (Maguire and Bevan, 2002). However, this is not a simple task, especially when different users and stakeholders may have diverse requirements, opposed points of view, and/or conflicting priorities about a system, which might lead to organizational dissonances (i.e. *cindynics*) (Vanderhaegen and Carsten, 2017). For example, a requirement for a quality manager would be a strict workflow to be followed for ensuring that users will see all the steps, while the users would require flexibility and ask for shortcuts and alternative flows.

The success of a software project will not only require conflict resolution but also cohesive view of requirements (Ambler, 2002). That is a tricky problem because users and stakeholders might express requirements with diverse degree of formality. Informal statements expressed in natural language are more familiar for users, but the lack of formality might lead to verbose, incomplete, and ambiguous descriptions that are difficult to assess. Requirements expressed in a formal language require technical skills and make the communication of requirements with developers, stakeholders, clients and users, more difficult to follow. Such as variability of requirements specifications recalls the *anamorphosis* described by Vanderhaegen and Carsten (2017) in their *dissonances taxonomy*.

Chelimsky *et al.* (2010) proposed a semi-structured method called Behavior-Driven Development (BDD) that supports the specification and test of user requirements expressed in natural language. For that, BDD relies on a textual template that allows to specify requirements in the form of User Stories (Cohn, 2004), which are argued to be easily understandable for both technical and non-technical stakeholders. User Stories might describe “executable requirements”, i.e. requirements that can be directly tested from their textual specifications. Executable requirements contain the specification of the requirement and the scenario that can be automated by a test aimed at certifying whether (or not) the requirement is supported by the system. BDD also encompasses a scenario-based approach that benefit from an iterative cycle of producing-evaluating test scenarios in a final and implemented version of the system.

The BDD support for both test automation and system design has aroused interest from both academic and industrial communities. However, the BDD method does not offer a consistent vocabulary for the specification of interactive user’s tasks. In a previous work (Silva, Hak and Winckler, 2017a, 2017b), we have proposed an ontology describing

events, behaviors and user interface elements that could be used to promote consistency of requirements. We also have demonstrated that User Story written using terms of that ontology can be consistently employed for testing automation of diverse types of artifacts, including task models, low-fidelity prototypes as well as the final implementation of the interactive system. By running such automated tests over diverse artifacts, we can check the consistence between the specification of the system and its final implementation.

In this paper, we investigate in which extent non-technical stakeholders are able to specify their own User Stories using common behaviors described in our ontology. The benefits of non-technical stakeholders' involvement in requirements specification are largely recognized in the literature (Bano and Zowghi, 2013), including reducing requirements misunderstandings, besides providing faster feedback and more accurate acceptance conditions. For that, we have run an empirical study where we have invited a group of potential Product Owners (POs) to write their own User Story using a predefined template and the set of interactive behaviors provided by our ontology. We then analyze the User Stories according to: *(i)* adherence to a particular template, and *(ii)* compatibility with terms of the ontology. We also categorize each deviation from the proposed template committed by the participants. We suggest that User Stories written with our ontology can address multiple concerns related to the involvement of non-technical people in requirements specification activities, since requirements and acceptance testing can be specified in a natural and high-level language while keeping it consistent and aligned with the resultant software artifacts.

The next section of this article presents the background and the related works. Then, we present the research questions, procedures and methods employed in the study. Section 4 presents a case study, the profile of participants that took part in the study, the business context and the exercise proposed to the participants. Section 5 presents the

results, highlighting the set of User Stories written by the participants, the adherence analyses considering stories and scenarios, and our discussion of such results. Section 6 presents our general findings and implications, as well as the threats to validity of this study. Finally, the last section concludes with our final remarks and points out future investigation opportunities in this research.

2. Background and Related Works

2.1. User Stories and Scenarios

The terms User Stories and scenarios have a large meaning in the literature (Silva, Hak and Winckler, 2016a, 2016b). It is argued that the Systems and Software Engineering use of scenarios is predominately concerned with physical manifestations of systems work (MacLeod, 2008). In BDD, User Stories and scenarios are typically used in agile projects to describe user requirements in a task-oriented way. This technique was proposed by Cohn (2004) and provides in the same artifact a narrative, briefly describing a feature in the business point of view, and a set of scenarios to give details about business rules and to be used as acceptance criteria giving concrete examples about what should be tested to consider a given feature as done.

User Stories are typically written by POs or Requirements Engineers (with the assistance of POs). POs are stakeholders that master the current business process and could integrate a specialized group for specifying user requirements to maintain or develop a new software system in a given domain (Schwaber, 2004). To help on this task, different formats and templates have been proposed for User Stories over time (Wautelet *et al.*, 2014). The authors have investigated the suitability of different templates for User Stories and how they could be improved to set an agreement in their semantics and

methodological elements, which could help to improve communication between stakeholders. The most widely used template, called BDD story (North, 2019), is presented in Figure 1.

```
Title (one line describing the story)

Narrative:
As a [role]
I want [feature]
So that [benefit]

Acceptance Criteria: (presented as Scenarios)

Scenario 1: Title
Given [context]
  And [some more context]...
When [event]
Then [outcome]
  And [another outcome]...

Scenario 2: ...
```

Figure 1. Template for a BDD story (North, 2019).

A User Story is thus described with a title, a narrative, and scenarios that define the acceptance criteria. The title provides a general description of the story and it refers to the main feature. The narrative describes the main feature and the expected benefits for both individuals playing a role in the system and for the overall business. The acceptance criteria are defined through a set of scenarios, each one with a title and three main clauses: “*Given*” (provides the context in which the scenario will be actioned), “*When*” (describes events that trigger the scenario), and “*Then*” (present the outcomes, used to determine whether or not the system behaves as expected). Each one of these clauses can include an “*And*” statement to provide multiple contexts, events and/or outcomes. Each statement in this representation is called *step*.

The adoption of User Stories to specify user requirements has experienced an substantial growing in the last years (Kassab, 2015). Many authors have studied their effectiveness (Lucassen *et al.*, 2016), dependencies (Gomez, Rueda and Alarcón, 2010; Trkman, Mendling and Krisper, 2016), and quality criteria (Lucassen *et al.*, 2015).

Recent design methods rely on User Stories for reducing communication gaps. For example, Valente et al. (2016, 2017) proposes to use User Stories as a mean to reduce the gap between the specification of business processes and user tasks. Nonetheless, a few studies have investigated the effective use of User Story along the development process. Wautelet *et al.* (2018) evaluated the experience and difficulties of modelers while using a unified model to visually identify the inter-dependencies between the elements of a User Story. Lucassen *et al.* (2018) proposes an alternative to plain-vanilla User Stories that embeds goal-oriented principles by emphasizing situation, motivation and the expected outcome. The authors evaluated the method with product managers, marketers and developers. Despite these efforts, none of the works found in the literature refers to the writing of User Stories by non-technical stakeholders.

2.2. A Behavior-Based Ontology for Interactive Systems

From empirical observation, we have found that User Stories specified in natural language often contain inconsistencies. For example, it is not rare to find scenarios that specify an action such as “*click*” on text fields, or “*select*” in a button. These observations motivated to investigate the use of a formal ontology for describing pre-defined behaviors that could be used to specify scenarios that address interactions with user interfaces (UIs) (Silva, Hak and Winckler, 2017a, 2017b). This ontology was built upon patterns for low-level behaviors that are recurrent when writing BDD stories for assessing user interface design artifacts used to build in interactive systems. On one hand, the ontology acts a controlled vocabulary of terms, thus removing ambiguities. On the other hand, the semantic of ontological terms determines the way the test should be implemented for assessing the interactive systems.

In that ontology, scenarios are represented as concepts borrowed from abstract state machines. A scenario played in a given UI is represented as a transition. States are used to represent the original and resulting UIs (after a transition occurs). Transitions have at least one or more conditions (represented in scenarios by the “*Given*” clause), one or more events (represented in scenarios by the “*When*” clause), and one or more actions (represented in scenarios by the “*Then*” clause).

Interactive behaviors describe how users might interact with the system whilst manipulating graphical elements of the UI. An example of behavior specification is illustrated by Figure 2. The specification of behaviors includes two aspects: (i) when the interaction is performed (using “*Given*”, “*When*” and/or “*Then*” clauses), and (ii) which graphical elements (i.e. *CheckBoxes*, *TextFields*, *Buttons*, etc.) are affected. Altogether, behaviors and interaction elements are used to implement the test of expected system behavior. In the example of Figure 2, the behavior “*I choose ‘<value>’ referring to ‘<field>’*” has two parameters: “<value>” and “<field>”. The first parameter is associated to data, whilst the second parameter refers to the interaction element supported by this behavior: “*Radio Button*”, “*CheckBox*”, “*Calendar*” and “*Link*”.

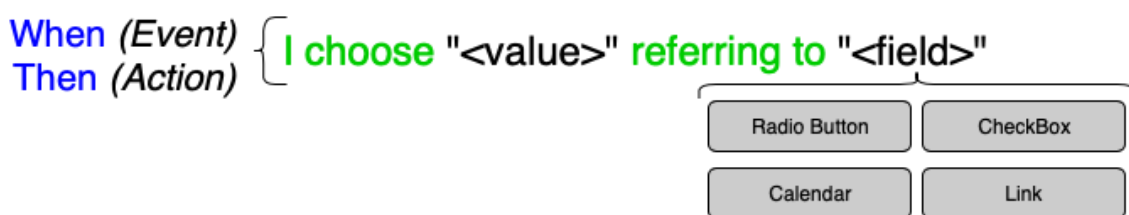


Figure 2. Structure of a behavior as specified in the ontology.

The ontological model is domain-free, which means it is not subject to particular business characteristics in the User Stories. The behaviors described by the ontology refer only to the interaction with elements of the user interface. The current version of the

ontology covers more than 60 interactive behaviors and almost 40 interaction elements for both web and mobile user interfaces. Table 1 exemplifies some of these interactive behaviors, the transition component during which they can be triggered (shadow in grey) and the set of corresponding interaction elements.

Behavior	Transition			Interaction Elements
	C	E	A	
<i>choose</i> \equiv <i>select</i>				<i>Calendar, Checkbox, Radio Button, and Link</i>
<i>chooseByIndexInTheField</i>				<i>Dropdown List</i>
<i>chooseReferringTo</i>				<i>Calendar, Checkbox, Radio Button, and Link</i>
<i>chooseTheOptionOfValueInTheField</i>				<i>Dropdown List</i>
<i>clickOn</i>				<i>Menu, Menu Item, Button, and Link</i>
<i>clickOnReferringTo</i>				<i>Menu, Menu Item, Button, and Link</i>
<i>doNotTypeAnyValueToTheField</i> \equiv <i>resetTheValueOfTheField</i>				<i>Text Field</i>
<i>goTo</i>				<i>Browser Window</i>
<i>isDisplayed</i>				<i>Browser Window</i>
<i>setInTheField</i> \equiv <i>tryToSetInTheField</i>				<i>Dropdown List, Text Field, Autocomplete, and Calendar</i>
<i>typeAndChooseInTheField</i> \equiv <i>informAndChooseInTheField</i>				<i>Autocomplete</i>
<i>willBeDisplayed</i>				<i>Text</i>

Table 1. Example of interactive behaviors described in the ontology. *Transition: (C)ontext, (E)vent, (A)ction.*

3. Methodology

3.1. Study Design

In order to investigate how non-technical stakeholders write User Stories, we have designed an empirical study around two research questions:

RQ 1. *Are participants able to read a basic User Story template and use it to write their own stories?*

To answer this research question, we measure the adherence of the User Stories produced by the participants to the structure of a User Story template.

RQ 2. *Is the vocabulary used by the participants to write their own User Stories similar to the vocabulary described in the ontology?*

To answer this research question, we compare the vocabulary used by the participants of the study with the predefined interactive behaviors modelled in the ontology.

The analysis of adherence of the scenarios was performed in two parts: the first part concerned the analysis of the User Story (*narrative* section); the in second part we analyze the related scenario (*scenario* section). For assessing the adherence, we observe possible gaps between the steps specified by the participant and the steps proposed by the ontology. The adherence was assessed using a 7-point scale ranging from null adherence (scale 0 ○○○○○○) to full adherence (scale 6 ●●●●●●).

3.2. Procedure

First of all, we have selected a case study so that we can compare the User Stories provided by participants with expertise in the same application domain. The case study selected concerns the management of business trips and it is fully described in section 4.

Participants were recruited in a convenience sample based on their expertise on the application domain of the case study (i.e. the management of business trips), availability and heterogeneity of profiles. Before each session of interviews, participants gave their written consent to take part in the study. With the agreement of participants, all the interviews were fully audio recorded.

The study was conducted as a semi-structured interview, as follows:

- Information gathering about the participants' profile, tasks and opinions about the software system in use.
- Introduction of User Stories to the participants (but not the ontology).

- Writing exercise where participants are asked to write their own User Stories.

3.2.1. Interviews

The interview focused on the participants' profile, previous experience with the management of business trips, and personal impressions about the current support systems. Six questions covered information about the profile including gender, age, education, for how long they were involved with that job, whether they had previous experiences in that kind of job before, and finally a general and open question about an overview of their job and daily activities. The details of this part are described in the next section "Participants' Profile" hereafter.

After that, we collected the participants' impressions about the system they use for booking the business trips. A total of 16 questions have been made at this second part of the interview concerning both factual and interpretation points. They were asked about how booking demands are processed and threatened along a life cycle in the travel department, and about their personal opinion regarding constraints and improvement opportunities in the current workflow as well as about the current system they use daily for processing the booking requests.

3.2.2. Introduction to User Stories

Prior to the exercise of writing the User Stories, participants were introduced to the structure and to the main components of a typical User Story based on the BDD format proposed by North (2019) as shown in Figure 1. Then, an example of User Story describing a searching feature of a one-way flight for a general business trip has been presented. Both the template and the example were given in the native language of the participants, which means French. The equivalent/translated English version of the User

Story template is given in Figure 1 (Section 2.1). The equivalent/translated English version of the example of a User Story is shown in Figure 3. It concerns a search for a one-way ticket for a business trip.

```
Title: Flight Tickets Search  
  
Narrative:  
As a frequent traveler  
I want to be able to search tickets, providing locations and dates  
So that I can obtain information about rates and times of the flights.  
  
Scenario: One-Way Tickets Search  
Given I go to the page "Find flights"  
When I choose "One way"  
And I type "Paris" and choose "Paris, Charles de Gaulle (CDG)" in the field "From"  
And I type "Toulouse" and choose "Toulouse, Blagnac (TLS)" in the field "To"  
And I choose "2" in the field "Total number of passengers"  
And I choose "12/15/2016" in the field "Depart"  
And I click on "Search"  
Then will be displayed the list of available flights
```

Figure 3. Example of a User Story presented to the participants translated to English.

3.2.3. The Exercise

For this exercise, participants were asked to write on paper a single User Story with one scenario for describing a feature they have faced recently when using the current software system for booking the business trips. The participants had no access to the system and had to remember themselves scenarios they have recently experienced. We did not mention the existence of common and predefined interactive behaviors in the ontology which were supposed to be used for writing the stories, although the example of User Story we provided to them had been written following such behaviors presented in the ontology. This decision was made because one of the goals of this study was to investigate the ability of non-technical core POs to specify their own User Stories and in which extent the interactive behaviors described in the ontology would be perceived as useful enough to be reproduced by the participants.

After getting the User Stories produced by the participants, the results were analyzed firstly by one of the authors (who conducted the interviews) and then crosschecked by an independent reviewer. Consolidated results and analyses are presented in Section 5. We have categorized each deviation from the proposed template committed by the participants when writing their User Stories. Such categories have been defined as *adherence problems* and have been classified under the Meyer's seven sins (Meyer, 1985). They are described as follows.

- **Lack of statement or keyword** (*Silence*), refers to clause or keyword present in the template, and not used by the participant.
- **Understatement** (*Silence*), refers to statements/behaviors specified following the structure presented in the template, but with less information than necessary.
- **Misspecification** (*Noise*), refers to statements/behaviors that have been misspecified according to the structure defined in the ontology.
- **Wrong information** (*Contradiction*), refers to statements which states a correct template structure, but presents wrong (or partially wrong) information for that statement.
- **Minor writing complement** (*Silence*), refers to the need of minor complements (or modifications) in the phrase in order to comply with the template structure or clarify the behavior's meaning.
- **High-level of abstraction** (*Wishful Thinking*), refers to behaviors specified in such a high-level of abstraction which not allow to assume the actual expected interaction on the UI.
- **Epic behavior** (*Overspecification*), refers to behaviors that encompass a wide number of implicit interactions. This kind of behavior should typically be broken

into several low-level interactive behaviors. The concept of *epics* has been introduced by Cohn (2004) and it refers to a large User Story that cannot be delivered as defined within a single iteration or is large enough that it can be split into smaller User Stories.

4. The Case Study

The case study concerns the management of business trips. It was settled at the travel department of the Toulouse Institute of Computer Science Research (IRIT), Toulouse, France, which handle multiple business trip requests from over 700 researchers.

The management of business trips for researchers is supported by two systems. The first one is named booking system, which is used by researchers for searching their flights and getting a quotation of rates for a given itinerary. The second one is a help desk system used to mediate the communication of booking requests between researchers and travel accounting departments. In short, the help desk system allows to approve or refuse travel quotations based on the budget. Both systems are currently in operation. The case study focuses on the researchers' request for booking a business trip. Figure 4 presents a screenshot of this system which both researchers and the department team have access.

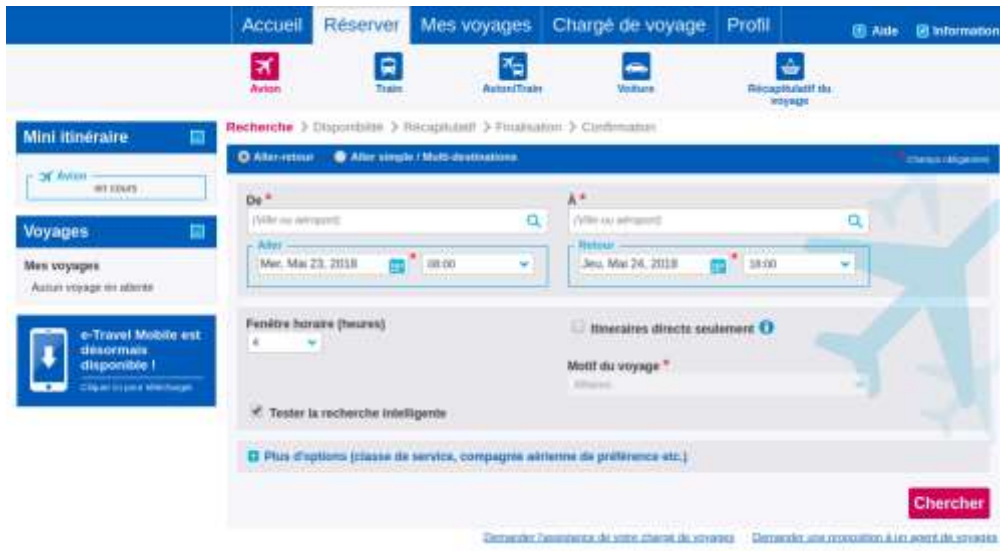


Figure 4. The system for booking business trips.

The overall booking process is illustrated by Figure 5 through a BPMN model (Business Process Model And Notation™ (BPMN™), 2011) is based on descriptions collected by the interviews. The process starts with a researcher searching for a flight. Search parameters (such as departure and arrival cities or airports, date of departure and return, timeframe, etc.) are processed by the system that creates a list of matching flights. The researcher makes a choice according to the list of available flights. When he/she confirms his/her choice, passenger and flight data are saved by the system and the booking is put on hold. At this point, the researcher needs to create a ticket in the help desk management system in order to request the authorization of payment by the travel department. When the ticket is created, the travel department team process the payment, checking whether the research has enough budget for the trip. If the budget is sufficient, the payment is accepted, and the travel department team authorizes the booking. Approved trips are processed, and researcher receives his/her electronic ticket. For this case study, we focused on the sub process highlighted in red in Figure 5.

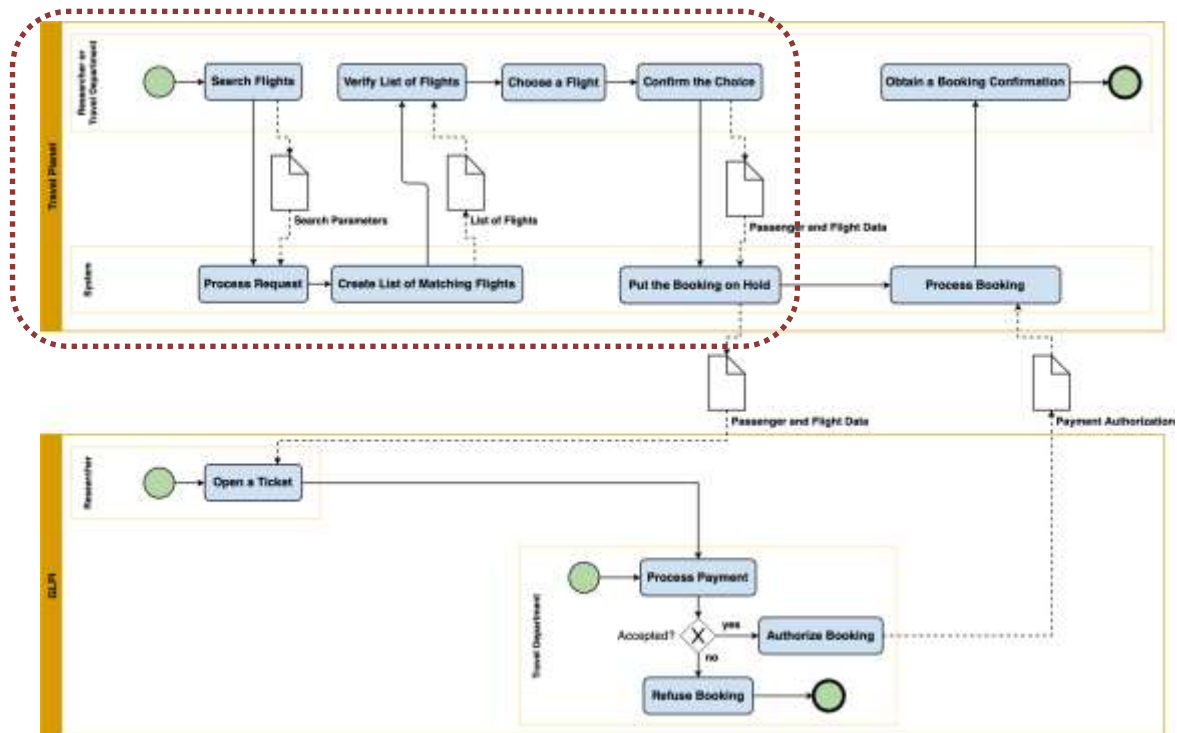


Figure 5. BPMN model for the case study.

4.1. Participants' Profile

The four participants who took part in the case study are users of the applications described in the case study and representative of POs in the domain of management of business trips. Table 2 summarizes the participants' profile. We notice therein a homogeneity in their level of education, with P3 and P4 having completed only the secondary level, while P1 has completed one year more of undergraduate studies, and P2 two years more. We notice as well P1 is the most experienced participant with almost twice the experience of the other 3 participants. Although P1, P2 and P3 have also the same level of seniority at this charge in the institute (about 4 years in average), P4 had been hired only 1 month prior to this study, so his/her participation was interesting to compare his/her view with possible work habits acquired by the older employees. We had a predominance of women (3 of 4 participants) with a good range of ages, from 25 to 52. Finally, the participants had no knowledge about agile methods nor agile requirements.

Participant	Gender	Age	Education	Experience (Years in total)	Experience (Years in the institute)
P1	Female	50	SSL+1	10	4
P2	Female	30	SSL+2	6	3
P3	Female	52	SSL	4	4
P4	Male	25	SSL	4	1 month

Table 2. Participants' Profile.

5. Results

Each interview lasted about 1 hour, including the presentation of the User Stories and the writing exercise part that took in average 20 minutes. During the interview, participants highlighted they manage about 400 travel demands per year, being a means of 12 per week in off-peak seasons, and 12 per day in peak seasons. They are in general pretty satisfied about the current system's support, nonetheless they frequently need to contact the researchers asking for complementary information about the trip. As far as new features are a concern, a participant pointed out that having a list of departure times for the less expensive flight rates could be very interesting. Another participant pointed out the need of a feature to book several trips for a group in the same demand. They almost unanimously pointed out that features for searching multi-destination trips have certainly a room for improvement. In the last part of the interview, focus of this study, we captured the User Stories written by the participants.

5.1. User Stories Writing

Figure 6, 7, 8 and 9 present the translated versions of the User Stories written by the participants in their first attempts. Figure 6 brings additionally an example of the User Story handwritten by the participant P1. As stated in Section 4.2, the participants were free to choose the scenario they wanted to specify.

Preamble .

En tant que : Invités .

Je veux : des billets d'avion avec horaires et vols .
définis .

Afin de : réserver des billets .

1

Scenario: Recherche du billet demandé .

Etant donné : je vais sur le site . (SIMBAB/TRAVEL)

Quand je choisis le vol demandé

(destination et horaires)
Et : je choisis type de voyageur - "Invités"
TOULOUSE / PARIS .

ET. RECHERCHE → Retour même jour 19h00
ALORS - JE CHOSIS le vols. désirés
~~Et je choisis le voyageur et~~

2

→ Quand .. On se saisi les données concernant
le voyageur - (nom, prénom, date naissance
téléphone, email)
éventuellement les cartes de réductions
(- flying blue et Carte abonnement)

ALORS - Billet en demande de validation .

3

Narrative:

As a guest

I want airline tickets with defined time and flights

So that I can book tickets

1

Scenario: Searching demanded tickets

Given I go to the site SIMBAB/TRAVEL

When I choose the demanded flight (destination and times, TOULOUSE/PARIS, departure 7 a.m., return 7 p.m. at the same day)

And I choose type of traveler "Guest"

And I search

Then several propositions

And I choose the desired flights

2

When I inform the data concerning the traveler (name, given name, birthdate, phone, mail), and eventually the loyalty card (Flying Blue and Season Ticket)
Then ticket waiting for validation

3

Figure 6. User Story written by P1.

As a frequent traveler
I want to search for tickets, providing locations and dates for a multi-destination trip
So that I can obtain information about rates and flight times

Scenario: Multi-destination searching
Given I go to the page "Searching Flights"
When I choose "Multi-destinations"
And I type "Paris" and choose "Paris, Charles de Gaulle" in the field "Departure"
And I type "Rio de Janeiro" in the field "Destination"
And I choose "15/02/17" in the field "Departure Date"
And I choose "20/02/17" in the field "Return Date"
And I type "Rio de Janeiro" in the field "Departure"
And I type "Porto Alegre" in the field "Destination"
And I choose "17/02/17" in the field "Departure Date"
And I choose "19/02/17" in the field "Return Date"
And I click on "Search"
Then will be displayed the list of available flights

Figure 7. User Story written by P2.

As a travel manager
I want to check travel authorizations
So that I can ensure the confirmed bookings

Scenario: Listing travel authorizations
Given I go to the tab "Travel Authorization"
When I type the "Booking Reference"
And I check if the request is well registered
Then at this time, I can know for sure (or not) the request has been taken into account
And it's shown a tab: authorized / non-authorized

Figure 8. User Story written by P3.

As an intern
I want to book a flight to Paris departing on May 2nd until May 10th
So that I can attend a seminar

Scenario:
Given I'm going to book my flight
When I provide all the information
And I choose search by fares
Then all the available flights for the date are classified by ascending order of fares.

Figure 9. User Story written by P4.

The participant P1 (Figure 6) presented a User Story to describe the process of booking trips for a guest, i.e. an external person, normally a researcher from outside the institute. We notice clearly that the first participant has chosen to describe the US in a high level, free of format, not necessarily paying attention to the ontology pattern step presented in the example. Thus, each step of scenarios could be identified as domain-dependent behaviors, i.e. behaviors that make direct reference to jargons used for booking flights. In the first identified point (1 in Figure 6), we can see the user states a narrative concerning a guest searching for airline tickets with defined time and flights in order to book tickets. Here we notice as well, the user has committed a mistake when identifying the role that benefits from the story. In fact, he/she identified that the guest would be the right role for this story when indeed the account managers of the travel department would be the beneficiaries, once it is them that would perform the booking using the system on behalf of the guest.

In the second identified point (2 in Figure 6), we notice the first scenario he/she identified. The scenario specifies the use of two systems for booking business flights. It simulates in a high-level a travel from Toulouse to Paris departing at 7 a.m. and returning at 7 p.m. at the same day. So, he/she informs this trip concerns a guest and based on the submitted search, he/she chooses the desired flights. At this point (3 in Figure 6), our user mixed a second scenario with the first one. He/she continues specifying actions for informing traveler's data and putting the ticket on hold, waiting for validation.

The participant P2 (Figure 7) reported a story for booking a multi-destination trip. We notice here that, unlike the first one, the second participant has chosen to describe the User Story closely paying attention to the ontology pattern step presented in the example. Thus, each step of the scenario could be identified as domain-independent behaviors, i.e. behaviors that refers to the actions on the user interface, without mentioning jargons used

for booking flights. In the first part of the story, we can see the user states a narrative concerning a frequent traveler searching for a multi-destination ticket in order to obtain rates and flight times. In the second part, a scenario for searching return flights from Paris to Rio de Janeiro with a stopover in Porto Alegre is presented. We can see the user clearly understood the structure of the scenario, once he/she adjusted the sequence of steps to cover a multi-trip data entrance with different cities and dates.

The participant P3 (Figure 8) reported a story for checking travel authorizations. In such a story, a travel manager checks travel authorizations in order to ensure that a given booking has been effectively taken into account. For that, a scenario for listing travel authorizations is specified. Therein, once the user goes to the tab “Travel Authorization”, types the booking reference and checks if the request is well registered, then, according to him/her, at this time, he/she is able to ensure whether the request has been taken into account or not. The resultant behavior of the system is to show a tab with a message signaling that the booking is authorized or non-authorized. For this user, we noticed a medium-level of adherence of the language patterns defined in the ontology.

The fourth and last participant P4 (Figure 9) reported a story in the role of a travel department’s intern. He/she describes a research of flights to Paris for attending a seminar from 2nd until 10th May. The participant however has mistakenly informed data details for a specific scenario in the narrative section of the story. As a consequence, when specifying a scenario for this story, he/she supposedly makes reference to the data already informed previously in the wrong section (“*When I provide all the information*”). The scenario also features a search of flights classified by their ascending order of fares. We can notice, in general, the participant had difficulties to understand the structure of the stories. It makes his/her User Story hardly adherent to the implicit proposed language patterns.

Considering the seven levels of Nielsen's linguistic model of interaction (Nielsen, 1986), the stories produced by the participants contain elements that could be classified from the level 1 (goal) until the level the level 5 (lexical).

5.2. Adherence Analyses

Below we present 4 tables (Tables 3, 4, 5 and 6) detailing, for each participant, the behaviors specified by him/her, the adherence of each behavior in the scale presented in the methodology, and a section of comments, where we classify the type of adherence problem identified (if any) and strive the reasons for such a kind of problem. Additionally, we suggest corrections for the problems identified, and when interactive behaviors are concerned, the necessary correction to meet the interactive behaviors in the ontology.

User Story Specification – Participant P1:

Behaviors Specified by the Participant	Possible Correction	Adherence	Comments
-	Title: Booking flights for guests	○○○○○	Lack of statement or keyword. Participant did not title the story.
Narrative:	-	●●●●●	Participant correctly used the keyword.
As a guest	As a travel manager	●●●○○	Wrong information. Participant correctly identified a role, but mistakenly specified the guest as the role who would benefit from the story, when actually it would be the travel manager.
I want airline tickets with defined time and flights	I want to search airline tickets with defined time and flights	●●●●○	Understatement. Participant only forgot the action he/she expects from the system (lack of a verb)
So that I can book tickets	So that I can book tickets for guests	●●●●○	Minor writing complement. Participant did not complement the benefit specifying for whom tickets will be booked.
Scenario: Searching demanded tickets	-	●●●●●	Participant correctly used the keyword with a name for the scenario.
Behaviors Specified by the Participant	Behaviors Defined in the Ontology	Adherence	Comments
Given I go to the site SIMBAB/TRAVEL	Given I go to “SIMBAB/TRAVEL”	●●●●○	Minor writing complement. Participant added the term “the site” in the behavior “I go to” which is not present in the ontology.
When I choose the demanded flight (destination and times, TOULOUSE/PARIS, departure 7 a.m., return 7 p.m. at the same day)	When I choose “Toulouse” in the field “Departure”	●●○○○	Epic behavior. Participant did not break the actions in multiple steps, having informed all the required data for searching in brackets. The behavior “I choose” is nonetheless adherent to the ontology.
	And I choose “Paris” in the field “Destination”		
	And I choose “same day” in the field “Departure Date”		
	And I choose “same day” in the field “Return Date”		
	And I choose “7 a.m.” in the field “Departure Time”		

	And I choose “7 p.m.” in the field “Return Time”		
And I choose type of traveler “Guest”	And I choose “Guest” in the field “Type of Traveler”	●●●●○	Misspecification. Participant did not inform “Type of Traveler” as a field name.
And I search	And I click on “Search”	●●●●○	Understatement. Participant omitted the type of behavior that will trigger the searching.
Then several propositions	Then will be displayed “List of Flights”	●●○○○○	High-level of abstraction. Participant omitted the expected system behavior, only informing that the result will be “several propositions” of flights. “Several propositions” indeed will be proposed in a “List of Flights”, so this is the expected system output behavior.
And I choose the desired flights	And I choose “the desired flights”	●●●●○	Misspecification. Regardless being possible to specify a sequential input behavior in a “Then” clause, the participant is actually describing a second scenario, where he/she provides passenger’s data to effectively book the flight (an input behavior). Considering this step as part of a second scenario, the behavior of choosing the desired flight is highly adherent to the ontology.
When I inform the data concerning the traveler (name, given name, birthdate, phone, mail), and eventually the loyalty card (Flying Blue and Season Ticket)	When I inform “name” in the field “Passenger’s Name”	●●○○○○	Epic behavior. Once more, the participant did not break the actions in multiple steps, having informed all the required data in brackets. The behavior “I inform” is nonetheless adherent to the ontology.
	When I inform “given name” in the field “Passenger’s Given Name”		
	When I inform “birthdate” in the field “Passenger’s Birthdate”		
	When I inform “phone” in the field “Passenger’s Phone”		
	When I inform “mail” in the field “Passenger’s Mail”		

	When I inform “loyalty card” in the field “Passenger’s Loyalty Card”		
Then ticket waiting for validation	Then will be displayed “Ticket waiting for validation”	●●●○○○	Misspecification. Once more, the participant omitted the expected system behavior, only informing that the expected result will be “ticket waiting for validation”, without describing which system’s behavior would be responsible for doing this action. Considering this is meant to be a system behavior, we should note that a behavior defining a status verification for tickets is typically a domain-specific behavior, i.e. it only refers to (and would make sense for) booking systems, so it is not covered by the ontology.

Table 3. User Story Specification – Participant P1.

User Story Specification – Participant P2:

Behaviors Specified by the Participant	Possible Correction	Adherence	Comments
-	Title: Multi-destination flight search	○○○○○○	Lack of statement or keyword. Participant did not title the story.
-	Narrative:	○○○○○○	Lack of statement or keyword. Participant did not use the keyword for describing the story.
As a frequent traveler	-	●●●●●●	Participant correctly identified the role.
I want to search for tickets, providing locations and dates for a multi-destination trip	-	●●●●●●	Participant correctly defined a clear feature.
So that I can obtain information about rates and flight times	-	●●●●●●	Participant correctly defined a clear business benefit.

Scenario: Multi-destination searching	-	●●●●●	Participant correctly used the keyword with a name for the scenario.
Behaviors Specified by the Participant	Behaviors Defined in the Ontology	Adherence	Comments
Given I go to the page “Searching Flights”	-	●●●●●	Participant correctly used the behavior “I go to the page”.
When I choose “Multi-destinations”	-	●●●●●	Participant correctly used the behavior “I choose”.
And I type “Paris” and choose “Paris, Charles de Gaulle” in the field “Departure”	-	●●●●●	Participant correctly used the behavior “I type and choose in the field”.
And I type “Rio de Janeiro” in the field “Destination”	-	●●●●●	Participant correctly used the behavior “I type in the field”.
And I choose “15/02/17” in the field “Departure Date”	-	●●●●●	Participant correctly used the behavior “I choose in the field”.
And I choose “20/02/17” in the field “Return Date”	-	●●●●●	Participant correctly used the behavior “I choose in the field”.
And I type “Rio de Janeiro” in the field “Departure”	-	●●●●●	Participant correctly used the behavior “I type in the field”.
And I type “Porto Alegre” in the field “Destination”	-	●●●●●	Participant correctly used the behavior “I type in the field”.
And I choose “17/02/17” in the field “Departure Date”	-	●●●●●	Participant correctly used the behavior “I choose in the field”.
And I choose “19/02/17” in the field “Return Date”	-	●●●●●	Participant correctly used the behavior “I choose in the field”.
And I click on “Search”	-	●●●●●	Participant correctly used the behavior “I click on”.
Then will be displayed the list of available flights	Then will be displayed “the list of available flights”	●●●●○	Minor writing complement. Participant only forgot quotation marks to indicate (as a variable) that “the list of available flights” is the output expected from the system.

Table 4. User Story Specification – Participant P2.

User Story Specification – Participant P3:

Behaviors Specified by the Participant	Possible Correction	Adherence	Comments
-	Title: Checking Travel Authorizations	○○○○○○	Lack of statement or keyword. Participant did not title the story.
-	Narrative:	○○○○○○	Lack of statement or keyword. Participant did not use the keyword for describing the story.
As a travel manager	-	●●●●●●	Participant correctly identified the role.
I want to check travel authorizations	-	●●●●●●	Participant correctly defined a clear feature.
So that I can ensure the confirmed bookings	-	●●●●●●	Participant correctly defined a clear business benefit.
Scenario: Listing travel authorizations	-	●●●●●●	Participant correctly used the keyword with a name for the scenario.
Behaviors Specified by the Participant	Behaviors Defined in the Ontology	Adherence	Comments
Given I go to the tab “Travel Authorization”	Given I go to “Travel Authorization”	●●●●●○	Minor writing complement. Participant added the term “the tab” in the behavior “I go to” which is not present in the ontology.
When I type the “Booking Reference”	When I type “XXX” in the field “Booking Reference” OR When I type “Booking Reference” in the field “Booking Reference Field”	●●●●●○	Understatement. Participant has omitted either the field name or the value that will be affected by (or affect) this behavior.
And I check if the request is well registered	Then will be displayed “Request well registered”	●●○○○○	Misspecification. The participant did not identify that the information about the booking registration will be provided by the system as an output. For that, a “Then” clause should be used instead of a “When”. Besides that, he/she also specified a domain-dependent behavior, without

			identify how the checking is supposed to be made. He/she could instead use a common interactive behavior presented in the ontology such as “will be displayed”.
Then at this time, I can know for sure (or not) the request has been taken into account	-	○○○○○○	Misspecification. This is not an interactive behavior, but rather a cognitive task. This could also be considered as a business benefit of this story, and as such, it has been correctly specified in the clause “So that” in the beginning of the story.
And it’s shown a tab: authorized / non-authorized	And will be displayed “Authorized” OR And will be displayed “Non-Authorized”	●●●○○○	Misspecification. This step brings the expected output of the system. The participant expects to see a tab with a message signaling whether the booking is authorized or not. This behavior has been put in a “Then” clause, indicating the participant actually understood that showing some information after his/her interaction is a system’s output. However, the participant did not realize that he/she is supposed to inform a valid *or* an invalid state, i.e. he/she should have specified a scenario in which the system would present an authorized booking, and another (if he/she wants) specifying a scenario in which the system would present an unauthorized booking.

Table 5. User Story Specification – Participant P3.

User Story Specification – Participant P4:

Behaviors Specified by the Participant	Possible Correction	Adherence	Comments
--	---------------------	-----------	----------

-	Title: Searching flights to Paris	○○○○○	Lack of statement or keyword. Participant did not title the story.
-	Narrative:	○○○○○	Lack of statement or keyword. Participant did not use the keyword for describing the story.
As an intern	-	●●●●●	Participant correctly identified the role.
I want to book a flight to Paris departing on May 2nd until May 10th	-	●●●○○	Wrong information. Participant mixed a feature description with data for specifying a testable scenario.
So that I can attend a seminar	-	●●●●●	Participant correctly defined a clear business benefit.
Scenario:	Scenario: Searching demanded tickets	●●●○○	Lack of statement or keyword. Participant did not title the scenario but used the appropriate keyword.
Behaviors Specified by the Participant	Behaviors Defined in the Ontology		
Given I'm going to book my flight	-	●○○○○	Understatement. Participant did not identify how or where the activity of booking will be performed in the system. This step is described more as an intent than as an actual behavior.
When I provide all the information	When I inform "..."	●○○○○	High-level of abstraction. Participant did not describe which kind of information should be provided for the scenario. The supposed data to be used here was mistakenly put when specifying the feature in the narrative.
And I choose search by fares	And I click on "Search by fares" OR And I choose "Search by fares" And I click on "Search"	●●●●○	Misspecification. Supposing the system provides different buttons for different types of search, the participant could simply have used the behavior "click on" (supported by buttons) instead of the behavior "choose". Otherwise, "Choose by fares" is a domain-dependent behavior, so for specifying a domain-independent

			behavior, the participant should rather have informed which option he/she would choose (or select) to “search by fares” and then submitting the search by clicking on the respective button, for instance.
<p>Then all the available flights for the date are classified by ascending order of fares</p>	<p>Then will be displayed “List of available flights”</p>	<p>●○○○○○</p>	<p>Misspecification. Again, the participant did lean on a domain-dependent behavior. To specify an action for verifying the arrangement of a list, it would be necessary an ontological behavior allowing to classify datasets in ascending (or even descending) order.</p>
	<p>-</p>		

Table 6. User Story Specification – Participant P4.

Figure 10 presents the assessment of each statement in the User Story specification. Therein, we isolated each one of the statements presented in the template and analyzed, for each participant, the dispersion of results in each degree of adherence and analyzed, for each participant, the dispersion of results in each degree of adherence in the 7-point scale proposed. Such a dispersion has been calculated as a median of the adherence for each stratum proposed in the study design.

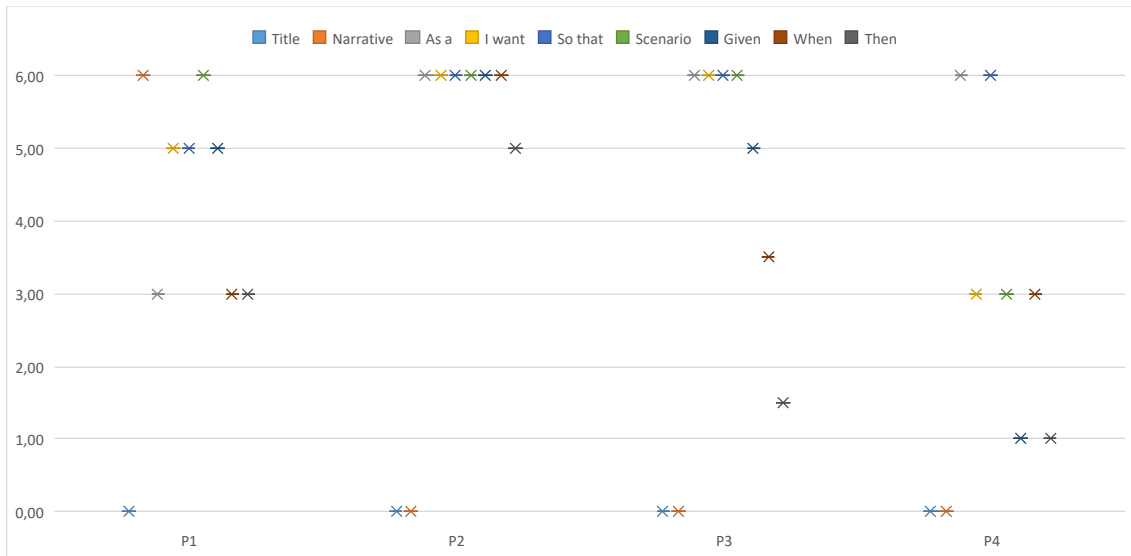


Figure 10. Assessment of each statement in the User Story specification (y-axis: degree of adherence in the 7-point scale proposed).

Figures 11 and 12 present, for each participant, the number of occurrences in each stratum ranging from a null adherence of statements to a full adherence of them. In the first chart (Figure 11), we consolidate only the statements concerning the structure of the User Story, but not addressed as interactive behaviors by the ontology (*narrative* section). Figure 12, on the other hand, consolidates only the statements referring to the interactive behaviors addressed by the ontology (*scenario* section). Finally, Figure 13 presents the consolidated assessment of the User Stories based on the data from all the four participants. The chart summarizes the total amount of occurrences in each stratum of the adherence scale.

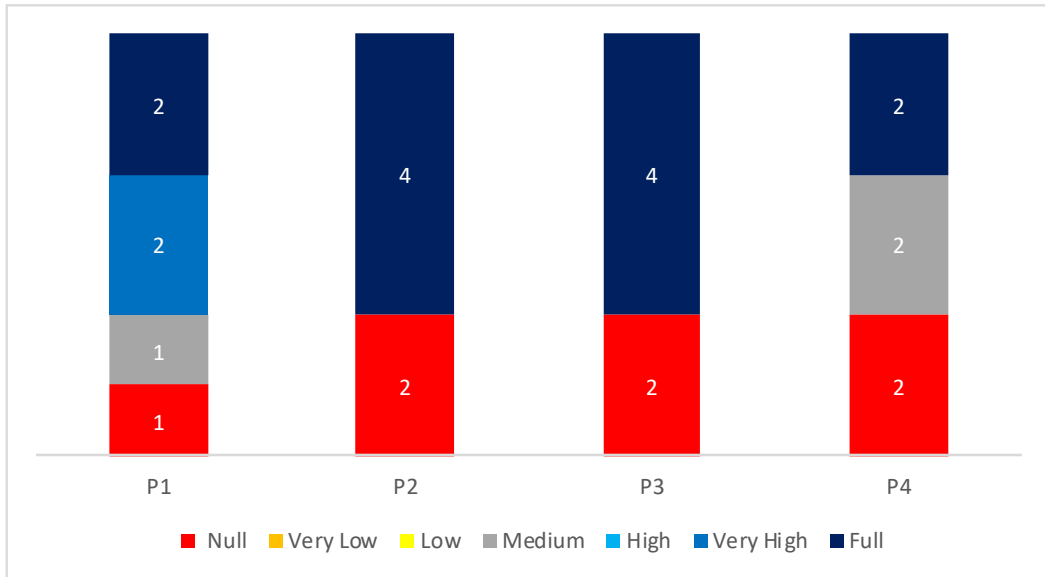


Figure 11. Assessment of the *narrative* section for each participant (y-axis: number of occurrences in each stratum).

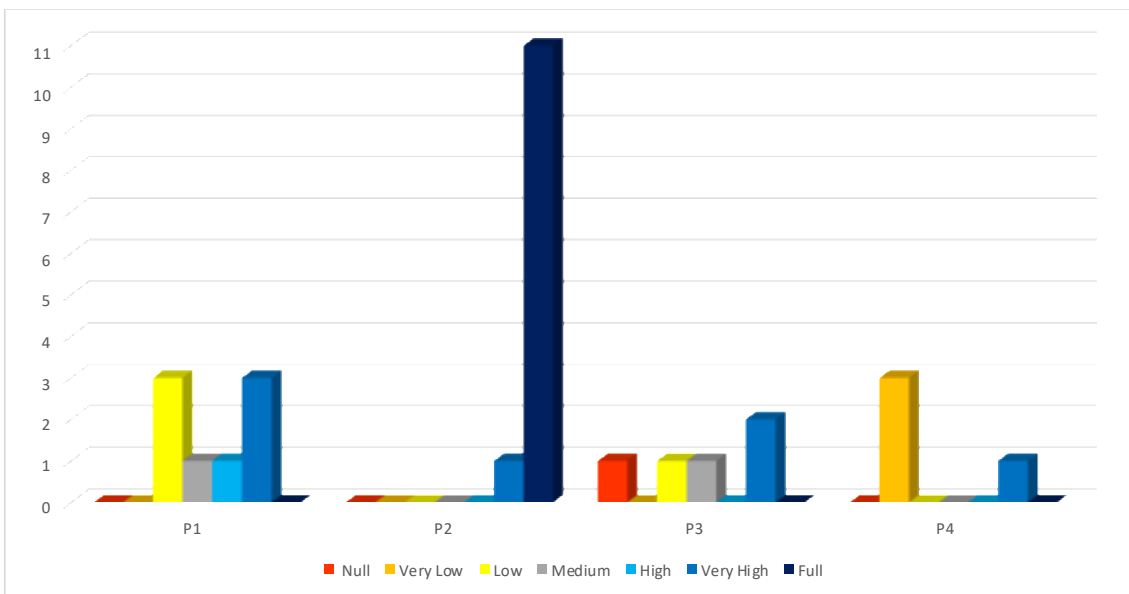


Figure 12. Assessment of the *scenario* section for each participant (y-axis: number of occurrences in each stratum).

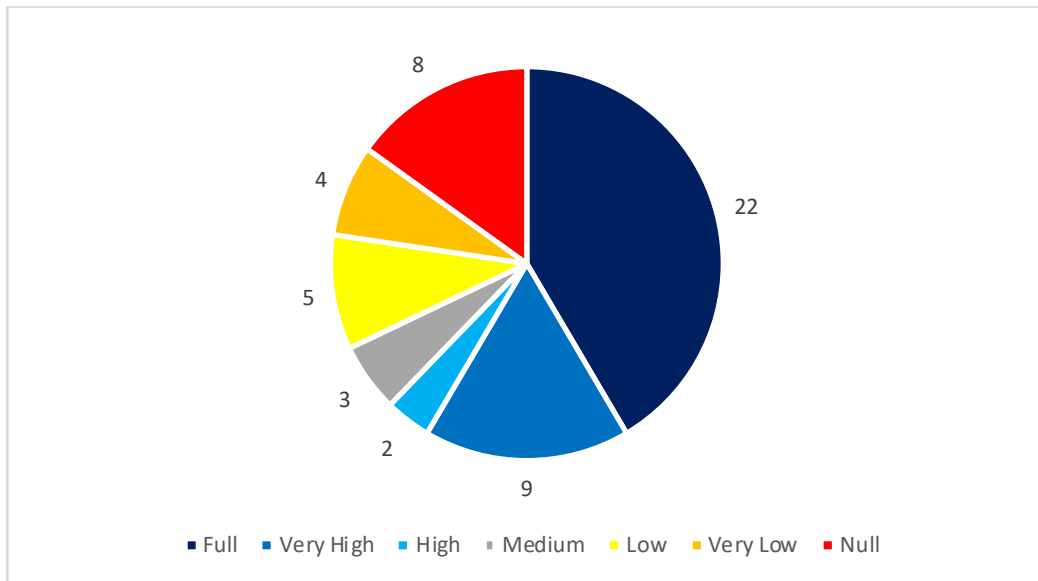


Figure 13. Consolidated assessment of the User Stories (number of occurrences in each stratum).

5.3. Findings Supporting the Research Questions

Hereafter we report the findings related to the two research questions.

RQ 1. *Are participants able to read a basic User Story template and use it to write their own stories?*

Concerning the assessment of the *narrative* section of the User Stories, we notice the majority of participants neglected in titling and using the keyword “*Narrative*” in the beginning of the stories. Only P1 used the keyword, but even him/her did not title the story. We are not sure about the main reasons for that. In a first approach, it seems more like a lack of attention from the participants. All participants, except P1, identified a correct role (statement “*As a*”) for the stories. P1 correctly identified a role, but mistakenly specified the guest as the role who would benefit from the story, when actually it would be the travel manager. This fact is important to notice because it could be part of the travel manager activity to mentally represent a “*persona*” to build a sequence of events

important for the user journey. In fact, 3 of the participants (P1, P2 and P4) described a kind of user journey through the User Story. So maybe this is a cognitive task required by the travel manager job. This, in principle, might work against the User Story writing, especially for the role description.

Concerning the feature description (statement “*I want*”), we noticed a very good rating of this statement, with participants ranging from 5 to 6 in our scale, except P4. P4 has mixed the feature description with data for specifying a testable scenario. Concerning the business benefit expected from the feature (statement “*So that*”), all participants shared a very good rating as well, ranging likewise from 5 to 6.

We have also observed in these charts that the stories produced by P2 and P3 had identical results, both with a majority of full adherence and a medium-to-low number of null adherence statements. These last ones due to the lack of “*Title*” and “*Narrative*” sections of the story. P1 had a low level of null (absence of title) and medium (wrong information when identifying the role) adherent statements, and a clear majority of very-high and full adherent statements. P4 had an equal mix of null, medium and full adherent statements, with problems varying from absence of keywords or sections until the presence of wrong information.

RQ 2. Is the vocabulary used by the participants to write their own User Stories similar to the vocabulary described in the ontology?

Concerning the adherence to the ontology vocabulary in the User Story specification, i.e. the adherence in the section “*Scenarios:*”, we have noticed all the participants titled their scenarios, except P4, that regardless use the right keyword, not titled his/her scenario. For the statement “*Given*”, we have observed a tendency in users specifying more information to define where they are going to access some feature. The

ontology has specified a generic behavior named “*I go to*” and a variation to “*I go to the page*”. However, while P1 used this convention, P2 and P3 have specified respectively, “*I go to the site*” and “*I go to the tab*”, so at this point, somehow the ontology could be enriched to recognize those variants as well. P4, on the other hand, specified a very generic behavior (“*Given I’m going to book my flight*”), not identifying how or where the activity of booking will be performed in the system. This step is described more as an intent than as an actual behavior. For this statement then, P1, P2 and P3 scored a very high adherence (between 5 and 6), while P4 scored a very low adherence (1).

Concerning the statement “*When*”, we notice a mid-range adherence (between 3 and 3,50) for P1, P3 and P4, and a full adherence (6) for P2. P1 produced what we classify as epic behaviors, providing in a same step, several independent actions to be performed on the UI. P3 shared a well formulated step with a misspecification. The participant either confused an output information (that should be specified in a “*Then*” statement) or specified a domain-dependent behavior, not supported by the ontology scope. P4 specified a behavior with a high-level of abstraction without describing which kind of information should be provided, along with a domain-dependent behavior.

Finally, concerning the statement “*Then*”, we notice a pretty low adherence (between 1 and 1,50) for P3 and P4, a mid-range adherence (3) for P1 and a very high adherence (5) for P2. P3 specified a kind of cognitive task in his/her first “*Then*” statement, defining much more a business benefit than an interactive task. His/her second “*Then*” statement brings a misspecification that despite specifying the expected output of the system, it does not comply with a single valid or invalid state, expressing both states in the same expected output. P4 wrote a single misspecified “*Then*” statement, indicating once more the use of a domain-dependent behavior. P1 specified “*Then*” statements with a high-level of abstraction, along with small misspecifications, being one of them related

to the use of a new input interaction, and another related to the use of a domain-specific behavior. P2 committed a really minor writing mistake, only forgetting to use quotation marks to indicate a variable in the interactive behavior.

We have also observed P2 wrote a very high-adherent story with only some minor deviations, especially when describing the narrative. In contrast, P4 wrote a very low-adherent story with the majority of statements classified as having a very low adherence. P1 had half of low and medium adherent statements along with half of high and very high ones. P3 had a slight majority of statements flirting with the low-level stratum (a mix of null, low and medium) and the remaining ones classified as very high adherence.

Looking at the consolidated assessment of the User Stories, we notice however a large majority of statements classified as full or very high adherence to the template. From a total of 53 statements, 33 (62,26%) were classified in the top stratum (full, very high, and high adherence), 3 (5,66%) in the medium stratum, and the remaining 17 (32,08%) in the bottom stratum (low, very low, and null adherence).

5.4. Adherence Problems

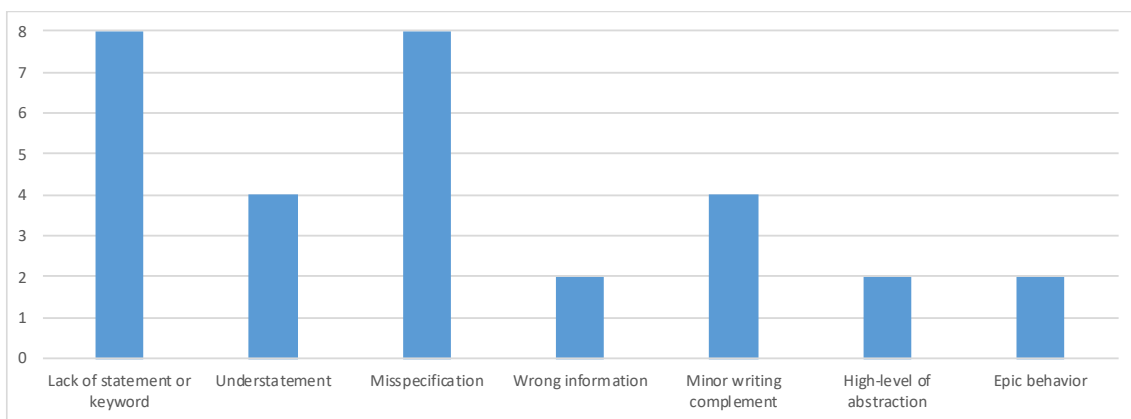


Figure 14. Number of occurrences in each category of adherence problems.

We have also analyzed the types of adherence problems found in the stories specified by the participants. As explained in Section 3, we have categorized 7 types of problems as follows: lack of statement or keyword, understatement, misspecification, wrong information, minor writing complement, high-level of abstraction, and epic behavior. Figure 14 brings the number of occurrences in each category.

In a total of 30 adherence problems identified, we can observe in the chart that the most common types of adherence problems have been the “lack of statement or keyword” and the “misspecification” with 8 occurrences each. It is more than 50% of the problems found (53,33%). “Wrong information”, “high-level of abstraction”, and “epic behavior” were, on the other hand, the types of adherence problems less observed in the participants’ User Stories. With a total of 2 occurrences each, they represent singly no more than 7% of occurrences (6,67%). “Understatement” and “Minor writing complement” complete the set, with each type reporting 4 occurrences, i.e. 13,33% of occurrences each.

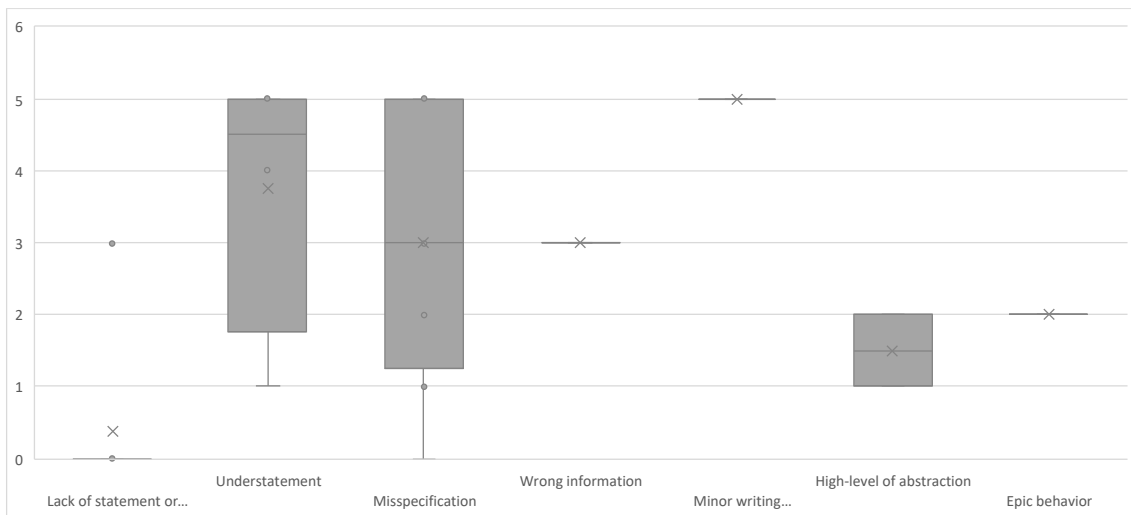


Figure 15. Boxplot of each type of adherence problems identified in participants’ User Stories (y-axis: degree of adherence in the 7-point scale proposed).

Figure 15 brings the boxplot of each type of adherence problems. Y-axis brings the degree of adherence in the 7-point scale proposed. We observe therein that the category “misspecification” had the largest dispersion, ranging from 0 (null, with the lower quartile near 1) to 5 (very high, coinciding with the upper quartile), with a median (and a mean) at the medium stratum of adherence problems. “Understatement” had the second largest dispersion, ranging from 1 (low, with the lower quartile near 2) to 5 (very high, coinciding with the upper quartile) with a median at the top stratum (4,5) and the mean near the level 4 of adherence. “High-level of abstraction” comes next with a dispersion between 1 (very low, coinciding with the lower quartile) to 2 (low, coinciding with the upper quartile), with median and mean in 1,5. “Lack of statement or keyword”, “wrong information”, “minor writing complement”, and “epic behavior” had no dispersion, and achieved respectively a median of 0, 3, 5 and 2. Equal results have been observed for the mean of these types of adherence problems (“lack of statement or keyword” had a mean slightly above the median). Outliers have been observed for “lack of statement or keyword” with just an occurrence of an adherence problem classified in the medium stratum with all the others classified in the null stratum.

6. Discussion

6.1. Individual Performance

Analyzing the individual performance of the participants, we noticed P1 and P3’s stories performed primarily at the top stratum (very high and full adherences for P1, and full adherence for P3) for the *narrative* section. For the *scenario* section however, P1 mixed a performance of half occurrences at the top stratum (high and very high adherences) and half at the medium to low stratum (medium and low adherences), while P3’s story

performed primarily at the medium to low stratum (null, low and medium adherences for 2/3 of occurrences) with the remaining occurrences being classified as very high. P1 and P3 had the largest number of behaviors marked as “misspecification”, confirming a particular difficulty to assimilate some structures of User Stories in the proposed template, mixing primarily the writing of some domain-dependent behaviors with “understatement” and “minor writing complement”.

“Epic behaviors” have been only specified by P1. In the context it has been made, a sequence of data input in a form, this error could signalize the need of tables to enter a set of data in forms. This kind of solution has been proposed by the FIT Framework¹, however it is not covered by our ontology so far. “High-level of abstraction” however has been observed in stories written by P1 and P4. P4’s low performance (2/3 of occurrences classified in the null and medium adherences for the *narrative* section, and the clear majority of occurrences classified as very low for the *scenario* section) could find an explanation in the participant’s lack of experience in the business processes of the institute (just a month), despite having 4 years of experience working for other companies.

Analyzing the best performance and the highest adherence of P2’s stories (primarily at the top stratum – very high and full adherences for the *narrative* section – with an overwhelming majority of full adherent statements – with a single very high adherence – for the *scenario* section), and being the second younger participant besides the second more experienced one, we wonder about the role played by the sum of age and

¹ <http://fit.c2.com/>

experience factors in the willingness and commitment to adopt new ways of work. P2 had clearly the better performance with the lowest ratio age/experience in the group.

6.2. General Performance

Based on the results presented above, we can highlight some important findings about the writing profile of User Stories used by the participants. The wide dispersion of adherence problems classified as “misspecification” means the participants’ stories had a varied level of compliance for the problems found in this category, ranging from slightly mistaken identifications of fieldnames to specification of highly domain-dependent behaviors. The specification of domain-dependent behaviors was one of the most frequent issues. Despite the high number of misspecifications, participants seemed to understand the purpose of scenarios. They were able to describe the expected behavior for the system even when they did not use the set of predefined interactive behaviors. This fact is confirmed by the medium to low adherence in the “*When*” and “*Then*” statements of the story, where typically reside the most interactive behaviors in a scenario, and consequently, where the ontology is more used to specify them.

Concerning the general assessment of the User Stories, we notice a clear concentration of occurrences in the top stratum (62,26%), which indicates an overall very good adherence of User Stories to the proposed template and a limited usage of our predefined interactive behaviors presented in the ontology. Analyzing each statement of the stories individually, we also notice a clear concentration of occurrences in the top stratum, exception made for the aforementioned “*When*” and “*Then*” statements which are dispersed mostly between the medium and the low stratum, and for “*Title*” and “*Narrative*” statements that were almost always omitted by the participants, which occasioned consequently a null adherence for both of them with only one exception.

Despite the high dispersion of problems related to “Understatement”, they presented a median at the top stratum (4,5), which means the level of noncompliance for this kind of problem is very low. So, we conclude that participants made just slight deviations from the proposed template. “Lack of statement or keyword”, despite the high number of occurrences, was primarily found in the “*Title*” and “*Narrative*” statements that were frequently omitted by the participants, which explains the prevalence of null adherence for this type of problem. “High-level of abstraction” and “epic behaviors” presented problems with a low level of adherence to the proposed template once these kinds of problems are associated to descriptions with a low level of interaction details, which is opposed to what is defined in the ontology. As expected, “minor writing complement” had a very high rate of adherence with behaviors presenting only minor deviations from the proposed template.

6.3. Opportunities for Improvement

These findings bring us some opportunities for improving our current set of interactive behaviors in the ontology. As stated before, the adoption of tables with data examples together with the ontology could reduce the workload of describing input of data in forms and stimulate a complete specification by users. The ontology could also be enriched to recognize variants for the same interactive behavior. Participants of this study specified some behaviors very close to the ontology statements, but with minor variants. Even mapping synonyms for some specific behaviors, it does not provide any kind of semantic interpretation, i.e. behaviors must be specified on stories exactly as they were defined in the ontology. Further studies on Natural Language Processing (NLP) techniques might help to improve the process of specification adding more flexibility to write scenarios that could be semantically interpreted to meet the behaviors described in the ontology.

Another aspect to be considered is the high number of domain-dependent behaviors specified by the participants. This points out the need of considering a still higher level of descriptions for our behaviors. Domain-specific behaviors have the disadvantage of being dependent on the jargon used for each type of business processes, which would implicate in developing different ontologies for different business processes, with each one encompassing the proper jargon of each domain. Domain-specific ontologies nonetheless could act as a top layer in a multi-layer ontology architecture to allow the use of multiple domain ontologies associated to the current domain-independent ontology, which would remain describing only the fundamental interactive behaviors for a given environment.

6.4. Threats to Validity

Generalization of results. We have selected a representative group of participants as Product Owners (POs) in a system for booking airline tickets for business trips. Such kind of system has usually a strong search-based feature, once they are centered in providing and comparing rates, times and availability of flights given a set of provided parameters. However, as the ontology in which we based our analyses is designed for domain-independent interactive behaviors, we assume our results would be reproduced in other interactive systems domains. The profile and previous experiences of the participants could, nonetheless, bring different results. Studies involving Product Owners previously introduced to User Stories and/or test automation could bring different and less frequent adherence problems.

Sample size. We have conducted this study with 4 participants that could eventually assume a role of Product Owners in a typical scenario of software development. Our results are certainly limited to the profile and experience of these 4

participants. Studies conducted with a bigger sample could bring different adherence problems and/or reduce the variability of occurrences when looking to the whole group. It could eventually bring more homogeneous results.

Absence of training. This study has been conducted without training. Studies involving prior training in the vocabulary used in the ontology would certainly bring different results due to the background knowledge. However, such results would not capture the factor of users choosing their own vocabulary to express their interaction needs. This factor is useful to identify the suitability of the predefined interactive behaviors to naturally express the user's intents.

Possible interpretation bias. Due to the empirical nature of this study, interpretation bias should be considered as a threat to validity. To mitigate this threat, the results were cross-checked by an independent reviewer, expert in human factors engineering. The independent reviewer analyzed the raw data provided by the participants to the study. Then the independent reviewer performed a qualitative data analysis of the evidences in the stories written by each participant. Additionally, the reviewer commented on a pre-scaling of the adherence of each statement and provided a set of expected modifications and recommendations. Finally, this review process resulted to a convergence in the scaling of the adherence of each statement presented in the article.

7. Conclusion and Future Works

This article presented a study that we have conducted to evaluate the writing of User Stories by potential Product Owners. When analyzing the adherence of the User Stories produced by the participants, the study has shown they had an overall good rating concerning the statements and the structure, and a moderate-to-high usage of the implicit vocabulary to describe interactive behaviors in the ontology. The results suggest a high

level of adherence of statements when the User Story structure is considered (*narrative* section and *scenario/given/when/then* clauses), i.e. the participants were able to read a basic User Story template and use it to write their own stories (RQ 1). The vocabulary of the interactive behaviors described in the ontology was, to some extent, reproduced by the participants even without a specific prior training (RQ 2). Indeed, all the interactive behaviors presented in the example have been used by the participants at least once, and other behaviors in the ontology (not presented to the participants) have also been employed, intuitively. This finding might suggest that the ontology provides a somewhat natural vocabulary for stakeholders. Finally, we identified 7 types of adherence problems in the participants' User Stories, being "lack of statement or keyword" and "misspecification" the most common ones.

The results presented in this article have also supported the development of a broader approach aiming at automating the assessment of user interface design artifacts based on the user requirements expressed in the User Stories. This approach has been previously published by us (Silva, 2016; Silva and Winckler, 2016, 2017, Silva, Hak and Winckler, 2016b, 2016a) and so far has been used to assess user interface design artifacts. With the support of the vocabulary proposed by the aforementioned ontology, we are also designing a domain-specific language (DSL) for specifying user requirements in a higher level of abstraction, which could guide POs in the specification of their own User Stories.

As future works, we will investigate the impact of training for writing User Stories using the ontology. We suggest better adherence level of the stories will be reached with prior training of the participants. We will also include a larger sample of participants with more diverse expertise and roles with respect to the use of the application. Finally, we are willing to investigate other application domains, which can contribute to address new challenges for the human performance with complex systems such as Cyber-Physical &

Human Systems (CPHS). Actually, CPHS may lead to *intrinsic dissonances* with end users representations (Vanderhaegen, 2017) that could be partially covered by the approach described in this paper.

Acknowledgments

We would like to thank the anonymous participants of this study and the travel department of the Toulouse Institute of Computer Science Research (IRIT) for authorizing the study.

References

- Ambler, S. (2002) *Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process*. 1st edn. Wiley.
- Bano, M. and Zowghi, D. (2013) ‘User Involvement in Software Development and System Success: A Systematic Literature Review’, in *EASE '13: Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering*, pp. 125–130. doi: 10.1145/2460999.2461017.
- Business Process Model And Notation™ (BPMN™)* (2011) *Object Management Group*. Available at: <http://www.omg.org/spec/BPMN/2.0/> (Accessed: 1 December 2017).
- Chelimsky, D. *et al.* (2010) *The RSpec Book: Behaviour Driven Development with RSpec, Cucumber, and Friends*. Pragmatic Bookshelf.
- Cohn, M. (2004) *User Stories Applied for Agile Software Development*. Addison-Wesley.
- Gomez, A., Rueda, G. and Alarcón, P. P. (2010) ‘A systematic and lightweight method to identify dependencies between user stories’, in *XP 2010*, pp. 190–195. doi: 10.1007/978-3-642-13054-0_17.
- Kassab, M. (2015) ‘The changing landscape of requirements engineering practices over the past decade’, in *5th International Workshop on Empirical Requirements Engineering, EmpiRE 2015 - Proceedings*. IEEE. doi: 10.1109/EmpiRE.2015.7431299.
- Lucassen, G. *et al.* (2015) ‘Forging High-Quality User Stories: Towards a Discipline for Agile Requirements’, in *2015 IEEE 23rd International Requirements Engineering*

- Conference, *RE 2015 - Proceedings*, pp. 126–135. doi: 10.1109/RE.2015.7320415.
- Lucassen, G. *et al.* (2016) ‘The Use and Effectiveness of User Stories in Practice’, in *Requirements Engineering: Foundation for Software Quality*. Springer, pp. 205–222. doi: 10.1007/978-3-319-30282-9_14.
- Lucassen, G. *et al.* (2018) ‘Jobs-to-be-Done Oriented Requirements Engineering: A Method for Defining Job Stories’, in *International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2018)*. Springer International Publishing, pp. 227–243. doi: 10.1007/978-3-642-37422-7.
- MacLeod, I. S. (2008) ‘Scenario-based requirements capture for human factors integration’, *Cognition, Technology and Work*, 10(3), pp. 191–198. doi: 10.1007/s10111-007-0099-3.
- Maguire, M. and Bevan, N. (2002) ‘User Requirements Analysis: A Review of Supporting Methods’, in *IFIP World Computer Congress*. Kluwer Academic Publishers, pp. 133–148.
- Meyer, B. (1985) ‘On Formalism in Specifications’, *IEEE Software*, 2(1), pp. 6–26. doi: 10.1109/MS.1985.229776.
- Nielsen, J. (1986) ‘A Virtual Protocol Model for Computer-Human Interaction’, *International Journal of Man-Machine Studies*, 24(3), pp. 301–312. doi: 10.1016/S0020-7373(86)80028-1.
- North, D. (2019) *What’s in a Story?* Available at: <https://dannorth.net/whats-in-a-story/> (Accessed: 1 January 2019).
- Schwaber, K. (2004) *Agile Project Management with Scrum*. Microsoft Press.
- Silva, T. R. (2016) ‘Definition of a Behavior-Driven Model for Requirements Specification and Testing of Interactive Systems’, in *Proceedings - 2016 IEEE 24th International Requirements Engineering Conference, RE 2016*, pp. 444–449. doi: 10.1109/RE.2016.12.
- Silva, T. R., Hak, J.-L. and Winckler, M. (2016a) ‘An Approach for Multi-Artifact Testing Through an Ontological Perspective for Behavior-Driven Development’, *Complex Systems Informatics and Modeling Quarterly*, (7), pp. 81–107. doi: 10.7250/csimq.2016-7.05.

- Silva, T. R., Hak, J.-L. and Winckler, M. (2016b) ‘Testing Prototypes and Final User Interfaces Through an Ontological Perspective for Behavior-Driven Development’, in *6th International Working Conference on Human-Centred Software Engineering, and 8th International Working Conference on Human Error, Safety, and System Development (HCSE 2016 and HESSD 2016)*, pp. 86–107. doi: 10.1007/978-3-319-44902-9.
- Silva, T. R., Hak, J.-L. and Winckler, M. (2017a) ‘A Behavior-Based Ontology for Supporting Automated Assessment of Interactive Systems’, in *Proceedings - IEEE 11th International Conference on Semantic Computing, ICSC 2017*, pp. 250–257. doi: 10.1109/ICSC.2017.73.
- Silva, T. R., Hak, J.-L. and Winckler, M. (2017b) ‘A Formal Ontology for Describing Interactive Behaviors and Supporting Automated Testing on User Interfaces’, *International Journal of Semantic Computing*, 11(04), pp. 513–539. doi: 10.1142/S1793351X17400219.
- Silva, T. R. and Winckler, M. (2017) ‘A Scenario-Based Approach for Checking Consistency in User Interface Design Artifacts’, in *IHC’17, Proceedings of the 16th Brazilian Symposium on Human Factors in Computing Systems*, pp. 21–30. doi: 10.1145/3160504.3160506.
- Silva, T. R. and Winckler, M. A. A. (2016) ‘Towards Automated Requirements Checking Throughout Development Processes of Interactive Systems’, in *22nd International Working Conference on Requirements Engineering – Foundation for Software Quality, REFSQ 2016*, pp. 1–2.
- Trkman, M., Mendling, J. and Krisper, M. (2016) ‘Using business process models to better understand the dependencies among user stories’, *Information and Software Technology*. Elsevier B.V., 71, pp. 58–76. doi: 10.1016/j.infsof.2015.10.006.
- Valente, P. *et al.* (2016) ‘Bridging Enterprise and Software Engineering Through an User-Centered Design Perspective’, in *Web Information Systems Engineering – WISE 2016*, pp. 349–357. doi: 10.1007/978-3-319-48743-4.
- Valente, P. *et al.* (2017) ‘The Goals Approach: Agile Enterprise Driven Software Development’, in *Complexity in Information Systems Development*, pp. 201–219. doi: 10.1007/978-3-319-52593-8.

Vanderhaegen, F. (2017) ‘Towards increased systems resilience: New challenges based on dissonance control for human reliability in Cyber-Physical&Human Systems’, *Annual Reviews in Control*. Elsevier, 44, pp. 316–322. doi: 10.1016/j.arcontrol.2017.09.008.

Vanderhaegen, F. and Carsten, O. (2017) ‘Can dissonance engineering improve risk analysis of human–machine systems?’, *Cognition, Technology and Work*. Springer London, 19(1), pp. 1–12. doi: 10.1007/s10111-017-0405-7.

Wautelet, Y. *et al.* (2014) ‘Unifying and Extending User Story Models’, in *International Conference on Advanced Information Systems Engineering (CAiSE 2014)*. Springer, pp. 211–225. doi: 10.1007/978-3-319-07881-6_15.

Wautelet, Y. *et al.* (2018) ‘On Modelers Ability to Build a Visual Diagram from a User Story Set: A Goal-Oriented Approach’, in *International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2018)*. Springer International Publishing, pp. 209–226. doi: 10.1007/978-3-642-37422-7.