From Ideas to Expressed Needs: an Empirical Study on Early Requirements Evolution

1st Given Name Surname
dept. name of organization (of Aff.)
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
email address or ORCID

***Abstract*—Requirements are elicited from the customer and other stakeholders through an iterative process of interviews, prototyping, and other interactive sessions. Many communication phenomena may emerge in these early iterations, that lead initial ideas to be transformed, renegotiated, or reframed. Understanding how this process takes place can help in solving possible communication issues as well as their consequences. In this work, we perform an exploratory study of descriptive nature to understand in which way requirements get transformed from initial ideas into documented needs. To this end, we select 30 subjects that act as requirements analysts, and we perform a set of elicitation sessions with a fictional customer. The customer is required to study a sample requirements document for a system beforehand and to answer the questions of the analysts about the system. After the elicitation sessions, the analysts produce user stories for the system. These are compared with the original ones by two researchers to assess to which extent and in which way the initial requirements evolved throughout the interactive sessions. Our results show that between 30% and 38% of the produced user stories include content that can be fully traced to the initial ones, while the rest of the content is dedicated to new requirements. We also show what types of requirements are introduced through the elicitation process, and how they vary depending on the analyst. Our work contributes to theory in requirements engineering, with empirically grounded, quantitative data, concerning the impact of elicitation activities with respect to initial ideas.**

***Index Terms*—requirements elicitation, user stories, information, requirements evolution**

I. Introduction

Requirements are elicited from the customer and other stakeholders through an iterative process of interviews, prototyping, and other interactive sessions [1]–[3]. The iterations transform the initial ideas of the customer into more explicit needs, normally expressed in the form of a requirements document to be used for system specification. Large part of the research has focused on requirements inspection [4]–[6], to identify possible defects of requirements documents, for example in terms of ambiguity [7], [8] or incompleteness [9], [10]. However, researchers observed that many communication problems may have already emerged in the early iterations before the requirements are written down in their final form [11], [12]. Relevant needs may have remained unexpressed, novel requirements may have been introduced, and others may have been discarded through early negotiation. Understanding how this process takes place can help to solve possible communication issues as well as their consequences.

In this work, we perform an exploratory study of descriptive nature to understand in which way requirements get transformed from initial ideas into documented needs. To this end, we recruit 58 subjects that will act as requirements analysts, and we perform a set of elicitation sessions with a fictional customer. The fictional customer is required to study a set of about 50 user stories for a system, which are regarded as the initial customer ideas for the experiment. Then, each analyst performs two requirements elicitation interview sessions with the customer, who is required to answer based on the user stories, and on novel ideas that can be triggered during the conversation. The sessions are separated by a period of 2 weeks, in which the analyst is working on the data collected in the first interview through notes, diagrams, or mockups. After the elicitation sessions, each analyst documents the requirements into 50 to 60 user stories. The produced user stories from a sample of 30 subjects are compared with the original ones by two researchers to assess to which extent and in which way the initial requirements evolved throughout the interactive sessions.

Based on the analysis, we provide the following main findings:

- Only between 30% and 38% of the produced user stories include content that can be fully traced to the initial ones;
- Most of the requirements produced after elicitation and documentation—specifically between 54% and 61%—are refinements of the initial ideas, while between 13% and 20% are related to completely novel ideas;
- Most of the roles involved—between 59% and 75%—are the same as the ones initially planned, but between 16% and 30% are entirely novel roles;
- The relevance given to certain requirement categories is different between initial ideas and documented needs;
- The original roles are mostly preserved in the analysts' user stories, but the distribution of the analysts' stories among roles differs;
- There are recurrent new requirements and roles across analysts.

The paper is organized as follows. Section II summarizes the

related work. Section III describes the adopted methodology, Section IV presents the results of our analysis and Section V discusses their implications. Section VI describes the threats to the validity of our results and how they have been mitigated and Section VII concludes the paper.

II. Background and Related Work

Our work focuses on the early stages of requirements evolution, where the customers' initial ideas are discussed and refined to produce a first requirements document. This document is based on both the elicitation process focused on collecting the customer's needs, and the creativity of the analysts, which can influence the interviews and the subsequent analysis.

A. Requirements Evolution

Requirements evolution is a well-recognized phenomenon that can have critical effects on software systems [13]–[15].

Zowghi and Gervasi [16] analyze how the initial incomplete knowledge about a system evolves and identify consistency, completeness, and correctness (the "three Cs") as the driving factors. The main idea behind this conclusion is that the goal of an analyst is to produce a complete, consistent, and consequently correct set of requirements. Thus, the analysts keep working using the collected knowledge and their expertise trying to get closer to the three Cs at every step of the process. Grubb and Chechik [15] focus on the early stage of requirements evolution, but they look at the modeling phase. In particular, they propose to use goal model analysis to help stakeholders to answer *what if* questions to support the evolution of a system considering different scenarios, as well as the customer in understanding trade-offs among different decisions. In their approach, the authors augment goal models with the capability of explicitly modeling time to provide a more useful analysis for the stakeholders. Other authors have looked into the concept of *pre-requirements* [17], intended as information available prior to requirement specification—-including system concepts, user expectations, the environment of the system—-and their tracing with expressed needs.

Requirements keep evolving also when the system is deployed. Carreño and Winbladh [18] suggest the idea that user feedback is a critical factor for requirements evolution and needs to be used for this purpose. The authors created a system to automatically extract topics from user feedback and generate new requirements for future versions of the app. Feedback in the form of app reviews is analysed by a stream of works from Maalej and his team (e.g., [19], [20]). Along the same line of research, Guzman *et al.* [21] proposes to use the information mined in Twitter to guide the evolution of requirements. Additional similar approaches are discussed by Khan *et al.* [22] and by Morales-Ramirez *et al.* [23].

Ali *et. al.* [24] identify in "assumptions" one of the main reasons behind requirements evolution. The authors develop a system to monitor the assumptions and evolve the model of the systems every time they are violated.

B. Creativity in RE

Creativity plays an important role in many requirements engineering activities. This includes the requirements evolution process [25]–[27].

According to Sternberg, creativity can be described as *the ability to produce work that is both novel and appropriate* [28]. Nguyen [29] states that creativity can be attributed to five factors: product, process, domain, people, and context. To analyze the impact of creativity in discovering new requirements, Maiden *et al.* [30] performed a study consisting of a series of creative workshops to discover new requirements for an Air Traffic Management System. This work provides empirical evidence of the impact of creativity and creative processes in identifying new requirements. Inspired by these seminal contributions on creativity and RE, the CreaRE workshop was established[1], and it is currently at its 10th edition, indicating the interest of the community in the topic. Several techniques have been experimented in the literature; among them, the EPMcreate technique [31], theoretical frameworks for understanding creativity in RE [29], platforms to support collaboration for distributed teams [32], toolboxes for selecting the appropriate creativity technique [33], [34], and the use of combination of goal modeling and creativity techniques [35], [36].

C. Contribution

With respect to the literature on requirements evolution, our work is among the first ones that focuses on early requirements elicitation performed with traditional interviews, which are extremely common in practice [3], [37]. The closest work to ours is the contribution of Hayes *et al.* [17], focusing on pre-requirements information. Their concept of pre-requirement is analogous to our notion of "initial idea". However, their goal is to aggregate and automatically cluster pre-requirements from multiple stakeholders, and support traceability. Instead, in our paper we want to evaluate how the initial ideas get transformed through the elicitation and documentation process.

Compared to work on creativity our study also differs from the literature. Instead of providing a novel technique to stimulate creativity, it gives quantitative evidence on the impact of early elicitation and documentation activities. In particular, it shows that (a) creativity takes place as a natural phenomenon without introducing specific triggering techniques; (b) a quantitative evaluation is possible, and can be used to compare different creativity techniques for requirements elicitation.

III. Research Design

A. Research Questions

The overarching objective of this research is to explore in which way requirements evolve from ideas to expressed needs. In this study, the overarching goal is addressed by considering the following main research question: *What is the difference between the initial customer ideas and those documented by an analyst after requirements elicitation?*

[1] https://creare.iese.de

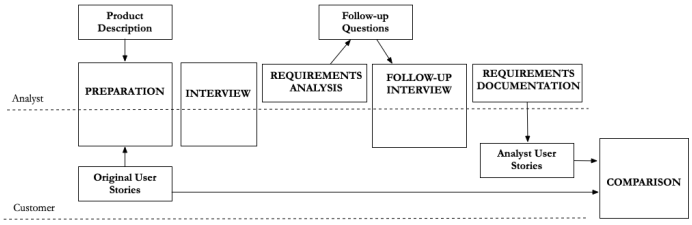

Fig. 1. Steps of the experiment.

This is further decomposed into the following sub-questions:

- RQ1: *How much is the difference in terms of documented requirements and roles with respect to initial ideas?*
- RQ2: *What is the relevance given to the different categories of requirements and roles with respect to initial ideas?*
- RQ3: *What are the emerging categories and roles and what is their impact on the requirements?*

B. Data Collection

To answer the questions, we perform an experiment with one fictional customer (1st author of the current paper), and a set of 58 different student analysts recruited from $< omitted >$. The steps of the experiment (approved, together with the recruitment process, by the ethical committee of $< omitted >$) are described in the following. Please refer to Fig. 1 for an overview of the steps.

1. Preparation Analysts are given a brief description of a product to develop and are asked to prepare questions for a customer that they will have to interview to elicit the products' requirements. The product is a system for the management of a summer camp. The brief description has an initial part that describes the company current practice followed by a set of briefly described needs about (1) managing the information, registration, and activities of the participants, (2) giving the participants' guardians the opportunity to register and follow their children, (3) managing the employees' performance and schedule, (4) communicating with parents and employees, and (5) managing facilities.
The fictional customer is required to study a set of about 50 user stories for the system, which are regarded as the initial customer ideas for the experiment. The initial user stories are taken from the dataset by Dalpiaz [38], file `g21.badcamp.txt`.

2. Interview Each analyst performs a 15 minutes interview with the customer, possibly asking additional questions with respect to the ones that they prepared. The fictional customer answers their questions based on the set of user stories that describe the product, and that are not shown to the analysts. Overall, the customer is required to stick to the content of the user stories as much as possible. However, he is allowed to answer freely when he does not find a reasonable answer in the user story document, to keep the interview as realistic as possible, and capture novel ideas that emerge in the dialogue. The students are required to record their interviews, and take notes.

3. Requirements Analysis Based on the recording and their notes, the analysts have to: (a) perform an initial analysis of the requirements, and based on this analysis (b) produce additional questions for the customer to be asked in a follow-up interview. The initial analysis can be performed with the support of a graphical prototype, use cases, or written form. Analysts can adopt the method they find more suitable.

4. Follow-up Interview Then, they perform a follow-up interview with the customer, which also lasts 15 minutes, to ask the additional questions prepared. During the interview, they can use the graphical prototype, the use cases, or any material produced as a support to ask questions to the customers. In practice, they can show the material to the customer and discuss based on the material.

5. Requirements Documentation After the second interview, they are required to write down from 50 to 60 user stories for the system. This choice aims to enable comparison with the original set. About 50 user stories are also the typical number in the dataset by Dalpiaz [38], which we deem representative of user story sets used for research purposes.

6. Comparison The user stories of 30 randomly selected analysts are manually inspected by two researchers (1st and 2nd author) to identify:

- User stories that express content that was already entirely present in the initial set of user stories (marked as "existing", **E**).
- User stories that express content that is novel with respect to the initial set of user stories, but that belongs to one of the existing high-level categories of the initial set (marked as "refinement" **R**).
- User stories that express content that is novel, and belonging to a novel category not initially present (marked as "new" **N**).
- The name of novel categories of user stories introduced.
- Recurrent themes in **R** and **N** stories.
- Roles that were used also in the original stories ($\mathbf{E}_\rho$).
- Roles that represented a refinement of roles used in the original stories ($\mathbf{R}_\rho$).
- Roles that were novel and never considered in the original stories ($\mathbf{N}_\rho$).

This process is carried out by means of a template spreadsheet that is used to annotate the user stories. Given a list of user stories produced by one of the analysts, a researcher went through the list, and marked each user story with E, R, or N. Whenever a user story was marked with N, the researcher was asked to report the name of the new category identified. An excerpt of the data analysis for one of the user story documents is reported in Fig. 2.

Subsequently, the roles used in the stories were extracted and marked as E_ρ, R_ρ or N_ρ. For the case of E_ρ and R_ρ roles, the researcher also indicated the corresponding role in the original story. For N_ρ, the role was added to the list of roles.

Validity Procedure. To extract the original categories and roles and ensure the validity of the procedure, the following

USER STORY	Customers	Facilities	Personnel	Camp	Communication	New Feature
As the Admin, I want the interface for this app to be simple to use, so that there is minimal time required to start using it.						N - Usability
As an employee, I want this app to easily reschedule my week, so that if fellow employee wishes to swap shifts or needs to be covered I can quickly and seamlessly do so.			R			
As the Admin, I want to be able to send alerts parents regarding their children, so that if one needs immediate attention, I am able to contact the parents quickly.					E	
As the Admin, I want this app to run on any phone platform, so that employees with either an Android or iPhone can use it.						N - Portability
As the Admin, I want this app to allow students to have usernames so that counselors and other campers can tag someone and highlight what is going on in real time.	E				R	

Fig. 2. Extract of a compiled spreadsheet.

tasks were performed by the researchers (all three authors) before the comparison activity:

- **Definition of existing user story categories:** The initial set of user stories used as preparation material for the interviewee has been analyzed by the three authors independently to identify the emerging categories.
- **Annotation of a sample set of user stories:** To validate the extracted categories each researcher independently used the identified categories to label two sets of user stories, for a total of 107 user stories.
- **Preliminary check of agreement on the annotation:** the researchers accessed the other categories and labeled user stories of the other researchers and then met in a 1 hour and 30 minutes meeting to preliminary reconcile the disagreements. The meeting provided as an output a set of guidelines to consolidate the different labels in high-level categories.
- **Consolidation of the categories and schema:** After the meeting, the 2nd author used the guidelines to consolidate the categories and re-labeled the original user stories using these categories. The final categories were discussed among the authors and finalized. The spreadsheet model shown in Fig. 2 and used for the comparison procedure was produced in this phase.
- **Application of the consolidated schema:** Using the spreadsheet model the 1st and 2nd authors independently labeled more than 100 user stories and then met to reconcile. The disagreement on the category to assign was minimal and only related to the novelty within a category and not to the assignment in the categories The 3rd author analyzed the spreadsheet model and approved the consolidated schema.

To further ensure the validity of the procedure, as each set of user stories was analyzed only by one researcher, during the analysis, the user stories that raised doubts were marked to be discussed in a subsequent meeting between the 1st and the 2nd authors This meeting was broken over two days for a total of more than 6 hours.

The analysis of the roles did not require a similar effort as the original user stories clearly identified three well-separated roles.

C. Data Analysis

Data analysis is carried out based on the results of the comparison activity and consists of a quantitative hypothesis testing activity, and thematic analysis on the content followed by an analysis of frequencies of themes.

For hypothesis testing, we consider the following study variables, oriented to give a quantitative representation of the concepts of initial customer ideas and documented ideas, as well as their differences.

The dependent variables of the study are:

- **Conservation rate:** rate of produced user stories that include content that can be traced directly to the original set of user stories. More formally, given a set of user stories produced by a certain analyst, let e be the number of user stories that are marked as **E** by the researchers for some of the categories, let T be the total number of user stories for the analyst, the conservation rate c is defined as $c = e/T$.
- **Refinement rate:** rate of produced user stories including content that can be traced to the *categories* of the original set of user stories, but that provide novel content within one or more of those categories. Formally, the refinement rate r is $r = en/T$, where en is the number of user stories marked as **R**.
- **Novelty rate:** rate of produced user stories including content that is novel, and cannot be traced to existing categories. Formally, the novelty rate is $v = n/T$, where n is the number of user stories marked as **N**.

For the roles, we have analogous variables, Role Conservation, Refinement and Novelty Rate, defined respectively as: $c_\rho = e_\rho/T_\rho$, $r_\rho = en_\rho/T_\rho$, and $v_\rho = n_\rho/T_\rho$, where:

1) T_ρ is the number of roles identified by the analyst;
2) e_ρ, en_ρ, n_ρ are the roles that can be traced to the initial roles, the roles that express a refinement of the original ones, and the new roles, respectively.

Based on these rate variables, we want to see the interval, for which we can state that, for a confidence level of 0.95 ($\alpha = 0.05$), the rate variable X is comprised between a certain lower bound L and a certain upper bound U. To this end, for each rate variable X, we consider two null hypotheses to be rejected, namely:

	c	r	v
median	0.3485714286	0.5431162652	0.1739467563
average	0.3374381865	0.5768463618	0.1651505205
stdev	0.1102982621	0.1068841256	0.1055906819
	c_ρ	r_ρ	v_ρ
median	0.7083333333	0.1833333333	0.25
average	0.668023088	0.1666955267	0.2347041847
stdev	0.2100729279	0.1834055271	0.1872880153

TABLE I
MEDIAN, AVERAGE, AND STANDARD DEVIATION FOR THE RATES.

- $H^L_{NULL}(X)$: The X rate is lower than or equal to L
- $H^U_{NULL}(X)$: The X rate is greater than or equal to U

Our goal is to identify the values of L and U for each rate variable. We perform statistical hypothesis testing by means of the one sample T-test—or Wilkoxon signed-rank test if the normality hypothesis is not satisfied by the data. We repeat each test multiple times with decreasing values of L, for H_L and increasing values of U, for H_U until a $p-value$ lower than 0.05 is reached. At each iteration, we increase (decrease) by 0.01, so our results are consider a granularity of 1%. By performing these iterations, the first values of L and U for which we obtain $p-value \leq 0.05$, are those that we consider as statistically significant for our final statements, which will have the following form:

- The X rate is comprised between L and H

The labeling procedure and theme extraction performed by the 1st and 2nd authors have produced additional information that can help to answer RQ2 and RQ3. In particular, we are interested in analyzing the following indicators with respect to the original idea and between different analysts:

- **Categories and Roles Frequency**: percentage of user stories belonging to each category and role.
- **Emerging Categories, Themes, and Roles** in the analysts' user stories and their frequency.

IV. Execution and Results

A. RQ1: How much is the difference in terms of documented requirements and roles with respect to initial ideas?

In Fig. 3 and Fig. 4, we report the plots of the values of the different rates for each analyst. Instead, in Table I we report the statistics in terms of median, mean values, and standard deviation. We see that in general the higher values are observed for the refinement rate, followed by the conservation rate and by the novelty. Conversely, for roles, conservation rate dominates over novelty and refinement ones. Looking at Fig. 3 and Fig. 4, we intuitively see that variations for each rate are quite high from an analyst to the other. This suggests that each individual analyst produces different user stories in terms of content, and thus, depending on the analyst, different systems may be developed. In some cases, analysts lean more towards the refinement of user stories in the existing categories, while in others focus on completely novel features, as one can observe, e.g., for analysts 7 and 16. In other cases, e.g., analysts 2 or 12, the elicitation process tends to be more conservative, with less space for creativity.

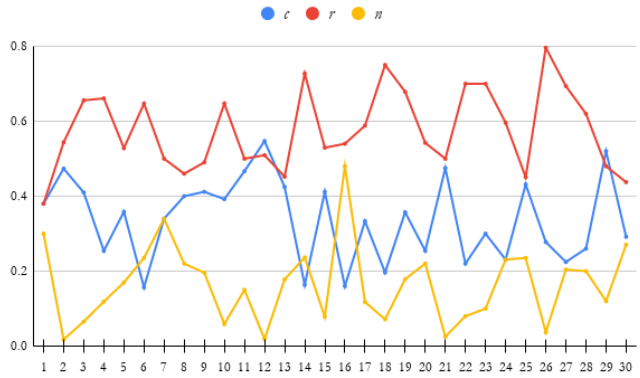

Fig. 3. Values of the different rates for each analyst.

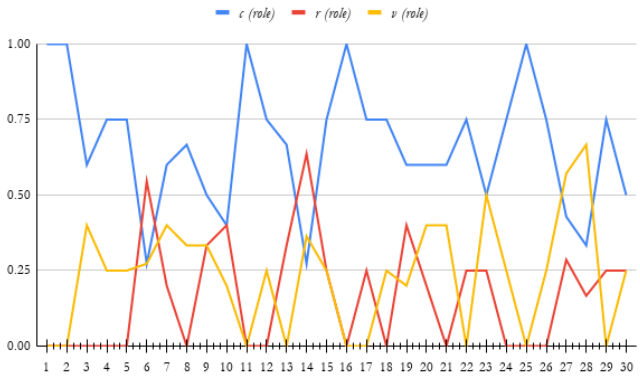

Fig. 4. Values of the different role rates for each analyst.

In the following, we answer RQ1 by means of hypotheses testing.

a) Conservation Rate: The Shapiro-Wilk Test for r indicates that the data can be considered to follow a normal distribution—more specifically *p-value* = 0.34, indicating that we cannot reject the hypothesis stating that data are normally distributed. Thus, a one-sample t-test is conducted. We consider values of μ_0 (i.e., the reference value for testing the hypotheses) in the set $[0.3, 0.31, \ldots, 0.38]$, which are in a neighborhood of the observed sample mean (0.34). By varying the values, and performing hypothesis testing we can identify the values of μ_0 that allow us to (1) reject or not reject the null hypothesis $H^L_{NULL}(c)$ for the lower bound (i.e., the null hypothesis is: *The conservation rate is lower or equal than* μ_0), (2) reject or not reject the null hypothesis for the upper bound (i.e., *The conservation rate is greater or equal than* μ_0). The procedure is carried out in a similar way for the other rates, for which, in the following, we only give the reference values.

b) Refinement Rate: Test for normality is passed with *p-value* = 0.32. The T-test is therefore applied, for values of μ_0 in $[0.54, 0.55, \ldots 0.61]$, in a neighborhood of the average of 0.58.

μ_0	t	p-value	μ_0	V	p-value
Consevation Rate			**Role Conservation Rate**		
0.38	**-2.1135**	**0.02164**	**0.75**	**57.5**	**0.03827**
0.37	-1.617	0.05836	0.74	170	0.09985
0.31	1.3625	0.09176	0.6	223	0.0517
0.3	**1.8591**	**0.03659**	**0.59**	**323**	**0.03133**
Refinement Rate			Role Refinement Rate		
0.61	**-1.6989**	**0.05002**	**0.21**	**152**	**0.04772**
0.6	-1.1865	0.1225	0.2	146	0.09383
0.55	1.3757	0.08972	0.13	234	0.4917
0.54	**1.8882**	**0.03452**	**0.12**	**318**	**0.03823**
Novelty Rate			Role Novelty Rate		
0.2	**-1.8077**	**0.04052**	**0.3**	**152**	**0.04897**
0.19	-1.289	0.1038	0.29	168	0.09276
0.14	1.3046	0.1011	0.17	303	0.0738
0.13	**1.8233**	**0.03929**	**0.16**	**321**	**0.03435**

TABLE II
RESULTS OF THE TESTS PERFORMED TO IDENTIFY THE UPPER AND LOWER BOUNDS OF THE DIFFERENT RATES.

c) Novelty Rate: Test for normality is passed with *p-value* = 0.09. The T-test is therefore applied, for values of μ_0 in $[0.13, 0.14, \ldots, 0.20]$, as average is 0.17.

d) Role Conservation Rate: Test for normality is not passed (*p-value* = 0.03) and thus we apply the Wilkoxon signed rank test, for μ_0 in $[0.59, 0.60, \ldots, 0.75]$, as average is 0.67.

e) Role Refinement Rate: Test for normality is not passed (*p-value* = 0.0002). The Wilkoxon signed rank test is applied, for μ_0 in $[0.12, 0.13, \ldots, 0.21]$, considering that average is 0.17.

f) Role Novelty Rate: Test for normality is not passed (*p-value* = 0.0.006). The Wilkoxon signed rank test is applied, for μ_0 in $[0.16, 0.17, \ldots, 0.3]$, as average is 0.24.

In Table II we report the values identified for upper and lower bounds for each rate, together with the *p-values* obtained. For reference, we also report the closest neighboring values tested for which significance is not obtained, and the associated *p-values*.

Based on the tests results, the following statements can be given, with *p-value* ≤ 0.05:

- The conservation rate c is between 30% and 38%
- The refinement rate r is between 54% and 61%
- The novelty rate n is between 13% and 20%
- The role conservation rate c_ρ is between 59% and 75%
- The role refinement rate r_ρ is between 12% and 21%
- The role novelty rate n_ρ is between 16% and 30%

B. RQ2: What is the relevance given to the categories of requirements and roles with respect to initial ideas?

To answer RQ2, we analyze how the relevance given to each category and each role (i.e., the percentage of the stories belonging to a certain category or role) change in the analysts' stories with respect to the original ones. This analysis will allow us to identify what is important for the analysts and to reflect on the meaning of these preferences.

Category	Mean	Std. Dev.	Min	Max
customers	26.27475	9.510693	6	54.71698
facilities	8.030648	5.240704	0	18
personnel	15.48588	7.368808	1.960784	29.82456
camp	9.845647	7.938568	0	30.13699
communication	31.79153	10.34756	9.589041	48.21429

TABLE III
DESCRIPTIVE STATISTICS FOR THE CATEGORY DISTRIBUTIONS IN THE PARTICIPANTS USER STORIES.

*1) **Analysis of Categories**:* Through the process described in Section III-B, the researchers identified 5 different categories:

- **Administrative procedure related to customers** (labeled as *customers*): this category includes features such as registration to a camp, creation of new campers and parents profiles, modification, and elimination of profiles.
- **Management of facilities** (*facilities*): this category includes the tracking of the facilities' status both in terms of usage and maintenance needs and the management of the inventory.
- **Administrative procedure related to personnel** (*personnel*): This category includes features related to assigning tasks to workers and evaluating them. In its more general interpretation, it can also include the ability to create and modifying personnel profiles and other administrative needs.
- **Individual camp management** (*camp*): This category includes features such as scheduling activities within a specific camp, managing its participants, and dealing with additional planning details.
- **Communication** (*communication*): This category includes everything related to communication from one-to-one messaging and broadcasting to a specific category of users to posting information online and in social media.

a) Initial Categories Distribution: In the original stories, and, thus, in the mind of the fictional customer, great relevance is given to the administrative activities on the customer side (i.e., stories belonging to *customers*), such as "*As a camp administrator, I want to be able to add campers so that I can keep track of each individual camper*". In particular, 64.15% of the total number of stories belong to this category, and the remaining is distributed among the other categories as follows: 5.66% belong to *facilities*, 9.43% to *personnel*, 7.55% to *camp*, and 15.09% to *communication*.

b) Categories Distribution in the Analysts' Stories: The category distribution of the analysts' stories is different with respect to the original stories. Indeed, looking at Table III, we observe that the mean of the percentage of stories belonging to *customers* is lower than half of the one in the original stories, while the mean of *communication* is double of the value for the original stories.

The box plots in Fig. 5, in which the black diamonds represent the percentage for each category in the original stories, show that the relevance given to *facilities*, *personnel*, and *camp* of the original stories is in line with the one given by

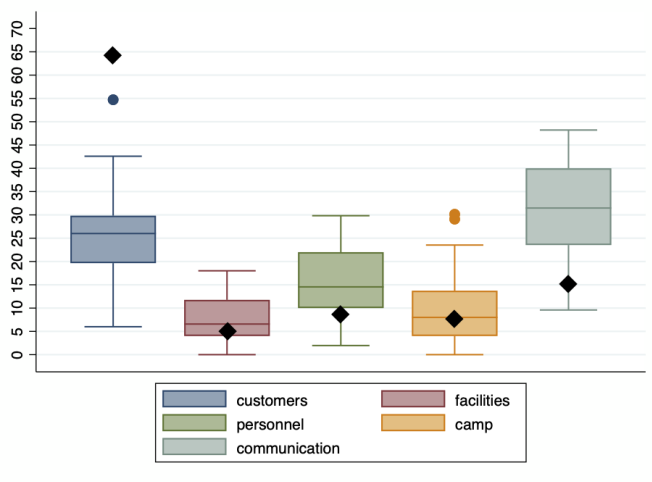

Fig. 5. Box plots of the distributions for each category, measured as the percentage of user stories in the category.

the analysts. Instead, for *customers* and *communication* there are strong differences. This suggests that the analysts focused their attention on aspects that were not the original focus of their customer.

c) Emerging Themes in the Analysts' Stories: Our analysis suggests not only a change of distribution among categories, but also a change of interest in the themes within the existing categories. For example, in *customers*, the original stories contain stories about managing consent forms. This theme is almost completely overseen by the analysts, with only a fourth of them marginally considering it. Conversely, more than 35% of the analysts include in the features a payment system. This is not part of the customer ideas, and the only feature related to payment in the original user stories was about storing the information about payments ("*As a parent I want to be able to see if I made all the necessary payments*") rather than having the option to pay for the camp through the app.

Another emerging theme is related to social media as a form of *communication*. This part is completely missing in the customer ideas and in the original user stories, in which the only story implicitly related to social media is "*As a parent I want to be able to share any photos the camp has taken of my child*". Despite this initial lack of interest in social media of the customer, more than 65% of the analysts explicitly mention social feed and posting (e.g., "*As the camp administrator I can tag students in posts about their accomplishments on our social feed so that their parents can share in their accomplishments*") in their stories after asking question about this topic in their interviews.

*2) **Analysis of Roles**:* Similar considerations can be done about the perspective considered in imagining the system, and, thus, the roles used in the user stories. In the original set of user stories, there are three roles: *camp administrator*, *parent*, and *camp worker*.

a) Initial Roles Distribution: The majority of stories is dedicated to the camp administrator (66.04% of the user

Role	Mean	Std. Dev.	Min	Max
Administrator	42.42049	15.73949	11.32076	83.33334
Worker	26.6092	12.8854	0	62.7451
Parent	16.3276	9.798852	0	32
New role	14.6427	14.11918	0	54

TABLE IV
DESCRIPTIVE STATISTICS FOR THE ROLE DISTRIBUTIONS IN THE PARTICIPANTS USER STORIES.

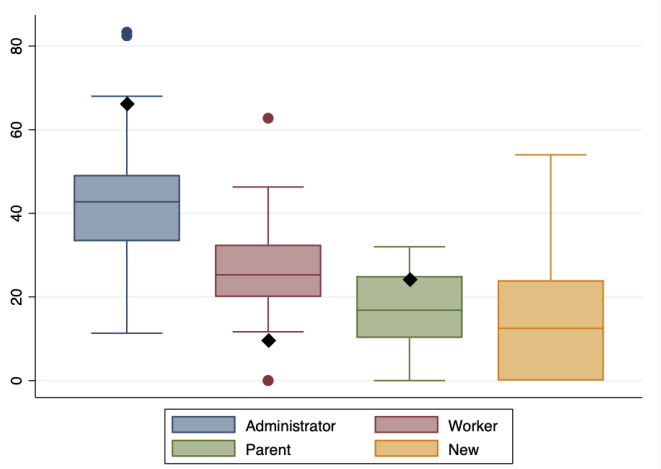

Fig. 6. Box plots of user story distribution for each role, measured as the percentage of user stories that used that role.

stories), followed by the parents of the participants (24.53%) and the camp worker (9.43%). This also helped the preparation of the fictional customer who was acting as the camp administrator as he had most of the available stories looking at the system to be developed from his perspective.

b) Roles Distribution in the Analysts' Stories: The analysts had only the chance to talk with the administrator and, nevertheless, 30% of them in their stories were dedicated to other roles, and only 16.67% of them had more than half of the stories focused on the administrator.

Table IV shows the descriptive statistics for the distribution of the analysts' user stories among the different roles. It is interesting to observe that the mean for camp workers is much higher than the one on parents. This could be connected to the decreasing attention to the category *customers* in the analysts' requirements. Notice that, while all the analysts considered the role of camp administrator, two of them did not consider the role of the camp worker, and four did not consider the role of parents.

As shown in the box plots in Fig. 6, in which the black diamonds represent the percentage for each role in the original stories, only the outlier focused more on the camp administrator role than the original stories. Similarly, the relevance given to the parents' perspective is in general much smaller than in the original stories.

The impact of new roles is limited (mean of 14.64%). All the analysts considered at least 2 of the original roles, and 80% of them considered all the 3 original roles. However, the relevance given to the perspectives and roles in the analysts'

user stories is considerably different than in the original ones.

C. RQ3: what are the emerging categories and roles and what is their impact on the requirements?

To answer RQ3, we analyze the new categories and new roles, their recurrence among analysts, and their weight within the set of stories of the analysts who included them. This analysis provides insight on what is generated by the analysts' expertise, background, preparation, and analysis.

*1) **Analysis of New Categories**:* In their stories, every analyst included new categories up to a maximum of 7 with a mean of 3.43. The more recurrent new categories are reported in Fig. 7. In addition, there are 6 other categories used by a single analyst that could not map over any other existing or newly created ones.

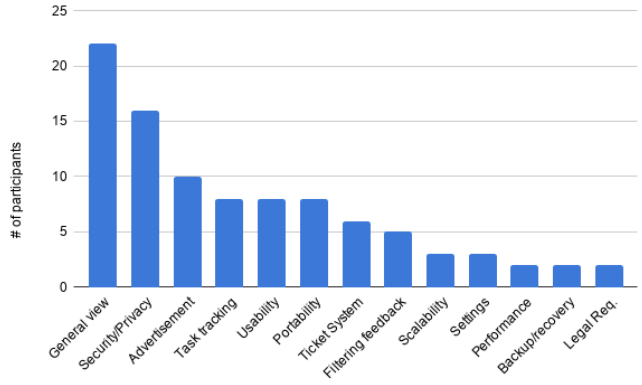

Fig. 7. Emerging categories distribution.

Table V reports descriptive statistics on new categories. The more recurrent new category, used by 73.33% of the analysts, is *General view*. All the features that act on the company, its high-level activities, and properties as a whole belong to this category. An example is "*As a business owner, I want to be able to analyze the data that is entered into the system so that I can see trends in the information that I receive.*". The mean usage is 6.18% (around 3.23 stories) with std. dev. 5.48 as the relevance given varies from 1.69% to 18.34%.

Another strongly emerged category is *Security/Privacy*. More than half of the analysts included requirements belonging to this category with an impact on the total of user stories that varies from 1.78%, one story, to 7.84%, 4 stories. An example of such a story is "*As an employee, I want this app to not have access to my personal phone data so that what I have on there stays private*". Notice that there is a fundamental difference between this category and the previous one. "General view" includes functional requirements that describe global operations at a company level, while "Security/Privacy" includes nonfunctional requirements that many of the analysts autonomously decided to investigate during their conversation with the customer.

Among the other nonfunctional requirements categories that emerged in the analysis are *Usability* and *Portability*, which

Category	Obs.	Mean	Std. Dev.	Min	Max
General view	22	6.182205	5.16035	1.666667	18.36735
Sec./Privacy	16	4.792629	2.893629	1.785714	10
Advertisement	10	2.650138	1.04596	1.666667	4
Tracking Sys.	8	8.383646	5.26647	1.960784	16
Usability	8	6.640473	5.218612	2	5.218612
Portability	8	4.213585	2.075225	2	8.333333

TABLE V
DESCRIPTIVE STATISTICS SUMMARIZING THE IMPACT OF THE MOST RECURRENT EMERGING CATEGORIES.

Category	Obs.	Mean	Std. Dev.	Min	Max
Participants	17	9.058824	4.762754	4	19
Visitors	8	3.625	2.13391	1	7
Users	5	5	4	2	12
Consultants	3	3.666667	3.785939	1	8

TABLE VI
DESCRIPTIVE STATISTICS SUMMARIZING THE IMPACT OF THE MOST RECURRENT EMERGING ROLES.

have both been considered by 8 participants (26.67% of the total). An example that belongs to the *Usability* category is "*As an employee, I wish for this app to be simple to use so that our older staffers can still use it*". Notice that, when present, usability is highly considered with an average of 6.64% stories. *Portability* has a lower impact on the stories of the analysts who use it (4.21%). An example in this category is "*As a camp administrator, I want users to be able to use the system on all platforms so that no users are excluded from seeing what the campsites have to offer.*".

In synthesis, we observe that new categories emerged for every analyst and some of them are highly predominant both in terms of number of analysts who considered them and in impact on the stories of the analysts who considered them. The new categories are almost equally divided into new functionalities and nonfunctional requirements.

*2) **Analysis of New Roles**:* Differently from the case of categories, not all the analysts added new roles to their stories, but still a big percentage did it (70%, which correspond to 21 analysts over 30). 5 new roles emerged from the analysis, namely, in order of frequency, participants (17), visitors (8), consultants (3), system administrators (2), and investors (1). Moreover, 5 analysts introduced the concept of "user" as a generic role. Table VI reports descriptive statistics on new roles (excluding system administrators and investors that appear just once in the stories of 2 and 1 analysts, respectively). The most frequent among these roles, participants, refers to the children participating in the camps and thus assumes that they all will have access to the system. When used, this role has a high impact with a mean of 9.06 (standard deviation of 4.762754). Notice that the analyst who used it more dedicated 35.85% of the stories to this role which represented the most significant role in the analyst's set.

Summarizing, we observe that many analysts consider additional roles with respect to the ones in the original set.

V. DISCUSSION

The analysis of the data collected in our study shows interesting trends in the analysts' behavior and provide food

for thoughts and suggestions on how to move forward in this research. In particular, our analysis shows that during the elicitation process there is an evolution of the original idea. While this evolution might be (partially) driven by the three Cs [16], our data show that it does not only goes in the direction of completing the existing information, but often changes the focus of the requirements and the roles, adds new functionalities that were not part of the initial ideas, and introduces nonfunctional requirements. In the following, we will discuss the takeaways of the data analysis performed in this study and what we have learned in the process of conducting it.

a) New Features: The idea of including new functionalities and perspectives might have been risen in the initial preparation of the analysts before the first interview and/or in the analysis phase between interviews. Before the interview, the analyst has read the brief description provided by the customer and prepared some initial questions. While some of these question have been generated using the provided description, others might have been originated from the analyst's knowledge and experience in the field and in designing software systems. Between interviews, the analysts reasoned on the collected information using their preferred tool (mockups, UML-like diagrams, textual notes). This analysis might have created curiosity and doubts on certain aspects of the system and suggested new functionalities for it.

Analysts' expertise in designing software system might suggest them functionalities that are usually included in similar systems, but were not part of the original idea.

This is in general a positive aspect, but, as observed in our analysis, it might be the cause of neglecting functionalities desired by the customer. So it is important that the analysts include both aspects and, if needed, prioritize the customer's primary needs.

Notice also the use of diagrams in the analysis phase might lead the analysts to make early design decisions that might "force" the introduction of new requirements. Making design choices too early in the development process is recognized by experts as one of the main analysts' mistakes [39].

Analysts might include requirements that are driven by early design choices.

This confirms the ambivalent role of models and prototyping tools that on one side can support the analysts in showing ideas to the customers [40] and on the other they might also cause harm by moving the focus on the model itself and anticipating decision that can compromise the quality of the produced system [37].

b) Nonfunctional Requirements: A considerable part of the new requirements introduced by the analysts are nonfunctional requirements (e.g., privacy, usability) and this could be be a sign of the analysts' maturity.

Analysts' training and expertise can drive them to ask questions related to nonfunctional aspects.

Indeed, as shown in [41], expert analysts use their belief structure for the construction of mental lists of topics they want to cover in the elicitation process and, among them, are nonfunctional requirements that require expertise and so are often not mentioned by the customers.

c) Distribution over Roles: In our study, there were three main roles, which were in general maintained in the analysts' user stories. However, while in the original specification (and thus in the mind of the customer) the mainly considered perspective was the one of the administrator, the majority of the analysts re-balanced the focus among different perspectives and introduced also additional ones. This is very interesting especially because the analysts only spoke with the camp administrator.

Analysts tend to successfully identify the main users of a system, but they might try to represent their perspective without talking to them or collecting enough data.

This can be mainly be caused by the desire of the analysts to show a complete understanding of the domain and that they value the most relevant stakeholders [42]. It is curious to observe that the majority of the analysts considered the perspective of the camps' participants and imagined them as users. Camps are usually available for a large range of ages including children attending kindergarten who most likely will not own a mobile device and so will not be a direct user of the system.

When analysts are not knowledgeable of the domain or do not prepare enough, they might include requirements that are inappropriate for the context.

d) Independently Created Requirements: While part of the new requirements in the analysts' user stories can be traced back to questions that the analysts asked during the interviews, some of the user stories were not originated in the interaction with the customer—as confirmed by the customer who also performed the analysis. This means that they were created by the analysts in their analysis. In this creative process, the analysts use their experience and the domain knowledge acquired in the elicitation process, including their preparation to it, to create new stories or manipulate the collected information. For example, the content of "*As a government employee, I want to be able to see safety records of equipment being used, so that I can effectively report potential issues*" was never discussed with the customer and the story has been derived by knowledge about the domain possessed by the analyst. The requirement is highly pertinent and represents some knowledge that the camp administrator most likely has and did not consider to share with the analysts.

While acquiring domain knowledge in preparation of an elicitation activity or analyzing the data collected, analysts might identify topics that can help to disclose tacit knowledge.

This add to the findings of Ferrari *et al.* [43] where the authors suggest that tacit knowledge can be disclosed

exploiting ambiguity, and is in line with Hadar *et al.* [44] as domain knowledge helps to identify the correct direction. However, this example shows also one of the negative impact of tacit knowledge, i.e., it makes difficult to listen to the users [44]. In this case, this problem is even more extreme because the analyst used their knowledge without checking it with the stakeholders.

e) Evaluating User Stories: A broader impact of our work is the possible use of our labelling process in RE education to support a systematic and objective evaluation of the students' work.

The labeling spreadsheet could be used as a tool to evaluate user stories in educational contexts.

In particular, the evaluation sheet can help to assess if the students sufficiently covered the area of interest of their customer, and identified some new interesting requirements. This is analogous to measure the "completeness" dimension in the Quality User Story Framework [45]. However, in this case, we are not focusing on the completeness of the project *per se*, but of the user stories with respect to the original ideas and provided information, similarly to the concept of *backward completeness* [10]. Overall, this process can be automated by means of NLP techniques leveraging text similarity (see, e.g., Abbas *et al.* for pointers to recent measures in RE [46]).

VI. Threats to Validity

a) Construct Validity: The main constructs of interest are "initial ideas", and "documented needs". These are somewhat vague concepts, which are strictly related to the notion of *pre-requirements* introduced by Hayes *et al.* [17]. We *reify* the intuitive meaning of these concepts through a form that is well defined in the literature, namely *user stories*. In analysing initial ideas—and their counterpart, documented needs—through the user story representation, we consider multiple rate variables that are related to subjective evaluations. We mitigate subjectivity threats through the triangulation process described in Sect. III-B.

The list of user stories that was used to reify initial ideas was written beforehand by other authors, and not by our fictional customer. In other terms, what we actually use is his understanding of the ideas of someone else, which is not the construct we are interested in. However, even in real contexts, the subject who speaks with a requirement analyst is often someone who has collected different ideas from other subjects. Furthermore, if we would ask our customer to write down user stories for his own initial ideas, the act of writing would be a further bias, as the documented ideas would not be "initial" anymore. Given these limitations, also due to the complexity of the considered problem, we argue that our design represents an acceptable trade-off between construct validity and potential bias that could be introduced with a different design.

b) Internal Validity: The initial user stories were studied by the fictional customer, and represent his interpretation of these initial ideas, which cannot be considered entirely faithful. Furthermore, given the repetition of interviews involving the same customer, a learning bias could not be entirely avoided. However, the fictional customer is a trained research assistant and was asked to provide uniform interviews to the different analysts. He was asked to stick as much as possible to the initial user stories, while limiting further elaboration to the questions asked by analysts, as it would happen in a real setting. We believe that this is a sufficient countermeasure in the adopted context to guarantee uniform treatments to all analysts.

c) External Validity: Given that this research is a *laboratory experiment*, intended as a study oriented to *identify the relationship between several variables or alternatives under examination* [47], external validity is inherently limited. A limited number of user stories was used compared to a real system, and a single system was considered which is not necessarily representative of all types of systems. Different results may be obtained in a more realistic settings—however, some variables could hardly be measured, especially if we wish to guarantee statistical significance. We used students instead of professionals in our experiment, as it is common and widely accepted in software engineering research [48]–[50]. Most of the involved participants also work as professional developers or analysts.

VII. Conclusion

Requirements start from unexpressed ideas to be transformed into documented needs, and eventually realised into (*satisfied by*) specifications and products. Understanding the evolution of requirements at their early stages can contribute to address potential issues that may emerge later in the development process. In this paper, we study early requirements evolution, when they pass from initial customer ideas into documented needs. To this end, we perform a laboratory experiment involving 30 subjects, and we quantitatively and qualitatively evaluate this evolution. Our study shows that the elicitation and documentation process can be regarded as a *co-creation activity* involving the contribution of analysts and customers alike. The process does not only complete the initial ideas, but often changes the relevance given to specific requirements and roles, adds new functionalities, and introduces nonfunctional requirements. Our work contributes to theory in RE, and should be regarded as an empirically grounded starting point to better understand the transition from ideas into products.

At this stage, we looked into the elicitation and documentation process as a black box, and did not investigate the impact of the different means (mock-ups, prototypes) used for the analysis, and their relationship with the results. Future work oriented to unpack this black-box will address this issue. In addition, we also aim to observe the further evolution of the requirements into the actual products developed, to have a complete trace of their transformation. This study can serve as a baseline to support future automated software engineering methods oriented to manage requirements evolution.

REFERENCES

[1] D. Zowghi and C. Coulin, "Requirements elicitation: A survey of techniques, approaches, and tools," in *Engineering and managing software requirements*. Springer, 2005, pp. 19–46.

[2] O. Dieste and N. Juristo, "Systematic review and aggregation of empirical studies on elicitation techniques," *IEEE Transactions on Software Engineering*, vol. 37, no. 2, pp. 283–304, 2010.

[3] D. M. Fernández, S. Wagner, M. Kalinowski, M. Felderer, P. Mafra, A. Vetrò, T. Conte, M.-T. Christiansson, D. Greer, C. Lassenius *et al.*, "Naming the pain in requirements engineering," *Empirical software engineering*, vol. 22, no. 5, pp. 2298–2338, 2017.

[4] E. Kamsties, D. M. Berry, B. Paech, E. Kamsties, D. Berry, and B. Paech, "Detecting ambiguities in requirements documents using inspections," in *Proceedings of the first workshop on inspection in software engineering (WISE'01)*, 2001, pp. 68–80.

[5] M. Bano, D. Zowghi, A. Ferrari, and P. Spoletini, "Inspectors academy: Pedagogical design for requirements inspection training," in *2020 IEEE 28th International Requirements Engineering Conference (RE)*. IEEE, 2020, pp. 215–226.

[6] G. S. Walia and J. C. Carver, "Using error abstraction and classification to improve requirement quality: conclusions from a family of four empirical studies," *Empirical Software Engineering*, vol. 18, no. 4, pp. 625–658, 2013.

[7] S. F. Tjong and D. M. Berry, "The design of sree—a prototype potential ambiguity finder for requirements specifications and lessons learned," in *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer, 2013, pp. 80–95.

[8] V. Gervasi, A. Ferrari, D. Zowghi, and P. Spoletini, "Ambiguity in requirements engineering: towards a unifying framework," in *From Software Engineering to Formal Methods and Tools, and Back*. Springer, 2019, pp. 191–210.

[9] C. Arora, M. Sabetzadeh, and L. C. Briand, "An empirical study on the potential usefulness of domain models for completeness checking of requirements," *Empirical Software Engineering*, vol. 24, no. 4, pp. 2509–2539, 2019.

[10] A. Ferrari, F. dell'Orletta, G. O. Spagnolo, and S. Gnesi, "Measuring and improving the completeness of natural language requirements," in *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer, 2014, pp. 23–38.

[11] A. Ferrari, P. Spoletini, B. Donati, D. Zowghi, and S. Gnesi, "Interview review: detecting latent ambiguities to improve the requirements elicitation process," in *2017 IEEE 25th International Requirements Engineering Conference (RE)*. IEEE, 2017, pp. 400–405.

[12] F. Salger, "Requirements reviews revisited: Residual challenges and open research questions," in *2013 21st IEEE International Requirements Engineering Conference (RE)*. IEEE, 2013, pp. 250–255.

[13] S. D. P. Harker, K. D. Eason, and J. E. Dobson, "The change and evolution of requirements as a challenge to the practice of software engineering," in *[1993] Proceedings of the IEEE International Symposium on Requirements Engineering*, 1993, pp. 266–272.

[14] M. Crowne, "Why software product startups fail and what to do about it. evolution of software product development in startup companies," in *IEEE International Engineering Management Conference*, vol. 1, 2002, pp. 338–343 vol.1.

[15] A. M. Grubb and M. Chechik, "Looking into the crystal ball: Requirements evolution over time," in *2016 IEEE 24th International Requirements Engineering Conference (RE)*, 2016, pp. 86–95.

[16] D. Zowghi and V. Gervasi, "On the interplay between consistency, completeness, and correctness in requirements evolution," *Information and Software Technology*, vol. 45, no. 14, pp. 993–1009, 2003.

[17] J. H. Hayes, G. Antoniol, and Y.-G. Guéhéneuc, "Prereqir: Recovering pre-requirements via cluster analysis," in *2008 15th Working Conference on Reverse Engineering*. IEEE, 2008, pp. 165–174.

[18] L. V. G. Carreño and K. Winbladh, "Analysis of user comments: An approach for software requirements evolution," in *2013 35th International Conference on Software Engineering (ICSE)*, 2013, pp. 582–591.

[19] W. Maalej, Z. Kurtanović, H. Nabil, and C. Stanik, "On the automatic classification of app reviews," *Requirements Engineering*, vol. 21, no. 3, pp. 311–331, 2016.

[20] D. Pagano and W. Maalej, "User feedback in the appstore: An empirical study," in *2013 21st IEEE international requirements engineering conference (RE)*. IEEE, 2013, pp. 125–134.

[21] E. Guzman, M. Ibrahim, and M. Glinz, "A little bird told me: Mining tweets for requirements and software evolution," in *2017 IEEE 25th International Requirements Engineering Conference (RE)*, 2017, pp. 11–20.

[22] J. Khan, L. Liu, L. Wen, and R. Ali, *Crowd Intelligence in Requirements Engineering: Current Status and Future Directions*, 03 2019, pp. 245–261.

[23] I. Morales-Ramirez, F. M. Kifetew, and A. Perini, "Speech-acts based analysis for requirements discovery from online discussions," *Inf. Syst.*, vol. 86, pp. 94–112, 2019. [Online]. Available: https://doi.org/10.1016/j.is.2018.08.003

[24] R. Ali, F. Dalpiaz, P. Giorgini, and V. Silva Souza, "Requirements evolution: From assumptions to reality," vol. 81, 01 2011, pp. 372–382.

[25] J. Robertson, "Requirements analysts must also be inventors," *IEEE Software*, vol. 22, no. 1, pp. 48–, 2005.

[26] J. Lemos, C. Alves, L. Duboc, and G. Rodrigues, "A systematic mapping study on creativity in requirements engineering," *Proceedings of the ACM Symposium on Applied Computing*, 03 2012.

[27] N. Maiden, S. Jones, I. K. Pitts, R. Neill, K. Zachos, and A. Milne, "Requirements engineering as creative problem solving: A research agenda for idea finding," 09 2010, pp. 57–66.

[28] R. Sternberg and C. U. Press, *Handbook of Creativity*. Cambridge University Press, 1999. [Online]. Available: https://books.google.com/books?id=d1KTEQpQ6vsC

[29] L. Nguyen and G. Shanks, "A framework for understanding creativity in requirements engineering," *Information and Software Technology*, vol. 51, no. 3, pp. 655–662, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584908001250

[30] N. Maiden and S. Robertson, "Integrating creativity into requirements processes: experiences with an air traffic management system," in *13th IEEE International Conference on Requirements Engineering (RE'05)*, 2005, pp. 105–114.

[31] A. Herrmann, L. Mich, and D. M. Berry, "Creativity techniques for requirements elicitation: Comparing four-step epmcreate-based processes," in *2018 IEEE 7th International Workshop on Empirical Requirements Engineering (EmpiRE)*, 2018, pp. 1–7.

[32] M. Mahaux, L. Nguyen, O. Gotel, L. Mich, A. Mavin, and K. Schmid, "Collaborative creativity in requirements engineering: Analysis and practical advice," in *IEEE 7th International Conference on Research Challenges in Information Science (RCIS)*, 2013.

[33] P. P. Grube and K. Schmid, "Selecting creativity techniques for innovative requirements engineering," in *2008 Third International Workshop on Multimedia and Enjoyable Requirements Engineering - Beyond Mere Descriptions and with More Fun and Games*, 2008, pp. 32–36.

[34] R. B. Svensson and M. Taghavianfar, "Selecting creativity techniques for creative requirements: An evaluation of four techniques using creativity workshops," in *2015 IEEE 23rd International Requirements Engineering Conference (RE)*, 2015, pp. 66–75.

[35] J. Horkoff, N. Maiden, and J. Lockerbie, "Creativity and goal modeling for software requirements engineering," ser. CC '15. ACM, 2015, p. 165–168.

[36] J. Horkoff and N. A. M. Maiden, "Creative leaf: A creative istar modeling tool," in *Proceedings of the Ninth International i* Workshop*, ser. CEUR Workshop Proceedings, vol. 1674. CEUR-WS.org, 2016, pp. 25–30.

[37] A. Davis, O. Dieste, A. Hickey, N. Juristo, and A. M. Moreno, "Effectiveness of requirements elicitation techniques: Empirical results derived from a systematic review," in *14th IEEE International Requirements Engineering Conference (RE'06)*, 2006, pp. 179–188.

[38] F. Dalpiaz. Requirements data sets (user stories). Mendeley Data, V1. [Online]. Available: http://dx.doi.org/10.17632/7zbk8zsd8y.1

[39] A. Whittenberger. (2014) The top 8 mistakes in requirements elicitation. [Online]. Available: https://www.batimes.com/articles/the-top-8-mistakes-in-requirements-elicitation.html

[40] A. Sutcliffe and P. Sawyer, "Requirements elicitation: Towards the unknown unknowns," in *2013 21st IEEE International Requirements Engineering Conference (RE)*, 2013, pp. 92–104.

[41] M. Pitts and G. Browne, "Stopping behavior of systems analysts during information requirements elicitation," *J. of Management Information Systems*, vol. 21, pp. 203–226, 06 2004.

[42] C. Pacheco and I. Garcia, "A systematic literature review of stakeholder identification methods in requirements elicitation," *Journal of Systems and Software*, vol. 85, no. 9, pp. 2171–2181, 2012.

[43] A. Ferrari, P. Spoletini, and S. Gnesi, "Ambiguity as a resource to disclose tacit knowledge," in *2015 IEEE 23rd International Requirements Engineering Conference (RE)*, 2015, pp. 26–35.

[44] I. Hadar, P. Soffer, and K. Kenzi, "The role of domain knowledge in requirements elicitation via interviews: An exploratory study," *Requir. Eng.*, vol. 19, no. 2, p. 143–159, 2014.

[45] G. Lucassen, F. Dalpiaz, J. M. E. M. van der Werf, and S. Brinkkemper, "Forging high-quality user stories: Towards a discipline for agile requirements," in *2015 IEEE 23rd International Requirements Engineering Conference (RE)*, 2015, pp. 126–135.

[46] M. Abbas, A. Ferrari, A. Shatnawi, and E. Paul, "Is requirements similarity a good proxy for software similarity? an empirical investigation in industry," in *The 27th International Working Conference on Requirements Engineering: Foundation for Software Quality*, 2021. [Online]. Available: http://www.es.mdh.se/pdf_publications/6142.pdf

[47] K.-J. Stol and B. Fitzgerald, "The abc of software engineering research," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 27, no. 3, pp. 1–51, 2018.

[48] M. Svahnberg, A. Aurum, and C. Wohlin, "Using students as subjects-an empirical evaluation," in *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*, 2008, pp. 288–290.

[49] D. Falessi, N. Juristo, C. Wohlin, B. Turhan, J. Münch, A. Jedlitschka, and M. Oivo, "Empirical software engineering experts on the use of students and professionals in experiments," *Empirical Software Engineering*, vol. 23, no. 1, pp. 452–489, 2018.

[50] I. Salman, A. T. Misirli, and N. Juristo, "Are students representatives of professionals in software engineering experiments?" in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, vol. 1. IEEE, 2015, pp. 666–676.