Systematizing requirements elicitation technique selection

Dante Carrizo [a,*], Oscar Dieste [b], Natalia Juristo [b]

[a] Atacama University, Avda Copayapu 485, Copiapó, Chile
[b] Technical University of Madrid, Boadilla del Monte, Madrid 28660, Spain

A B S T R A C T

Context: This research deals with requirements elicitation technique selection for software product requirements and the overselection of open interviews.
Objectives: This paper proposes and validates a framework to help requirements engineers select the most adequate elicitation techniques at any time.
Method: We have explored both the existing underlying theory and the results of empirical research to build the framework. Based on this, we have deduced and put together justified proposals about the framework components. We have also had to add information not found in theoretical or empirical sources. In these cases, we drew on our own experience and expertise.
Results: A new validated approach for requirements technique selection. This new approach selects tech-niques other than open interview, offers a wider range of possible techniques and captures more require-ments information.
Conclusions: The framework is easily extensible and changeable. Whenever any theoretical or empirical evidence for an attribute, technique or adequacy value is unearthed, the information can be easily added to the framework.

1. Introduction

The requirements engineering (RE) process is composed of [1]: requirements elicitation, analysis, specification, validation and management. Requirements elicitation covers the capture and discovery of stakeholder needs. Its aim is to identify information determining what features the software system should have. This activity is carried out recurrently throughout the requirements stage. It often takes place iteratively and is interlinked with other activities in this stage. Each of these iterations for capturing key information about requirements is called an elicitation session. Each session requires preparation, execution and later analysis. Requirements engineers have to select which elicitation technique to use in each session. According to Zowghi and Coulin [2], the preparation of an elicitation session entails: (i) understanding the application domain by exploring the policy, organizational and social aspects of the current environment, as well as system and development constraints; (ii) identifying requirement sources, that is, users, experts and any relevant project, process and system, as well as existing documentation like manuals, forms, reports and earlier requirement specifications; (iii) analyzing stakeholders and identifying key representative users; (iv) selecting techniques, approaches and tools for use in the requirements process; and (v) eliciting requirements from stakeholders and other sources. Our research focuses on the task of selecting techniques for eliciting requirements (iv).

Software engineers tend to choose a technique to apply on one of the following grounds [22]: it is the only technique they are acquainted with; it is their favorite technique for all situations; they are using a methodology that prescribes a particular technique; or they guess that the technique is effective under the existing circumstances. This subjective decision can bias the elicitation results, degrade the quality of the output requirements, and, ultimately, have an impact on the quality of the final software product [9].

In practice, when analysts set out to determine a software system's requirements, they very often use only one technique, interviews, to capture information, even though they are probably acquainted with several other methods [3]. This could be because they are unaware of the benefits of each technique, there is no methodological guidance for the elicitation process or it is standard procedure: in many cases, an elicitation method or technique is chosen not for its features or strengths, but simply on the grounds of history or familiarity [4].

* Corresponding author. Tel.: +56 52 206809; fax: +56 52 206683.
E-mail addresses: dante.carrizo@uda.cl (D. Carrizo), odieste@fi.upm.es (O. Dieste), natalia@fi.upm.es (N. Juristo).

According to Beyer and Holtzblatt [38], there is evidence, however, that traditional interviews are inadequate for the development of systems today. They propose techniques that reflect the typical master-apprentice relationship in which the stakeholder is a master of his job but does not necessarily have teaching skills to communicate actions to the developer. In addition, watching and experiencing stakeholders doing their job reveals issues, details, structure and work strategies that are difficult to capture clearly using interviews. Other authors, such as Maiden and Rugg, indicate that probably more than one elicitation technique may be necessary to get the full range of requirements for most complex software systems [16]. In short, it must be recognized that there are contextual conditions that hinder the use of interviews to elicit requirements, such as stakeholder inaccessibility, schedule coordination problems, problem domain complexity, stakeholder profile diversity, disagreement on the problem to be solved, deficient communication skills or cultural differences between requirements engineers and stakeholders.

Some reviews account for tens of elicitation techniques [5]. Many of these techniques have been imported from fields like cognitive psychology, anthropology, sociology and linguistics [6], and have been successfully used in knowledge engineering [7] and, lately, software engineering. Most requirements engineers are nonetheless unfamiliar with this range of techniques and miss the chance of optimizing requirements elicitation. This can be considered as another side of the well-known breach between theory and practice [8].

The differences between elicitation techniques are characteristically very large, the inference being that some are likely to perform better in some situations than in others. An elicitation technique's intrinsic features dictate how it should be applied, but is not enough to ascertain its adequacy. Methodological support to help requirements engineers to select the most adequate technique for the conditions that they face during elicitation could turn out to be very useful.

This research tackles the problem of selecting elicitation techniques based on the idiosyncrasy of each technique and the particularities of the development context. An appropriate choice of technique optimizes the productivity of the information captured in the elicitation sessions. This makes it possible to generate a more correct and complete specification which then results in a quality final product through a process with fewer holdups [9].

We aim to set up a framework that in practice helps requirements engineers to choose a suitable technique to elicit the key information at any time during development project requirements elicitation.

Section 2 discusses work related to the problem. Section 3 shows an overview of the proposed framework. Section 4 determines and defines the relevant contextual attributes for elicitation and their values, which is the initial information required to analyze technique fitness. Section 5 establishes the values of elicitation technique fitness for each key contextual attributes value. Section 6 presents the procedure for selecting the best techniques for a particular project. Section 7 shows an example in which the framework is applied. Section 8 presents the evaluation of the proposed framework through two experiments. Finally, Section 9 discusses the results and presents the final conclusions.

2. Related work

The requirements research field has expanded over recent decades. Even so, there is little research addressing how to support analysts in decision making on the choice of elicitation techniques, methods or tools for capturing information for specifying requirements.

Several papers describe elicitation techniques and provide some instructions on their use [5,10,11,6]. There are also some empirical studies comparing elicitation techniques. The shortage of experiments and the non-uniformity of the experimental conditions, variables and techniques that they study make it difficult to infer the application conditions for elicitation techniques [12]. Some research on elicitation techniques, mainly empirical, comes from the knowledge engineering area. Although these studies were carried out before 2000, their results continue to be valid.

After reviewing about three hundred articles and thirty books related to requirements engineering, we found that the first studies designed to prescribe techniques based on contextual attributes were conducted only ten years ago. This dearth of studies is a sign of how little research has focused on the selection of elicitation techniques. Elicitation technique selection is the central goal of only ten papers. Table 1 summarizes these studies. It uses a set of criteria for evaluating and comparing the related work. These criteria are:

- Discipline in which the proposal is applicable: proposals have been made in knowledge engineering and software engineering (some comparative studies of techniques have been conducted in the fields of economics, marketing, psychology). Proposal objectives are discipline dependent. Thus, the goal in knowledge engineering is to elicit knowledge from experts. The goal in software engineering is to capture relevant information for requirements specification. This area is more important for our purposes, because our proposal aims to support novice requirements engineers.
- Scope that the proposal aims to cover: proposals may be designed to help select techniques for broader processes like the software development process or the requirements process or specifically for requirements elicitation. The techniques may differ depending on the scope of use. For our study, the elicitation techniques have the distinctive trait of being user interaction intensive. On this ground, we are interested in the techniques used in the requirements elicitation activity.
- Type of information on which the proposal is based: proposals can be based on the expert opinions of their authors and/or on empirical studies. We believe that this type of research to support technique selection should be based primarily on empirical evidence as its recommendations are more reliable. The proposals should evolve towards this goal.
- Number of elicitation techniques covered: requirements engineers reckon with more and more alternatives for capturing requirements information, close to fifty at present. However, the proposals should consider, at least initially, a sizeable number of the most popular techniques.
- Types of elicitation process contextual factors accounted for: the contextual attributes of the elicitation activity have been grouped under five factors that influence technique effectiveness (elicitor, informant, problem domain, solution domain, and elicitation process). Proposals should, at least initially, consider all these contextual factors.
- Specification of the contextual attribute values: the possible values of the contextual attributes at any time during a project may or may not be available in the proposals. This is important as the workability of the method depends on the possibility of quickly and easily determining such values.
- Evolvability of the proposal: proposals may or may not offer facilities for updating the method. However, we consider that the determination of other influential attributes, the addition of other techniques or the incorporation of evidence on their effectiveness is essential for the validity and use of the proposal over time.

Table 1
Summary of related work.

Aspect	Proposals Value	Kim and Courtney [13]	Dhaliwal and Benbazat [14]	Byrd, Cossick and Zmud [15]	Maiden and Rugg [16]	Robertson and Robertson [17]	Davis and Hickey [18]	Lauesen [19]	Zowghi and Coulin [2]	Jiang and Eberlein [20]	Kheirkhah and Deraman [21]
Area	Software engineering			x	x	x	x	x	x	x	x
	Knowledge engineering	x	x								
Target	Global process									x	x
	Elicitation process	x	x	x	x	x	x	x	x		
Basic information	Theoretical	x	x	x	x	x	x	x	x	x	x
	Empirical										
	Both										
Technique coverage	Low (0–5)	x									x
	Medium (6–14)			x	x				x	x	
	High ($\geqslant$15)		x			x	x	x			
Contextual factors	Analyst/developer		x		x		x				
	Informant		x		x		x	x			
	Problem domain	x	x	x	x	x	x	x			x
	Solution domain		x				x				x
	Process		x	x	x		x		x	x	x
Attribute value	Available					x		x	x		
	Unclear	x	x	x	x		x			x	x
Upgrading	Easy	x		x	x		x	x	x		
	Hard		x			x				x	x
Adequacy concept	Product quality		x								
	Relevance	x		x	x	x	x		x	x	
	Effectiveness		x					x			
Practical Supporting Tool	Available										
	Probably						x				
	No	x	x	x	x	x		x	x	x	x

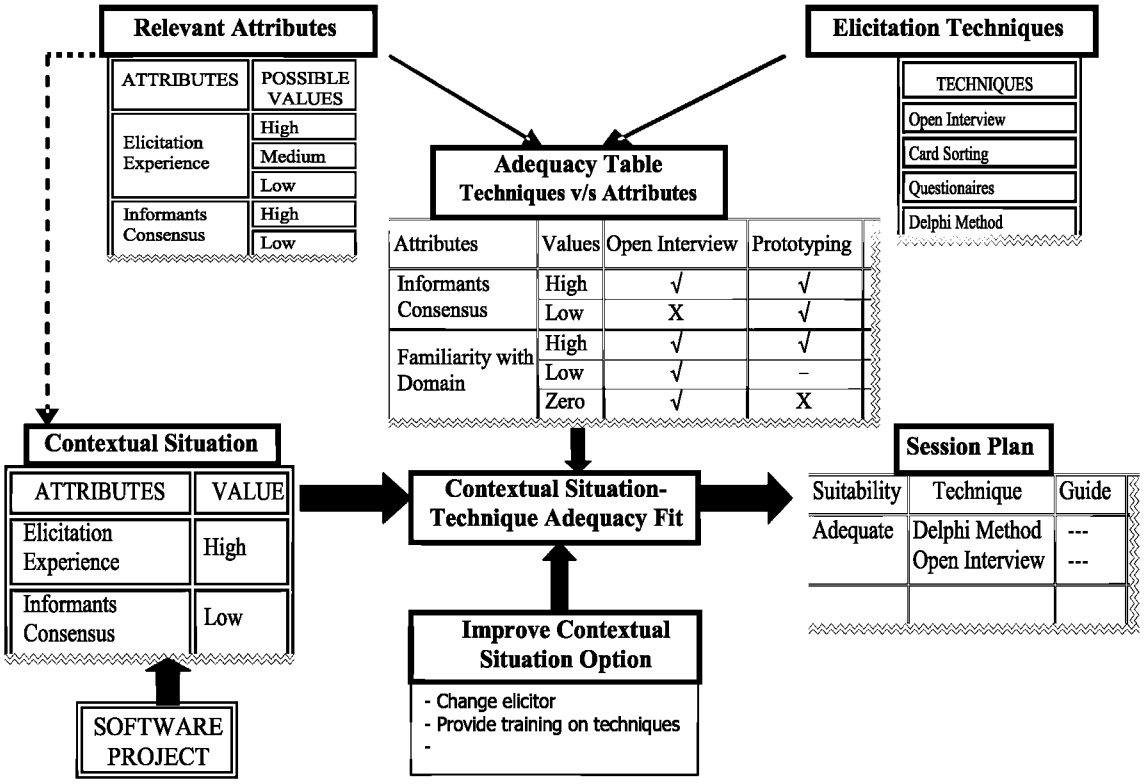

Fig. 1. Proposed technique selection method.

- Variable defined for measuring technique adequacy: software product quality, effectiveness, productivity or subjective measures of relevance are all examples of possible technique adequacy variables reported in the literature. We believe that work should have an objective metric that measures the effectiveness of the technique to capture a certain amount of relevant information to shape the requirements.
- Support software: research can end with a theoretical research proposal or advance towards a tool that helps practitioners decide about how to go about elicitation sessions. This is relevant to reduce the gap between practice and theory.

The above criteria are useful for defining the *ideal* proposal. This method should deal with the problem of selecting elicitation techniques to capture relevant information for putting together the software system requirements. It should be based preferably on empirical information and cover a wide range of techniques (over 15) and contextual factors (at least 4). The method should help to easily identify the possible values of the contextual factor attributes. Additionally, it should output the most effective techniques for an elicitation session, ordered or rated quantitatively to help requirements engineers with selection. Finally, the method should be implemented in a tool to help use and upgrade the method with new techniques or attributes.

We can see that Lauesen's [19] is the proposal that comes closest to this model. This proposal fails to meet three criteria: basic information, which is based solely on the authors' opinion; contextual factors, which consider only the informant and problem domain attributes; and practical supporting tool, as it does not provide support software. Davis and Hickey's proposal [18] falls short on four criteria: basic information, which is theoretical; attribute values, which are unclear; adequacy measure, which is subjective relevance; and practical supporting tool, which it does not appear to provide. Another four proposals are unsuccessful on five criteria [14,16,17,2].

Consequently, there is no elicitation requirements technique selection method that holds the minimal properties that we believe to be necessary to warrant consideration as a significant and systematic aid for requirements engineering practitioners. This is the gap that our research sets out to fill.

3. Approach overview

According to our vision of the elicitation process, a systematic way of selecting a technique in order to execute each elicitation session would be to follow the procedure shown in Fig. 1:

1. Identify *Contextual Situation*: determine the values associated with the attributes or features of the development context where the elicitation techniques are to be applied.
2. Situation-Technique *Adequacy Fit*: evaluate the adequacy of each technique for the problem context configuration.
3. Obtain a *Session Plan*: select one or more techniques in order of priority and application for the following session or sessions.

Contextual conditions (e.g., the amount and type of available information), vary in the course of the elicitation process, leading to a new selection starting from step 1.

So, elicitation technique selection relies on the following knowledge (also shown in Fig. 1):

- Which attributes influence the effectiveness of the elicitation techniques? The usual line of attack taken in earlier works to discover which attributes influence technique effectiveness has been to identify the characteristics of the elicitation techniques. We believe, though, that the important factor for technique selection is the context in which the technique is applicable. In other words, we sidestep the intrinsic technique characteristics and focus on the characteristics of the context in which the techniques are applicable. The intrinsic technique

characteristics are obviously what decide whether or not a technique is adequate for use in a particular situation, but they do not have to be formally specified in the selection. To be more precise, our approach primarily establishes the contextual characteristics that make the difference between technique effectiveness.

- How adequate is each technique for each key attribute value? Running experiments that turn up the required technique effectiveness information is not a feasible short-term option for settling the technique adequacy issue. Experimentation [23] on all the techniques and all the configurations of possible attribute values would be a long-term goal. In other words, the empirical information required by an elicitation technique selection method is not available at present. Pragmatically speaking, then, we have gathered the available theoretical and empirical information to build the measure of adequacy of each technique for each value of each attribute. For attribute-technique value combinations where no such information was available, we have applied our own experience and reasoning.
- How will the candidate techniques be evaluated to choose the most adequate technique for use in the next elicitation sessions? Our proposal collects the particular technique adequacy values for all attribute values for comparison and outputs a possible elicitation plan (the most adequate elicitation technique or techniques) for the session. If necessary, guidelines will be provided to improve the contextual situation and be able to use more elicitation techniques or improve proposal adequacy.

Each of these components will be developed in the following sections.

4. Determining key attributes

Before deciding on the most adequate elicitation technique to apply at any time during a project, it is first necessary to determine the contextual attributes that influence technique selection, as reported in Section 3. The procedure enacted to determine the key contextual attributes was (see Fig. 2):

1. Review the related literature in search of theoretical proposals and empirical studies directly reporting or inferring contextual attributes that possibly influence elicitation technique effectiveness.
2. Identify and group the candidate attributes, that is, classify the identified attributes by the factor to which they belong. We have defined five factors (elicitor, informant, problem domain, solution domain and elicitation process)
3. Analyze the candidate attributes by acceptance and rejection criteria (assessability, instrumentability and theoretical justifiability).
4. Determine framework attributes, that is, set up the final attributes by possibly merging, changing the name or adding attributes.

To identify the contextual attributes, we reviewed two types of studies: proposals of elicitation technique selection frameworks and empirical studies comparing elicitation technique effectiveness. We found six elicitation technique selection frameworks and 11 empirical studies comparing technique effectiveness. The framework proposals define attributes that their authors suggest are relevant for elicitation techniques selection. We use a broader definition for empirical studies, namely, any way of gaining knowledge by means of direct and indirect observation or experience. So, we consider empirical research ranging from higher quality studies (like randomized parallel or crossover intervention trials) to lower quality studies (non-randomized studies, longitudinal studies, case studies, etc.) [39]. We focus especially on experiments run to demonstrate a difference of effectiveness between certain techniques after altering a contextual attribute.

From these sources, we output a preliminary set of 34 possible influential attributes. Framework attributes can be aggregated or modified whenever any new evidence appears.

We grouped the identified attributes by five factors:

- *Elicitor:* Development team agent that elicits information on software system requirements. Other names, such as analyst or requirements engineer, are used in the literature to refer to this role.
- *Informant:* Human agent that has information regarding requirements. Informants can be customers, users and, generally, any software development stakeholders. Non-human sources have not been considered in this research.
- *Problem domain:* The problem that the software system under development is to solve.
- *Solution domain:* The software product being developed to solve the problem.
- *Elicitation process:* The requirements gathering process.

To decide whether an attribute should be selected to be part of the final attribute set we evaluated whether or not the attribute can be defined and justified using the assessability, instrumentability or theoretical justifiability criteria.

Assessability (A): Possibility of establishing ratings for the different attribute values. In other words, this criterion answers the question How possible is it to set a finite and manageable universe of admissible attribute value? The possible values are: low (L), medium (M) and high (H). For example, there are no proposals or mature theories about the universe of possible values for the Problem-Solving Methods attribute. Neither is it clear whether these values could be exclusive to one development or whether several

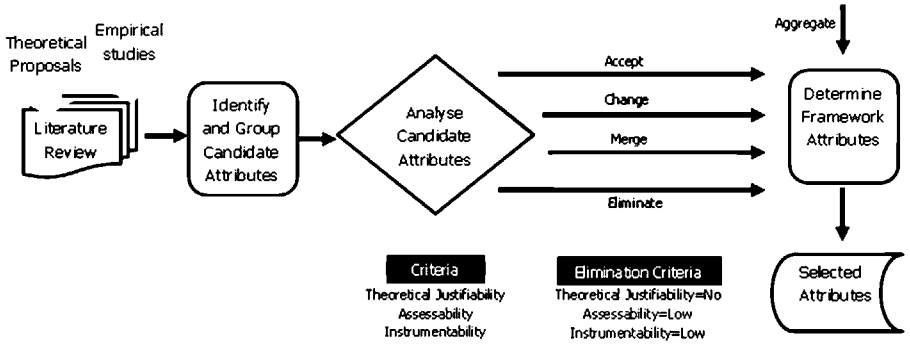

Fig. 2. Procedure followed to determine framework attributes.

could be used in any particular case. Therefore, we assign an assessability value of low to this attribute. Other examples of low assessability attributes are product type and development methodology. It is not easy to agree on a finite and universal number of possible values for product type or development methodology. Consensus is unlikely to be reached in the short term due to the diversity of opinions on this point. Not having a finite number of attribute values vitiates the evaluation function that will compare the contextual situation against technique adequacy. On the other hand, there is an overwhelming set of alternatives for defining the Type of Information attribute value depending on the ontology used. In this case, it is not a matter of values being hard to find, but of there being a lot of viewpoints for consideration. Assessability, then, does not appear to be simple. Thus, we assign an assessability value of medium to this attribute. Finally, distinguishing the values that the type of phenomena attribute could take looks to be a straightforward task. Researchers that propose this attribute define yes and no values to indicate whether or not the phenomena are observable. In this case, the criterion value would be high.

Instrumentability (I): Possibility of assigning a value to the attribute during a development project. In other words, this criterion answers the question How possible is it to determine a specific value for the attribute from the established universe of values? The possible values are: low (L), medium (M), and high (H). For example, for the perceived structuredness attribute, it is unlikely to be easy to assign a structuredness value to a particular problem to be solved by software development unless the information to be processed is known at the early elicitation stages. It is hard to generate a list of problems that are related to a particular structuredness level. Additionally, although there are domains with a set structuredness, the software systems to be developed in this domain can be very wide ranging. Therefore, we assign an instrumentability value of low to this attribute. Other examples of low instrumentability are cognitive problems and cognitive skills attributes. Cognitive problems or cognitive skills will require the administration and analysis of psychological tests to potential informants for assessment. However, this could turn out to be a nuisance and too costly for selecting a technique. A solution (evaluate cognitive problems or cognitive skills) should not be more costly and complex than the problem that it is to solve (select elicitation techniques). On the other hand, it can be hard in some developments to establish the value for the domain fields attribute. A transport system for a mining company can be considered to belong to the transport field and the mining field. So, we assign an instrumentability value of medium to this attribute. Finally, in the case of the requirements engineering experience attribute, requirements engineer experience would be easy to determine for either of the chosen rating types, time or number of activities, as the engineer's curriculum would be fully available throughout. Thus, we assign an instrumentability value of high to this attribute.

Attributes such as complexity of the problem would have a high assessability value, as we can establish the universe of possible values as complex, medium, simple, although their instrumentability would be low since it is hard to determine which value such attributes would take in a particular project. By contrast, the development methodologies attribute would have a low assessability value because it is hard to establish a specific number of possible accepted methodologies, although their instrumentability would be high because it would be easy to determine which methodology is used in a particular project.

Table 2
Analysis of influential contextual attributes.

Factor	Attributes	Proposing authors	Assessability	Instrumentability	Theoretical justifiability	Action
Elicitor	Requirements engineering experience (elicitation)	[22,14,24,25]	H	H	Y	C
	Technical knowledge of (training in) elicitation methods	[14]	H	H	Y	C
	Knowledge of (familiarity with) domain	[14]	H	H	Y	C
	Experience with elicitation methods (technique)	[14]	H	H	Y	C
	Cognitive problems	[15]	M	L	P	E
Informant	Number of users (per session)	[16]	H	H	Y	C
	Number of experts	[26]	H	H	Y	M
	User involvement (interest)	[22,24]	H	M	Y	C
	Location/accessibility		H	H	Y	+
	Availability of time		H	H	Y	+
	Expertise	[27,14]	H	H	Y	A
	Cognitive styles (articulability)	[14]	M	M	Y	C
	Personality variables	[14]	M	L	P	E
	Cognitive problems (consensus among informants)	[15]	M	M	Y	C
	Cognitive skills	[28]	M	L	P	M
Problem domain	Type of phenomena	[16]	M	L	P	E
	Type of information (to be elicited)	[29,30,13,16]	M	M	Y	C
	Type of heuristics	[31]	H	M	P	M
	Level of available information		M	M	Y	+
	Domain fields	[32]	L	M	P	E
	Perceived structuredness	[14,13]	M	L	N	E
	Problem definedness		H	M	Y	+
	Uncertainty	[14,33]	H	M	P	M
	Type of tasks	[7,34,28]	L	M	Y	E
	Domain entities	[15]	H	H	Y	M
	Confusedness	[33]	H	M	P	M
	Size	[13]	H	L	N	E
	Complexity	[13,35]	H	L	Y	E
SD[a]	Product type	[36]	L	M	Y	E
	Problem-solving methods	[14]	M	L	N	E
Elicitation process	Purpose of requirements	[16]	M	H	N	E
	Constraints (time)	[16]	H	H	Y	C
	Process time		M	H	Y	+
	Development methodology	[14]	L	H	P	E

[a] Solution domain.

Theoretical justifiability (TJ): Possibility of finding a justification for the attribute influencing elicitation technique effectiveness. The possible values for this criterion are: no (N), possibly (P), and yes (Y). For example, it is not clear for the purpose of requirements attribute whether the ultimate aim of the requirements might condition the use of one or other technique. The format of the requirements specification or documentation could vary depending on its purpose, but this does not mean that information extraction method will necessarily be different. In this case, we assign a theoretical justifiability value of no. Another example of an attribute with no clear justification is size. Problem size does not appear to make a difference to techniques, as elicitation takes place in sessions with time limits. If the problem is large, it can be simply construed as requiring more sessions. This way, there does not appear to be a clear justification for size being a relevant attribute for selecting the elicitation technique for use in the session. On the other hand, for the informant personality variables attribute, we might speculate that the intrinsic characteristics modeling each human being's personal and sociable behavior (such as sociability, empathy, sensibleness, introversion, etc.) could bear a relationship to the effectiveness of elicitation techniques, but this is hard to prove. In this case, we assign a theoretical justifiability value of possibly. Finally, for the type of tasks attribute, there is evidence in the literature that some techniques apparently work better for capturing information on certain types of diagnostic, planning, design tasks, etc. Therefore, we assign a theoretical justifiability value of yes to this attribute.

After analyzing and assessing these criteria for all the attributes, we make a decision about what to do with each attribute. These criteria are an objective and reasoned basis for deciding which attributes are likely to play a role in elicitation technique selection nowadays. If the proposed framework really is to be a useful guide for practitioners, the attribute values should be clearly distinguishable. Also value assignment in a particular development case should be fairly easy and quick.

The actions to be taken with respect to an attribute are (the acronyms between parentheses will be used later in Table 2):

- *Accept (A):* Approve as it is.
- *Eliminate (E):* For an attribute to be eliminated, its influence *should not be justifiable,* OR its assessability OR instrumentability should be *Low*.
- *Merge (M):* Fuse with another attribute because they are similar.
- *Change (C):* Modify name or description.

Apart from the above actions, we have decided to add (the respective symbol used in Table 2 is "+") some attributes that, based on our experience as practitioners and academics, appeared to be strangely absent from the evaluated papers. As the requirements community shares our implicit knowledge, many practitioners and academics will likewise feel the need to include these attributes. To be more precise, we found that Location/Accessibility, Availability of time, Level of available information, Problem definition and Process time had not been proposed as relevant attributes for elicitation technique selection. We think that Location/Accessibility, for example, clearly influences the elicitation process, as there are techniques designed to facilitate elicitation from informants that cannot attend a face-to-face session (e.g.: questionnaires). Likewise:

- Some techniques that are more applicable when the informant does not have much time to spare to participate in elicitation sessions or techniques (e.g. questionnaires vs. open interviews).
- The use of some techniques is conditioned by the fact that they require certain information (e.g. prototyping vs. open interview).
- Some techniques that require a clear understanding of the problem prior to the session, whereas others ease this understanding (e.g. open interview vs. Delphi method).
- Some techniques appear to work better at the start of the process than others (e.g. open interview vs. protocol analysis).

This procedure is far from ideal. We always run the risk of bias if opinions, even if they are based on experience, are used. Unfortunately, as discussed at more length in Section 5, the available empirical evidence is very limited, and the use of some sort of opinion-based recommendations is unavoidable. In fact, this paper pinpoints the empirical evidence that is missing. In any case, we firmly believe that this framework is just a starting point and not a final product. It is dynamic and must be updated with new emerging evidence as it is gathered.

Table 2 shows the analysis of influential contextual attributes, together with the sources of each attribute. For a detailed analysis, see the web appendix http://www.grise.upm.es/sites/extras/4/attributes.pdf.

As a result of this analysis to determine which attributes influence the selection of the most adequate techniques, 11 out of 29 candidate attributes were retained, although the names of most were modified appropriately (10 of 11 accepted attributes were renamed). Others were merged (6 of 29), and a few others were added (5). Briefly, we added five new attributes to the 11 accepted and modified attributes. This totaled 16 attributes: four for the elicitor factor, seven for the informant factor, three for the problem domain factor, zero for the solution domain and two for the elicitation process factor.

Finally, we established how the attributes can be rated in a real-life situation. Table 3 shows the selected set of attributes, the possible values of each attribute and the attribute description.

5. Elicitation technique adequacy

In this section we appraise the adequacy value of techniques for each value of the attributes that were established in Section 4. There are a good many techniques that are potentially useful for capturing requirements. To conform and check the feasibility of the framework, we chose an initial set of these techniques. However, others can be added over time, as can any new elicitation techniques.

We chose a total of fifteen techniques to set up the initial framework. We chose the most researched and commonly used elicitation techniques. Some of these techniques may have variants or different interpretations. We have used the most popular version of each technique. Appendix A gives a description of how we have construed each technique used in this paper, as well as the sources from which they were taken.

To determine technique adequacy, we reviewed the related literature to come up with a recommendation for technique use with each influential attribute. From the books and articles on requirements, we found two possible sources of information (see Table 4): expert opinion (light grey shading) and empirical results (dark grey shading). Expert opinion is a proposal or rational statement made by an expert in the field of elicitation techniques. We consider an expert to be a person with major scientific publications on elicitation techniques that reliably support their use prescriptions. They can, of course, be practitioners because they have had practical experience endorsing their claims. But they can also include academic experts that we know to have based their opinion on their huge experience. Empirical results are adequacy values or conclusions deduced from an empirical study.

Entries that are left blank indicate that there are no empirical or theoretical prescriptions on the use of the techniques for that attribute. In order to produce a comprehensive set of recommendations

Table 3
Selected influential contextual attributes.

Factor	Attributes	Description	Values	Description
Elicitor	Training in elicitation techniques	Elicitor's previous training and practice with each elicitation technique	High	Formal and practical training
			Low	Training without practice
			Zero	No knowledge at all
	Elicitation experience	Number of earlier projects in which the elicitor has carried out elicitation activities	High	More than 5 elicitation projects
			Medium	2–5 Elicitation projects
			Low	Less than 2 elicitation projects
	Experience with elicitation techniques	Number of earlier elicitation activities in which the elicitor has applied each technique	High	More than 5 technique applications
			Low	From 1 to 5 technique applications
			Zero	No application of the technique
	Familiarity with domain	Number of earlier projects in the domain carried out by or domain knowledge acquired by the elicitor	High	More than 2 projects or formal knowledge
			Low	From 1 to 2 projects or formal knowledge
			Zero	No knowledge at all
Informant	People per session	Number of individuals that can participate at the same time in the elicitation session	Individual	1 Individual
			Group	From 2 to 5 individuals
			Mass	More than 5 individuals
	Consensus among informants	Initial agreement among informants	High	Consensus
			Low	No consensus
	Informant interest	Informant's eagerness to participate in the elicitation sessions	High	Very interested
			Low	Not very interested
			Zero	Uninterested
	Expertise	Informant's expertise in the problem or work domain	Expert	More than 5 years in the domain or role
			Knowledgeable	From 2 to 5 years in the domain or role
			Novice	Less than 2 years in the domain or role
	Articulability	Informant's skill at explaining his or her knowledge	High	Explains knowledge very well
			Medium	Explains knowledge reasonably well
			Low	Does not explain knowledge clearly
	Availability of time	Time the informant has to spend on the sessions	High	Has enough time
			Low	Has less time than necessary
	Location/accessibility	Informant's physical location with respect to the elicitor	Far	In a different city from the elicitor
			Near	In the same city as the elicitor
Problem domain	Type of information to be elicited	Type of categorized information that the technique can elicit	Strategic	Elicits strategies, control, directives
			Tactical	Elicits processes, functions, heuristics
			Basic	Elicits concepts, attributes, elements
	Level of available information	Categorized information type that is available before the session	Upper	There is tactical/strategic information
			Lower	There is basic/tactical information
			Zero	There is no information
	Problem definedness	Clarity of the objectives and scope of the project	High	Well defined
			Low	Poorly defined
Elicitation process	Project time constraint	Relative time available in the project for applying the technique	High	Not enough time
			Medium	Enough time
			Low	More than enough time
	Process time	Pre-session stage of the elicitation process	Start	Elicitation of general definitions
			Middle	Elicitation of key requirements
			End	Elicitation of last information

of use to practitioners, we drew on our own experience to analyze the key characteristics of each technique with respect to the empty cells. This way, we deduced possible adequacy levels for the attribute/technique pairs about which no information was available.

Briefly, we obtained 158 expert opinions on how adequate a given technique is for a certain attribute value, and 19 results from empirical studies supporting the same number of prescribed uses described in the literature on technique adequacy.

For the 15 chosen techniques and the 16 selected attributes, totaling 240 required adequacy values, we found an expert opinion for 80 (33%) and empirical support for 17 (7%) of the required prescriptions.

For the other 143 cases (60%), no reference to their prescribed use was found. This result shows that, despite the pressing need for guidelines on elicitation technique selection, much more research still has to be done on the selection of the most adequate techniques for use during the elicitation process.

Table 5 shows the different technique adequacy levels, represented according to the following notation:

√: The technique is adequate for this attribute value. This means that the results of using this technique in the session would be better than for a technique with a lower adequacy level. Therefore, if the project has this attribute value, this technique should be given priority during selection.

−: The technique is indifferent for the attribute value in question. There is no guarantee that the results of applying this technique would be better than those obtained using others. While this technique is an option, more adequate techniques would be preferable.

×: The technique has a low adequacy level for the attribute value in question. This technique is not recommended for use under the circumstances described by this attribute value, because it is likely to produce worse results than other techniques.

Let us look, for example, at how to assess the articulability attribute for the protocol analysis technique:

Protocol analysis involves informants putting into words how they would reason out a proposed case. Consequently, session success depends on the informants' ability to express their reasoning. Informants that are very good at explaining what procedure they follow in certain situations will be able to describe quite clearly what actions and heuristics they apply. If informants have average expression skills, the results are likely to be equally as good, although some significant information may not be delivered. In

Table 4
Information about technique adequacy.

Factor	Attributes	Open-ended Interv.	Structured Interv.	Task Observation	Card Sorting/Ladd.	Questionnaires	Protocol Analysis	Repertory Grid	Brainstorming	Nominal Group T.	Delphi Technique	Participant Obs.	Prototyping	Focus Group	JAD Workshop	Scenarios/U. Cases
Elicitor	Training in Elicitation Techniques															
	Elicitation Experience															
	Experience with Elicitation Technique															
	Familiarity with Domain															
Informant	People per Session															
	Consensus among Informants															
	Informant Interest															
	Expertise															
	Articulability															
	Availability of Time															
	Location/Accessibility															
Problem Domain	Type of Information to be Elicited															
	Level of Available Information															
	Problem Definedness															
Process	Project Time Constraint															
	Process Time															

Expert Opinion
Empirical Result
None

this case, it is not so clear that the technique should be used. If informants find it difficult to express their knowledge, the effectiveness of the session will be at stake. Elicitors could capture wrong or shallow information that would not justify the high cost of the session. In this case, the technique appears to be inadequate.

Additionally, some researchers have related the adequacy of this technique to the experts' personality and skill at introspection and correctly verbalizing processes [34]. Similarly, in their review of elicitation techniques, Jitnah et al. claim that users that cannot satisfactorily describe what they do are likely to compromise the technique's success [37]. In an experiment on a number of elicitation techniques, Chao and Salvendy concluded that significant cognitive skills, including expressability, are associated with the protocol analysis technique [28].

Therefore, the protocol analysis technique is recommended for use with informants that have a high articulability score. If informant articulability is medium, the use of the technique is indifferent. This technique is not recommended for use with informants that have a low articulability rating.

Let us look at another example concerning the adequacy values for the familiarity with domain attribute and the questionnaires technique. In this case, there are neither prescriptions by researchers nor empirical evidence to substantiate adequacy values.

Previous information about the domain would appear to be a prerequisite for the use of questionnaires, as the questions that will be put to the informant have to be prepared by the elicitor beforehand. The questionnaire should have a clear goal. Strictly speaking, it can focus on any particular topic or on the clarification of several unrelated issues. In any case, a minimum knowledge of the domain by the elicitor provides a guarantee of the questionnaire being properly prepared. For this reason, elicitors that are familiar with the domain will find this instrument easier to prepare and administer.

If the elicitor is unfamiliar with the domain, the result of using this technique will depend on the information captured beforehand and on other conditions. For this reason, questionnaires are recommended for use if the elicitor has a high or low familiarity with the domain. If the elicitor is completely unfamiliar with the domain, the results of using the technique are likely to be uncertain, meaning that it is indifferent.

We analyzed the other attributes and techniques similarly, and the final values are listed in Table 5. For an exhaustive analysis, see http://www.grise.upm.es/sites/extras/4/adequacy.pdf. Thus, we provide an initial table of elicitation technique adequacy to be used by the selection procedure.

6. Selection of elicitation techniques

Now we have all that is necessary in order to propose a procedure for selecting the most adequate techniques for a particular elicitation session. Most approaches to elicitation technique selection are based on an attribute-technique adequacy table. Although this is a substantial aid, it has some shortcomings. With only an adequacy table, the decision on how to combine this information to choose the final technique is left to the engineer. A comprehensive solution to the selection problem should offer an elicitation plan containing more and less adequate techniques. Only then, and considering some other emerging conditions, will the requirements engineer be able to make a decision based on objective information about the technique to be applied in a particular session.

Table 5
Adequacy of elicitation techniques for contextual characteristics.

Factor	Attributes	Values	Open-ended interv.	Structured interv.	Task observation	Card sorting/ ladd.	Questionnaires	Protocol analysis	Repertory grid	Brain-storming	Nominal group T.	Delphi technique	Participant observ.	Proto-typing	Focus group	JAD workshop	Scenarios/ *U.* cases
Elicitor	Training in elicitation techniques	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	–	√	√	–	√	–	–	√	√	–	√	–	–	–	–
		Zero	–	–	–	×	–	×	×	–	×	×	–	–	×	×	–
	Elicitation experience	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Medium	√	√	√	√	√	–	√	–	√	√	√	√	√	–	√
		Low	–	–	√	–	–	×	–	–	√	√	√	–	–	×	–
	Experience with elicitation techniques	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Zero	–	–	√	–	√	–	–	–	√	–	–	–	–	×	–
	Familiarity with domain	High	√	√	–	√	√	–	√	√	–	√	–	√	√	√	–
		Low	√	√	–	√	√	√	√	√	–	√	√	–	√	√	–
		Zero	√	–	√	–	–	√	–	√	–	–	√	×	×	×	–
Informant	People per session	Individual	√	√	√	√	√	√	√	×	×	×	√	√	×	×	√
		Group	–	–	√	–	√	–	√	–	√	√	√	–	√	–	–
		Mass	×	×	–	×	√	×	√	√	√	√	√	–	√	√	–
	Consensus among informants	High	√	√	–	√	√	√	√	–	√	–	–	√	√	–	√
		Low	×	×	–	×	√	×	√	–	√	√	–	√	√	–	–
	Informant interest	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	–	–	√	√	√	–	√	–	√	√	–	√	√	–	√
		Zero	×	–	√	–	√	×	√	×	√	√	×	√	–	×	–
	Expertise	Expert	√	√	–	√	√	√	√	√	√	√	√	√	√	√	√
		Knowledgeable	√	√	√	√	√	–	√	√	√	√	√	√	√	√	√
		Novice	–	–	√	√	√	–	√	√	√	–	√	√	√	√	√
	Articulability	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Medium	√	√	√	√	√	–	√	√	√	√	√	√	√	√	√
		Low	–	–	√	√	√	×	√	×	×	√	–	√	×	×	–
	Availability of time	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	–	–	√	–	√	×	√	×	×	√	×	–	×	×	–
	Location/ accessibility	Far	–	–	–	–	√	×	√	×	×	√	×	×	×	–	×
		Near	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Problem domain	Type of information to be elicited	Strategic	√	–	–	×	√	–	×	√	√	√	√	–	√	√	–
		Tactical	√	√	√	×	√	√	×	√	√	√	√	√	√	√	√
		Basic	–	√	–	√	√	–	√	×	×	–	–	√	–	×	–
	Level of available information	Upper	√	√	√	×	√	√	×	–	–	√	√	√	√	√	√
		Lower	√	√	√	√	√	×	√	√	√	√	√	√	×	√	√
		Zero	√	×	√	×	×	×	×	√	√	×	√	×	×	√	√
	Problem definedness	High	–	√	√	√	√	√	√	√	√	√	√	√	√	×	√
		Low	√	–	√	×	√	×	×	√	√	√	–	√	√	√	–
Process	Project time constraint	High	–	–	×	√	√	×	√	×	×	×	×	×	×	×	√
		Medium	√	√	–	√	√	–	√	–	–	–	–	–	–	–	√
		Low	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	Process time	Start	√	–	√	√	–	×	–	√	√	–	√	–	√	√	–
		Middle	√	√	–	√	√	√	√	–	√	√	√	√	√	×	√
		End	√	√	×	×	√	–	–	×	×	–	×	–	–	×	–

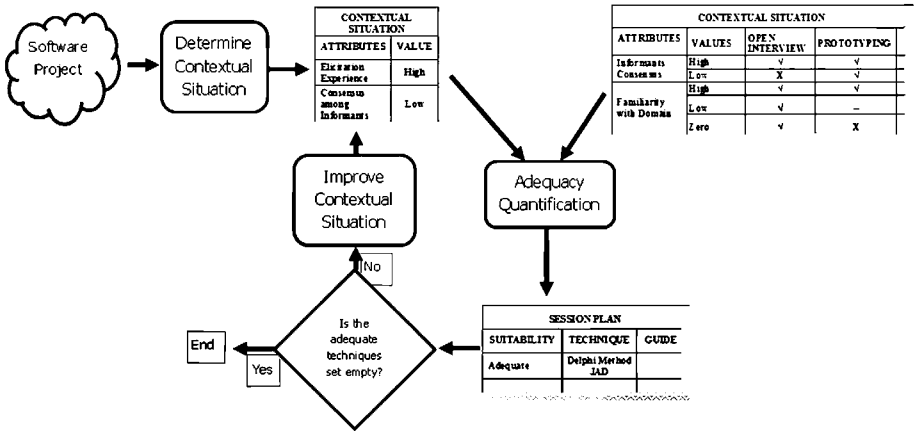

Fig. 3. Proposed elicitation technique selection procedure.

In our approach, we use the compiled technique adequacy table to specify the procedure for selecting the most adequate techniques for a software development project. This procedure consists of three main activities (see Fig. 3):

1. *Determine contextual situation:* The current conditions of a project define a contextual situation, that is, the particular values for the influential attributes.
2. *Quantify adequacy:* Each technique's adequacy level for the attribute values can be captured from the adequacy table. A quantified adequacy profile can be output for each technique by accounting for these adequacy levels and values. The adequacy values are expressed as the number of recommended ($\surd$), indifferent ($-$) and not recommended ($\times$) values. The result is a session plan, that is, a set of adequate and/or less adequate techniques for application in the next session.
3. *Improve contextual situation:* When there are no adequate techniques, some prioritized guidelines for modifying the attribute values are suggested, and the procedure starts over from Activity 1. Finally, the requirements engineer selects a technique for the next session from the output list of proposed techniques.

6.1. Determine contextual situation

The values of each attribute at any particular time constitute the contextual situation represented by a 16-tuple of baseline attribute/value pairs. The vector of values is $C = (c_1, c_2, c_3, \ldots, c_{16})$, where c_i is the value determined for attribute i in the current contextual situation.

Some attributes can have a default value (unvalued), either because the value has not yet been established or because it is intentionally defined as uninfluential. Also some attributes are likely to be multi-valued, that is, they have not a single value, but different values for each technique. This is the case, for example, of experience with elicitation technique, as the experience of a requirements engineer can vary among techniques.

For example, the vector C = (TET/{Zero, Low, Low, High,...}, EE/Low,...,), describes the following contextual situation. For the training in elicitation technique attribute, the elicitor has zero (value) training (attribute) in technique 1 (open interview), low training in technique 2 (structured interview), low training in technique 3 (task observation), high training in technique 4 (card sorting/laddering), and so on for the other techniques. The value of elicitation experience attribute is low, etc.

Table 6
Adequacy schema A structure.

Attribute	Values	Elicitation techniques			
		t_1	t_2	...	t_k
AT_1	$V_{1,1}$	a	a	a	a
	$V_{1,2}$	a	a	a	a
	$V_{1,3}$	a	a	a	a
...	...	a	a	a	a
	...	a	a	a	a
AT_{16}	$V_{16,1}$	a	a	a	a
	...	a	a	a	a
	$V_{16,m}$	a	a	a	a

Where t_k is requirements elicitation technique k, initially with $k = 1 \ldots 15$; AT_i is the influential attribute i, initially with $i = 1 \ldots 16$; $V_{i,j}$ is value j that can take attribute i, with $i = 1 \ldots 16$ and $j = \{2,3\}$; a is any of the defined adequacy values, that is, $a = \{-, \surd, \times\}$.

6.2. Quantify adequacy

The values for the influential project attributes are the input for the elicitation techniques adequacy quantification for a contextual situation. The quantification procedure uses the adequacy schema A. The structure of schema A is shown in Table 6 and represents the information listed in Table 5.

The adequacy quantification procedure can be represented by a selection function S_f that maps the contextual situation C to the adequacy schema A and outputs a session plan P. Plan P is a list containing the final adequacy quantifiers for each technique. These quantifiers are trinomials $\{a_1, a_2, a_3\}$ with values that match the sum of each technique adequacy value ($-$, $\surd$, $\times$) for all the attributes. Therefore:

$S_f(A,C) \rightarrow \{P | P$ is a list of techniques and adequacy quantifiers$\}$.
$P = \{(t_1, w_1), (t_2, w_2), \ldots (t_{15}, w_{15}) | w_k$ is a measure of the adequacy of technique k with $k = 1 \ldots 15\}$.
$w_k = \{\{a_1, a_2, a_3\}|$ where a_i is the sum of the adequacy values, $\{\surd, -, \times\}$ respectively, of all the attributes for technique k.

Table 7 is a view of Table 5 showing the adequacy values of four elicitation techniques for the values of the first two relevant attributes.

The selection function is implemented by three key functions: mapping function, aggregation function and prioritization function. First the contextual situation is mapped to the adequacy schema A, outputting the adequacy values of each technique for the attributes. Then, all the attributes are aggregated, that is, the

Table 7
Partial view of the adequacy table.

Factor	Attributes	Values	Open-ended interv.	Structured interv.	Task observation	Card sorting/ladd.
Elicitor	Training in elicitation techniques	High	√	√	√	√
		Low	–	√	√	–
		Zero	–	–	–	×
	Elicitation experience	High	√	√	√	√
		Medium	√	√	√	√
		Low	–	–	√	–

Table 8
Assigning adequacy values to quantifiers.

a	√	–	×
w	{1,0,0}	{0,1,0}	{0,0,1}

technique adequacy values of all the attributes are combined to output an overall rating. Finally, the elicitation techniques are prioritized according to their total calculated ratings.

6.2.1. Mapping function

The mapping function, which we denote as operator π, captures the adequacy values of each technique for a particular attribute from schema A determined considering the contextual situation value c_i. Using the detected adequacy values, each technique can be quantified by forming a pairwise list as follows:

$$\pi(c_i, A) = (t_1, w_{i,1}), (t_2, w_{i,2}), \ldots (t_k, w_{i,k})$$

For example, for the value of the first attribute c_1:

$$\pi(c_1, A) = (t_1, w_{1,1}), (t_2, w_{1,2}), \ldots (t_{15}, w_{1,15})$$

The values $w_{i,k}$ are the weights derived from the adequacy values a for the value c_i of attribute AT_i for the elicitation technique t_k, as shown in Table 8.

For the first value of the example contextual situation C = (TET/{Zero, Low, Low, High,...}, EE/Low,...,), the mapping in Table 7 would retrieve the adequacy values shown in grey in Table 9, that is:

$$\begin{aligned}&\pi([TET/\text{Zero}, \text{Low}, \text{Low}, \text{High}, \ldots], \text{Table 7})\\ &= (\text{Open Int.}, 0, 1, 0), (\text{Structured Int.}, 1, 0, 0),\\ &\quad (\text{T. Observation}, 1, 0, 0), (\text{Card Sorting/Ladd.}, 1, 0, 0)\end{aligned}$$

6.2.2. Aggregation function

The aggregation function, denoted by operator σ, combines the values of the different attributes for a contextual situation. The result is a list containing the sum of the quantifiers for each technique representing elicitation plan P.

$$\sigma(C) = \sigma(c_1, c_2, \ldots, c_{16}) = \pi(c_1, A) + \pi(c_2, A) + \cdots + \pi(c_{16}, A) =$$

$$= \{(t_1, w_{1,1}), (t_2, w_{1,2}), \ldots (t_{15}, w_{1,15})\} + \cdots + \{(t_1, w_{16,1}), (t_2, w_{16,2}), \ldots (t_{15}, w_{16,15})\}$$

$$= \{(t_1, w_{1,1} + \cdots + w_{16,1}), (t_2, w_{1,2} + \cdots + w_{16,2}), \ldots (t_{15}, w_{1,15} + w_{16,15})\},$$

that is:

$$P = \{(t_1, W_1), (t_2, W_2), \ldots (t_{15}, W_{15})\}$$

For example, the aggregation of values c_1 and c_2 of attributes AT_1 and AT_2, would be:

$$\sigma(c_1, c_2) = \pi(c_1, A) + \pi(c_2, A) =$$

$$= \{(t_1, w_{1,1}), (t_2, w_{1,2}), \ldots (t_{15}, w_{1,15})\} + \{(t_1, w_{2,1}), (t_2, w_{2,2}), \ldots (t_{15}, w_{2,15})\}$$

$$= \{(t_1, w_{1,1} + w_{2,1}), (t_2, w_{1,2} + w_{2,2}), \ldots (t_{15}, w_{1,15} + w_{2,15})\}$$

Table 10 shows the mapping of the first two attributes according to the example contextual situation C, that is:

$$\begin{aligned}&\sigma(\text{TET/Zero}, \text{Low}, \text{Low}, \text{High}, \ldots, \text{EE/Low})\\ &= \pi(\text{TET/Zero}, \text{Low}, \text{Low}, \text{High}, \ldots, \text{Table7}) + \pi(\text{EE/Low}, \text{Table 7}) =\\ &= (\text{Open Int.}, 0, 1, 0), (\text{Structured Int.}, 1, 0, 0),\\ &\quad (\text{T. Observation}, 1, 0, 0), (\text{Card Sorting/Ladd.}, 1, 0, 0)\\ &\quad + (\text{Open Int.}, 0, 1, 0), (\text{Structured Int.}, 0, 1, 0),\\ &\quad (\text{T. Observation}, 1, 0, 0), (\text{Card Sorting/Ladd.}, 0, 1, 0)\\ &= (\text{Open Int.}, 0, 2, 0), (\text{Structured Int.}, 1, 1, 0),\\ &\quad (\text{T. Observation}, 2, 0, 0), (\text{Card Sorting/Ladd.}, 1, 1, 0)\end{aligned}$$

Table 9
Mapping of the first value of the example contextual situation.

Factor	Attributes	Values	Open-Ended Interv.	Structured Interv.	Task Observation	Card Sorting/Ladd.
Elicitor	Training in Elicitation Techniques	High	√	√	√	√
		Low	–	√	√	–
		Zero	–	–	–	×
	Elicitation Experience	High	√	√	√	√
		Medium	√	√	√	√
		Low	–	–	√	–

Table 10
Partial view of the example contextual situation mapping.

Factor	Attributes	Values	Open-Ended Interv.	Structured Interv.	Task Observation	Card Sorting/Ladd.
Elicitor	Training in Elicitation Techniques	High	√	√	√	√
		Low	–	√	√	–
		Zero	–	–	–	×
	Elicitation Experience	High	√	√	√	√
		Medium	√	√	√	√
		Low	–	–	√	–

6.2.3. Priorization function

Finally, the prioritization function, denoted by the operator φ, orders list P by adequacy. In other words, this function generates a list of techniques ordered by the weights output by the aggregation function:

$$\varphi(P) = P'.$$

For the example, the list would be ordered as follows:

$$P' = (\text{T. Observation}, 2, 0, 0), (\text{Structured Int.}, 1, 1, 0), (\text{Card Sorting/Ladd.}, 1, 1, 0), (\text{Open Int.}, 0, 2, 0).$$

The techniques are divided into two groups depending on their x values. The "Not Recommended" values carry the heaviest weight in the decision-making process on technique adequacy, as their definition is more restrictive and tougher than the other values ($\surd$, –) for any one technique. The adequacy value can be:

- *Adequate:* Techniques that have no "Not Recommended" (×) values for this contextual situation. They are ordered by the number of "Recommended" ($\surd$) values.
- *Less adequate:* Techniques that have "Not Recommended" (×) values. In this case, the techniques are ordered not by the number of Recommended ($\surd$) values, but by the Not Recommended (x) values. Moreover, the order is not correlative with the number of (x) values. We established a priorization strategy.

The simplest option for calculating the adequacy level of each technique using the adequacy table (Table 5) is to weigh up the equally relevant contextual attributes and simply add up the indifferent and inadequate attributes. If there are no adequate techniques, however, a decision needs to be made on what action to take to improve this situation. To do this, it is necessary to decide which attribute should be changed. But not all attributes can be changed. The values assigned to some attributes are unchangeable, whereas others are more easily modifiable by taking some action. For this reason, we have clustered the attributes into four sets according to their *modifiability*, as shown in Fig. 4. Within each set, the attributes are ordered by priority.

These four types of modifiability sets are:

- *Fixed:* Attributes that are hard to modify and should be considered unchangeable for elicitation techniques selection. If this attribute group fails to unearth adequate techniques, there are no actions that can improve the situation. For this situation, there do not appear to be any adequate elicitation techniques.
- *Rigid:* Attributes whose value can be changed to get better elicitation technique selection results. However, the activities for improving this set of attributes generate a project cost, since these attributes are related to the people involved in elicitation. Therefore, an improvement means changing project participants. This is likely to lead to changes in the other stakeholders, and the elicitation techniques selection has to start again from scratch.
- *Scalable:* Attributes closely related to the people involved in elicitation but that do not imply personnel changes. Therefore, an improvement of these attributes does not affect other attributes. The project costs associated with these improvement actions could be high.
- *Flexible:* Ductile attributes which are easy to change to get more adequate elicitation techniques.

So, for example, a technique that failed on two flexible attributes is better than a technique that failed on a scalable attribute. The scale does not include failures on fixed attributes, as their values cannot be modified at discretion. This scale is shown in Table 11.

6.3. Improve contextual situation

The selection procedure delivers a list of techniques that are adequate for the contextual situation. If there are no adequate techniques, the procedure will show the less adequate techniques ordered according to the scale shown in Table 11, and we should consider the recommendations listed in Table 12 on improvements to be made to the contextual situation to get adequate techniques.

7. Example of elicitation technique selection

Finally, we are able to use the proposed framework. Here is an example. Let us suppose that a requirements engineer, who is quite

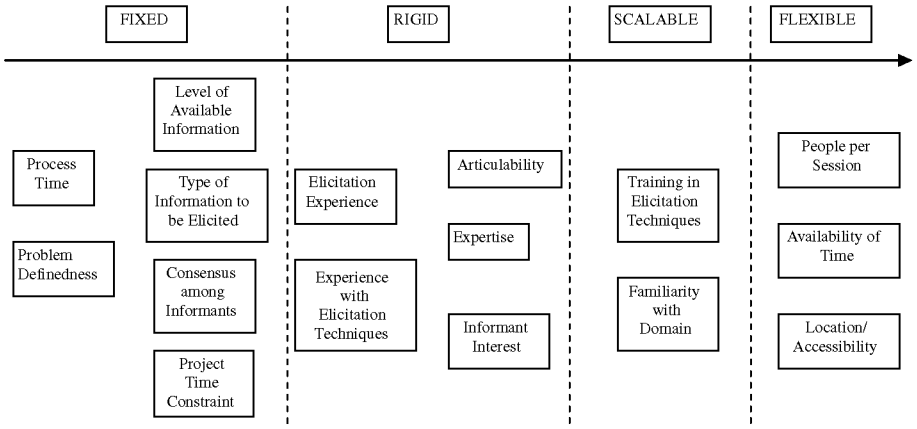

Fig. 4. Attribute modifiability sets and priority.

Table 11
Scale of less adequate techniques.

Scale attrib. No. (x)	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	13th	14th	15th	16th
Flexible (F)	1	2	3		1	2	3		1	2	3			1	2	3	1	>1
Scalable (E)				1	1	1	1	2	2	2	2		1	1	1	1	2	>2
Rigid (R)												1	1	1	1	1	1	>1

Table 12
Action guidelines for attribute improvement.

Attribute groups	Effects of changes	Improvement action	Recommendation output from selection procedure	Attribute	Action guide	Justification
Fixed	Unchangeable	None	There are no adequate techniques. The selection process can continue, but the recommended techniques will not be very appropriate for the case at hand	Process time Problem definedness Level of available information Type of information to be elicited Consensus among informants Project time constraint	None	This attribute value is unchangeable
Rigid	Changes affect other attributes and selection process needs to be restarted	Change of participant	Change the elicitation participants to be able to use adequate techniques. In this case, you should restart the elicitation process. If you decide not to make changes, you can continue with the process, but the application conditions will not be optimum for the recommended techniques	Elicitation experience Experience with elicitation techniciues	Change elicitor	A more experienced elicitor will increase the range of adequate techniques
				Articulability Expertise Interest	Change informant	A more articulate, expert or interested informant will increase the range of adequate techniques
Scalable	Changes affect neither other attributes nor the selection process	Learning activities	You can improve the current attribute conditions and be able to set adequate techniques. Continue with the selection process without improving attributes might cause the recommended techniques not being very adequate	Training in elicitation techniques	Train in elicitation techniques	A better trained elicitor will increase the range of adequate techniques
				Familiarity with domain	Get to know more about the problem domain	An elicitor that is more familiar with the problem domain will increase the range of adequate techniques
Flexible	Changes affect neither other attributes nor the selection process	Logistic action	You may want to expend effort and/or resources on actions aiming to improve the conditions of these attributes to get more adequate techniques	People per session	Reduce number of informants	There are more adequate techniques for individual informants than for group informants
				Availability of time	Facilitate agendas	There are more adequate techniques for people with time
				Location/accessibility	Facilitate meetings	There are more adequate techniques for participants that are close

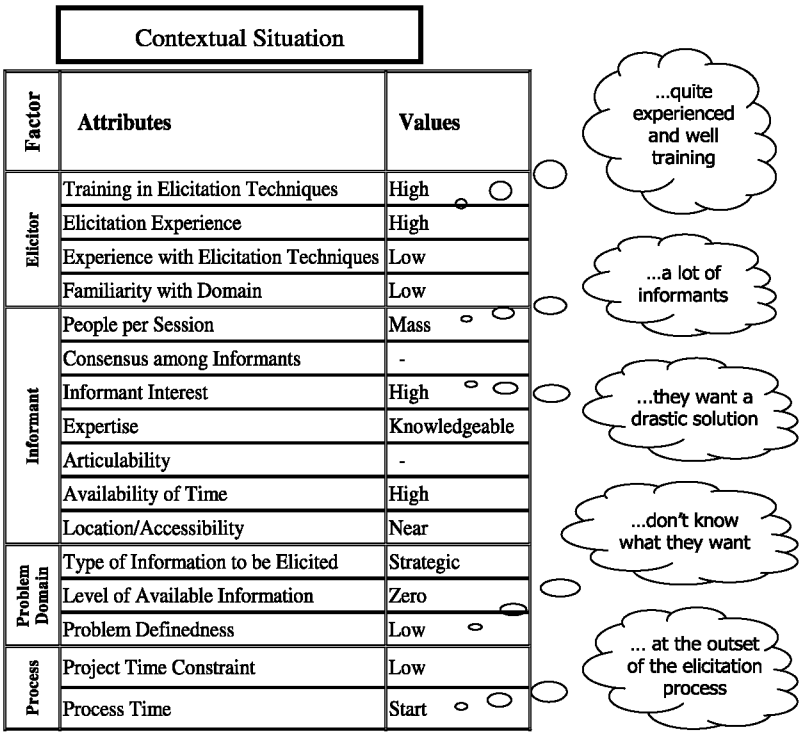

Fig. 5. Initial contextual situation.

experienced and well trained in elicitation techniques, has to specify the needs of a company that has a lot of potential informants from different organizational levels. The informants are not clear about exactly what they want but they know that they are looking for a drastic solution to all their current computer-related problems. This and other available information is used to state the contextual situation shown in Fig. 5 at the outset of the elicitation process.

This contextual situation is then mapped to the adequacy table (Table 5). The rows shaded grey in Table 13 represent this mapping. We get the adequacy of each technique by adding up the recommended, indifferent and not recommended values following the procedure defined above in Section 6.

Techniques are prioritized based on the result of adding up all the adequacy values for each technique, as shown in Table 14. In this case, we get six adequacy techniques, that is, techniques that do not score any *not recommended* value. Of these, we can single out two, Brainstorming and JAD Workshop, as being the most adequate, because they do not have *indifferent* values either.

For the next session, the elicitor wants to continue elicitation from the managers of the key areas. This means that the situation has changed to elicitation from an individual informant. The first manager is based at the company's productive facilities far away from the headquarters in the town center. Additionally, values have been assigned to other attributes after the first group session. For example, we found that stakeholders were divided and the manager in question was not very articulate. The new contextual situation is shown in Fig. 6.

Again we map this situation to the adequacy table in order to output technique adequacy as shown in Table 15.

And the techniques are prioritized by adequacy. In this case, no adequate techniques are output as they all have at least one *Not Recommended* value. As we necessarily have to select a technique for the session, we weigh up the less inadequate techniques. To do this, we consider Table 11. We first take the techniques that have one failure (that is, one *not Recommended* value) for the *flexible* group attributes, as they are the easiest to modify, then techniques that have 2, 3..., followed by techniques with 1 failure for the *scalable* attributes, then 1 for the *flexible* attributes and 1 for *scalable* attributes, and so on.

In our example, we have two techniques with one failure for a flexible attribute, namely Location/Accessibility. The result is returned to the elicitor with a recommendation on what action to take to improve the adequacy of the selected techniques. In this case, the recommendation is to facilitate agendas, for example, the elicitor should travel to the informant's place of work. The results of the selection are shown in Table 16.

8. Validation

In order to empirically validate our proposal, we have run two experiments to evaluate different aspects of the proposed solution. We set out to evaluate the effectiveness and efficiency of our proposal against the selection procedures based on elicitor experience. In particular, we want to check whether our approach is really qualitatively better for selecting techniques, and if whether this selection leads to better results in the requirements elicitation process. As shown in Fig. 7, the first experiment deals with the endemic problems of elicitation technique selection: overselection of the open interview, subjective technique selection and selection of a single technique. The second experiment aims to check whether the proposed framework selects more effective techniques than current requirements engineering methods and whether our framework satisfies users.

8.1. Experiment 1

The *main goal* of this experiment was to examine the influence of the method used to select the elicitation techniques regarding the following points:

Table 13
Mapping of the initial contextual situation to the adequacy table.

Grey rows represent the projection of contextual situation values on adequacy table
Black cells represent values of the attributes for current contextual situation

Factor	Attributes	Values	Open-Ended Interv.	Structured Interv.	Task Observation	Card Sorting/Ladd.	Questionnaires	Protocol Analysis	Repertory Grid	Brainstorming	Nominal Group T.	Delphi Technique	Participant Observ.	Prototyping	Focus Group	JAD Workshop	Scenarios/U. Cases
Elicitor	Training in Elicitation Techniques	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	–	√	√	–	√	–	–	√	√	–	√	–	–	–	–
		Zero	–	–	–	×	–	×	×	–	×	×	–	–	×	×	–
	Elicitation Experience	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Medium	√	√	√	√	√	–	√	–	√	√	√	√	√	–	√
		Low	–	–	√	–	–	×	–	–	√	√	√	–	–	×	–
	Experience Elicitation Techniques	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Zero	–	–	√	–	√	–	–	–	√	–	–	–	–	×	–
	Familiarity with Domain	High	√	√	–	√	√	–	√	√	–	√	–	√	√	√	–
		Low	√	√	–	√	√	√	√	√	–	√	√	–	√	√	–
		Zero	√	–	√	–	–	√	–	√	–	–	√	×	×	×	–
Informant	People per Session	Indiv.	√	√	√	√	√	√	√	×	×	×	√	√	×	×	√
		Group	–	–	√	–	√	–	√	–	√	√	√	–	√	–	–
		Mass	×	×	–	×	√	×	√	√	√	√	√	–	√	√	–
	Consensus Informants	High	√	√	–	√	√	√	√	–	√	–	–	√	√	–	√
		Low	×	×	–	×	√	×	√	–	√	√	–	√	√	–	–
	Informant Interest	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	–	–	√	√	√	–	√	–	√	√	–	√	√	–	√
		Zero	×	–	√	–	√	×	√	×	√	√	×	√	–	×	–
	Expertise	Expert	√	√	–	√	√	√	√	√	√	√	√	√	√	√	√
		Knowl.	√	√	√	√	√	–	√	√	√	√	√	√	√	√	√
		Novice	–	–	√	√	√	–	√	√	√	–	√	√	√	√	√
	Articulabilit y	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Medium	√	√	√	√	√	–	√	√	√	√	√	√	√	√	√
		Low	–	–	√	√	√	×	√	×	×	√	–	√	×	×	–
	Availability of Time	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	–	–	√	–	√	×	√	×	×	√	×	–	×	×	–
	Location/ Accessibility	Far	–	–	–	–	√	×	√	×	×	√	×	×	×	–	×
		Near	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Problem Domain	Type of Information to be Elicited	Strategic	√	–	–	×	√	–	×	√	√	√	√	–	√	√	–
		Tactical	√	√	√	×	√	√	×	√	√	√	√	√	√	√	√
		Basic	–	√	–	√	√	–	√	×	×	–	–	√	–	×	–
	Level of Available Information	Upper	√	√	√	×	√	√	×	–	–	√	√	√	√	√	√
		Lower	√	√	√	√	√	×	√	√	√	√	√	√	×	√	√
		Zero	√	×	√	×	×	×	×	√	√	×	√	×	×	√	√
	Problem Definedness	High	–	√	√	√	√	√	√	√	√	√	√	√	√	×	√
		Low	√	–	√	×	√	×	×	√	√	√	–	√	√	√	–
Process	Project Time Constraint	High	–	–	×	√	√	×	√	×	×	×	×	×	×	×	√
		Medium	√	√	–	√	√	–	√	–	–	–	–	–	–	–	√
		Low	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	Process Time	Start	√	–	√	√	–	×	–	√	√	–	√	–	√	√	–
		Middle	√	√	–	√	√	√	√	–	√	√	√	√	√	×	√
		End	√	√	×	×	√	–	–	×	×	–	×	–	–	×	–

Table 14
Prioritization of techniques for initial contextual situation after adequacy quantification.

Suitability	Technique	Number (√)	Number (−)	Number (x)
Adequate	Brainstorming	14	0	0
	JAD workshop	14	0	0
	Nominal group tech	13	1	0
	Participant observation	13	1	0
	Task observation	11	3	0
	Use cases/scenarios	9	5	0

- *Overselection of open interview*, that is, whether subjects select the open interview more often than other elicitation techniques that are equally applicable to the ongoing project.
- *Single candidate technique*, that is, whether subjects tend to select a single elicitation technique, although more than one is applicable.
- *Subjective selection*, that is, whether subjects select the elicitation technique to be applied objectively depending on the ongoing project features or subjectively depending on their preferences.

This would help to validate the systematicity of our framework, that is, whether subjects achieve similar results to the expert and whether these results are different from the restrictive/subjective selections made using other methods.

The *experimental subjects* were eleven software engineering master students. These students performed the requirements engineer role and selected elicitation techniques. They all had similar development experience (three to four years).

The *factor* or independent variable under study was the *method used to select techniques*. This factor had three levels: subjects' *own method* (the method that subjects used at the start of the course); *learned method* (where subjects take into account the information on elicitation techniques taught as part of the master's degree course: 6-h course unit on elicitation and 6-h course unit on elicitation techniques), and the *proposed framework* (the solution presented here). As the project type could possibly have an influence on this relationship, we considered three *statements of cases* from the *financial*, *technological* and *industrial* domains as a blocking variable.

In the *experimental design*, the eleven subjects were randomly assigned to three groups (with four, four and three members, respectively). First, the group that was to have fewer subjects was chosen by lot. It was group 2. Then these groups applied each method to a different case. This way, we managed to block the *case statement* variable, as shown in Table 17. Finally, a piece of paper was drawn from each of two bags at the same time: one contained the name of the subjects and the other contained the available groups.

The *experimental procedure* was as follows: subjects were given a different case statement at three different times during the course: at the beginning, after general training on techniques, and after training on our framework. At each time, they were asked to select which elicitation techniques to apply using the above three selection methods.

Three *response variables* (or dependent variables) were measured: the *number of selected techniques* and two Boolean (yes/no) variables indicating whether or not the *open interview was one of the chosen techniques* and whether or not the *subjects' and expert's selection matched*. In order to determine the value of

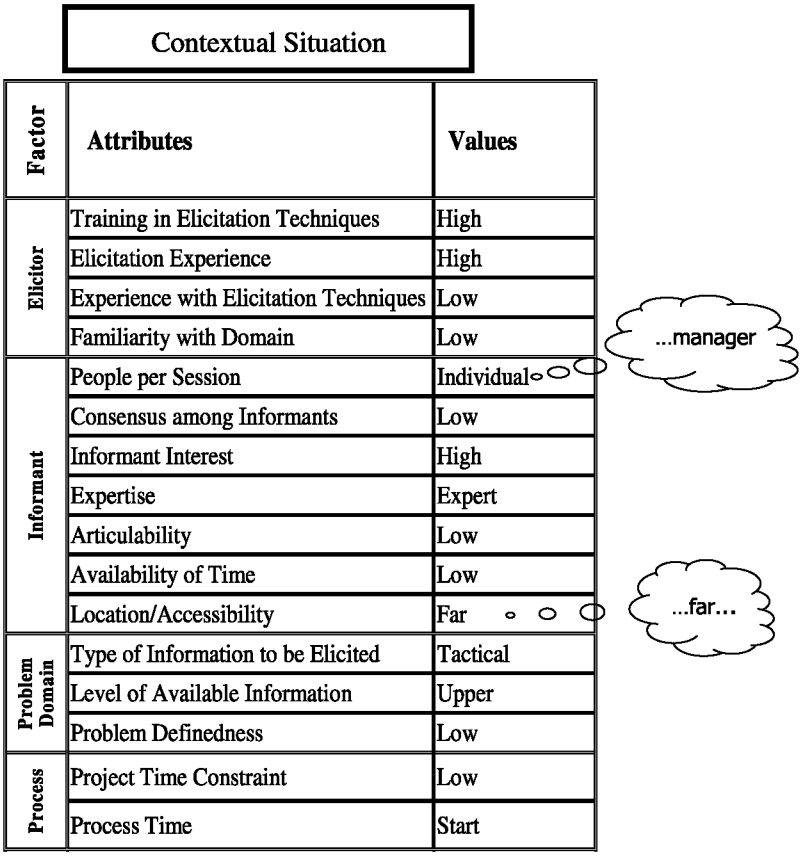

Fig. 6. Contextual situation of the second session.

Table 15
Mapping of the contextual situation of the second session to the adequacy table.

Grey rows represent the projection of contextual situation values on adequacy table
Black cells represent values of the attributes for current contextual situation

Factor	Attributes	Values	Open-Ended Interv.	Structured Interv.	Task Observation	Card Sorting/Ladd.	Questionnaires	Protocol Analysis	Repertory Grid	Brainstorming	Nominal Group T.	Delphi Technique	Participant Observ.	Prototyping	Focus Group	JAD Workshop	Scenarios/U. Cases
Elicitor	Training in Elicitation Techniques	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	–	√	√	–	√	–	–	√	√	–	√	–	–	–	–
		Zero	–	–	–	×	–	×	×	–	×	×	–	–	×	×	–
	Elicitation Experience	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Medium	√	√	√	√	√	–	√	–	√	√	√	√	√	–	√
		Low	–	–	√	–	–	×	–	–	√	√	√	–	–	×	–
	Experience Elicitation Techniques	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Zero	–	–	√	–	√	–	–	–	√	–	–	–	–	×	–
	Familiarity with Domain	High	√	√	–	√	√	–	√	√	–	√	–	√	√	√	–
		Low	√	√	–	√	√	√	√	√	–	√	√	–	√	√	–
		Zero	√	–	√	–	–	√	–	√	–	–	√	×	×	×	–
Informant	People per Session	Indiv.	√	√	√	√	√	√	√	×	×	×	√	√	×	×	√
		Group	–	–	√	–	√	–	√	–	√	√	√	–	√	–	–
		Mass	×	×	–	×	√	×	√	√	√	√	√	–	√	√	–
	Consensus Informants	High	√	√	–	√	√	√	√	–	√	–	–	√	√	–	√
		Low	×	×	–	×	√	×	√	–	√	√	–	√	√	–	–
	Informant Interest	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	–	–	√	√	√	–	√	–	√	√	–	√	√	–	√
		Zero	×	–	√	–	√	×	√	×	√	√	×	√	–	×	–
	Expertise	Expert	√	√	–	√	√	√	√	√	√	√	√	√	√	√	√
		Knowl.	√	√	√	√	√	–	√	√	√	√	√	√	√	√	√
		Novice	–	–	√	√	√	–	√	√	√	–	√	√	√	√	√
	Articulabilit y	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Medium	√	√	√	√	√	–	√	√	√	√	√	√	√	√	√
		Low	–	–	√	√	√	×	√	×	×	√	–	√	×	×	–
	Availability of Time	High	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
		Low	–	–	√	–	√	×	√	×	×	√	×	–	×	×	–
	Location/ Accessibility	Far	–	–	–	–	√	×	√	×	×	√	×	×	×	–	×
		Near	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Problem Domain	Type of Information to be Elicited	Strategic	√	–	–	×	√	–	×	√	√	√	√	–	√	√	–
		Tactical	√	√	√	×	√	√	×	√	√	√	√	√	√	√	√
		Basic	–	√	–	√	√	–	√	×	×	–	–	√	–	×	–
	Level of Available Information	Upper	√	√	√	×	√	√	×	–	–	√	√	√	√	√	√
		Lower	√	√	√	√	√	×	√	√	√	√	√	√	×	√	√
		Zero	√	×	√	×	×	×	×	√	√	×	√	×	×	√	√
	Problem Definedness	High	–	√	√	√	√	√	√	√	√	√	√	√	√	×	√
		Low	√	–	√	×	√	×	×	√	√	√	–	√	√	√	–
Process	Project Time Constraint	High	–	–	×	√	√	×	√	×	×	×	×	×	×	×	√
		Medium	√	√	–	√	√	–	√	–	–	–	–	–	–	–	√
		Low	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	Process Time	Start	√	–	√	√	–	×	–	√	√	–	√	–	√	√	–
		Middle	√	√	–	√	√	√	√	–	√	√	√	√	√	×	√
		End	√	√	×	×	√	–	–	×	×	–	×	–	–	×	–

this variable, we compared the selections by subjects against the selection made by an expert using the proposed framework.

The *data collected* in this first experiment are shown in Table 18. Notice that more than one technique can be selected, that is, subjects applying all three methods may output more than one candidate technique. This is why the values in the *Number of selected (candidate) techniques* column are greater than 1.

Below, we present the data analysis for each response variable. We used the SPSS v.15 statistical tool in all cases.

8.1.1. Overselection of open interview

The open interview is classed as overselected when subjects choose the open interview, which appears to be universally applicable, in preference to other elicitation techniques that are equally effective for the ongoing project. The overselection of open interview impoverishes the elicitation process, as it is an impediment to the use of other potentially more effective elicitation techniques.

In order to study the overselection of open interview, we have manipulated the statement of cases so that the open interview is

Table 16
Prioritization of Techniques for the Contextual Situation in the Second Session.

Suitability	Scale	Technique	Number (√)	Number (–)	Number (x)	Failed attribute	Action guide
Adequate	0(*x*)						
Inadequate	1(*x*)F	Prototyping	12	3	1	Location/accessibility	Facilitate agendas
		Use cases/scenarios	8	7	1	Location/accessibility	Facilitate agendas

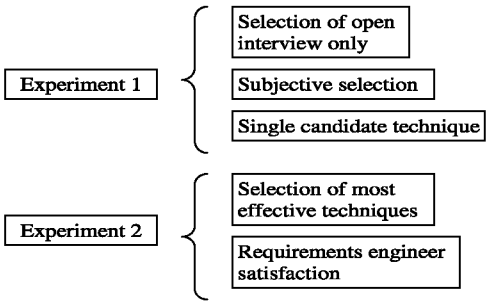

Fig. 7. Empirical validation of the proposed framework.

Table 17
Design of Experiment 1.

	Financial statement	Technological statement	Industrial statement
Own method	Group 1	Group 2	Group 3
Learned method	Group 2	Group 3	Group 1
Proposed framework	Group 3	Group 1	Group 2

not the best technique in any case. Consequently, subjects should not to consider open interview as the best technique, irrespective of the method (*own, learned, framework)* in use.

The relevant variable for determining the overselection of open interview is *Is Open interview chosen?* This variable takes a *Yes/No* value depending on whether the subject chooses *open interview* or *other techniques*. As Table 19 shows, six out of 11 subjects using their own method chose open interview, whereas only one subject using the other two methods (learned and our framework) selected this technique. Consequently, at first glance, it appears that subjects are inclined to choose open interview when they do not have specialized knowledge, acquired either through training (*learned* method) or a selection support strategy (*framework* method).

We can check the above perception statistically. Table 19 is a two-dimensional classification table, also known as a contingency table. Contingency tables are useful for testing the null hypothesis that the classification criterion used in the rows (i.e., whether the open interview was selected) is independent of the classification criterion used in the columns (selection method). To do this, we use Pearson's chi-squared test [40, p. 333].

The results of the chi-squared test for the selection methods ($\chi^2 = 8.250$, d.f. = 2) are shown in Table 20. A p-value of less than .05

Table 18
Collected data in Experiment 1.

Group	Subject	Method	Case statement	Number of selected techniques	Was open interview chosen?	Did subjects' and expert's selection match?
1	3	Own	Financial	3	Yes	Yes
1	11	Own	Financial	1	No	No
1	2	Own	Financial	1	Yes	No
1	5	Own	Financial	1	Yes	No
2	1	Own	Technological	1	No	No
2	10	Own	Technological	1	No	No
2	8	Own	Technological	1	Yes	No
3	6	Own	Industrial	1	No	No
3	4	Own	Industrial	2	No	Yes
3	9	Own	Industrial	1	Yes	Yes
3	7	Own	Industrial	1	Yes	No
2	1	Learned	Financial	1	No	No
2	10	Learned	Financial	1	No	Yes
2	8	Learned	Financial	2	No	Yes
3	6	Learned	Technological	1	No	No
3	4	Learned	Technological	1	Yes	No
3	9	Learned	Technological	1	No	No
3	7	Learned	Technological	2	No	Yes
1	3	Learned	Industrial	1	No	No
1	11	Learned	Industrial	1	No	No
1	2	Learned	Industrial	1	No	No
1	5	Learned	Industrial	1	No	No
3	6	Framework	Financial	6	No	No
3	4	Framework	Financial	4	No	No
3	9	Framework	Financial	4	No	Yes
3	7	Framework	Financial	5	No	Yes
1	3	Framework	Technological	3	No	No
1	11	Framework	Technological	3	No	Yes
1	2	Framework	Technological	3	No	Yes
1	5	Framework	Technological	6	No	Yes
2	1	Framework	Industrial	2	Yes	Yes
2	10	Framework	Industrial	5	No	Yes
2	8	Framework	Industrial	4	No	Yes

Table 19
Subjects' choice of open interview depending on the selection method and case.

Is open interview chosen?	Selection methods			Case statement		
	Own	Learned	Framework	Financial	Technological	Industrial
Yes	6	1	1	3	2	3
No	5	10	10	8	9	8

Table 20
Chi-squared analysis of the relationships between experimental conditions and the overselection of open interviews.

Experimental conditions (independent variables)	Chi-squared value	Degrees of freedom (d.f.)	p-Value
Selection methods	8.250	2	.016
Case statements	.330	2	.848

indicates that the probability of observing such large differences between values in the table cells (e.g., 10 Noes vs. 1 Yes for the Learned or Framework methods) is very small (in fact, less than 5%), i.e., such differences are not a matter of chance. The result is significant (p-value = 0.016 < 0.05), which means that the subjects that use more informed or structured methods choose techniques other than the open interview, whereas subjects that resort to the method with which they are familiar tend to choose this popular technique.

We can run a similar analysis to the above for case statements. Table 19 shows the data for analysis. In this case, we have found, as shown in Table 20, that there is no significant difference between the selection or otherwise of open interview for the case statements used in the experiment (χ^2 = .330, d.f. = 2, p-value = 0.848 > 0.05). This is a positive condition for the experiment, as cases can be omitted from the interrelationship, and we can focus on the relationship of interest, that is, the relationship between the selection method and selected techniques. This result does not mean that there is no dependency for any project type; the relationship certainly might exist.

8.1.2. Single candidate technique

The second point studied was whether subjects tend to select only one of a number of potentially applicable elicitation techniques, thereby limiting the range of possible elicitation techniques. To deal with the problem of underselection, we use the quantitative variable *number of selected techniques*. We want to find out whether the number of selected techniques depends on the selection method. The mean number of selected techniques, classified by *selection method and case statement*, is shown in Table 21. From the data shown in Table 21, it is clear that our method (*framework*) has a greater mean than the other two (4.09 as opposed to 1.18 and 1.27). In other words, subjects tend not to select many techniques when acting upon their own opinion (*own* method) or declarative learning in the classroom (*learned* method) to make the selection.

The most popular way of studying the relationship between a factor (that is, *selection method*) and a numerical dependent variable (*number of selected techniques*) is analysis of variance (ANOVA). ANOVA can only be applied if two conditions are met: sample normality and homogeneity of variances (homoscedasticity). Because there are not many data points (just 11 subjects per method), the sample was rather predictably not normal. This was confirmed using the usual, Kolmogorov–Smirnov and Shapiro–Wilk, normality tests. Consequently, ANOVA was not applicable.

When ANOVA is not applicable due to non-normality, we have the alternative of using a non-parametric analysis. In this case, the Kruskal–Wallis test, which is a non-parametric equivalent to the one-way ANOVA [41, p. 144], is the best option. This test is based on calculating a statistic called H, which is capable of testing the null hypothesis that the factor levels have the same effect on the response variable. The distribution of H can be approximated by the mean of the chi-squared distribution provided that each sample contains at least five observations (as in the case of this experiment).

Table 22 shows the result of the Kruskal–Wallis test. The parameter H is 23.956 in this case. The corresponding p-value can be approximated by means of a chi-square of the same value and two degrees of freedom (χ^2 = 23.956, d.f. = 2). The result is significant (p-value < .05). This means that at least one of the methods of selection (*own*, *learned*, *framework*) leads to an underselection of elicitation techniques.

We can identify which selection method leads users to select more (or less) techniques by running pairwise comparisons of methods (own-learned, own-framework, learned-framework). Since, as mentioned above, the data are not normal, we will have to use a non-parametric test. We have chosen the Mann–Whitney U test for two independent samples (although other tests, like the Wilcoxon W, would be equally applicable). We have applied a Bonferroni correction (which adjusts the significance criterion to 0.05/3 = 0.017) to prevent the problem of type I error inflation, a possibility when the same test is applied over again (in our case, three different times).

The results of applying the Mann–Whitney U test are shown in Table 23. There are no significant differences between the number of selected elicitation techniques using the own and learned methods (U = 59.5, p-value = 0.922 > 0.017), but there is a significant difference between the framework and own methods (U = 3.0, p-value = .0 < 0.017) and learned methods (U = 1.0, p-value = .0 < 0.017), respectively. This analysis again confirms that the framework method returns significantly more candidate techniques than the other two methods.

As in the case of the overselection of open interview, we have studied the effect of the case statement on the underselection of elicitation techniques. The related information is shown in Tables 21 and 22. In this case, we have found that there is no

Table 21
Summary of results on number of techniques variable.

	Selection methods			Case statement		
	Own	Learned	Framework	Financial	Technological	Industrial
Mean number of selected techniques	1.27	1.18	4.09	2.64	2.09	1.82

Table 22
Kruskal–Wallis analysis of the relationships between experimental conditions and the underselection of elicitation techniques.

Experimental conditions (independent variables)	Kruskal–Wallis H value	Chi-squared approximation	
		Degrees of freedom d.f.	p-Value
Selection methods	23.956	2	.000
Case statements	1.229	2	.541

Table 23
Mann–Whitney U analysis of the differences among selection methods.

Methods compared	Mann–Whitney U value	p-Value
Own-learned	59.5	.922
Own-framework	3.0	.000
Learned-framework	1.0	.000

significant difference between the number of selected techniques for the case statements used in the experiment ($H = 1.229$, $\chi^2 = 1.229$, d.f. = 2, p-value = 0.541 > 0.05).

8.1.3. Subjective selection

Finally, Experiment 1 examines whether subjects tend to select the best elicitation technique subjectively according to their personal preferences rather than based on objective project characteristics (i.e., what we refer to throughout the article as *contextual situation*). Subjective selection poses a threat to the elicitation process in that a suboptimal technique could be selected.

To study the problem of subjective selection, we use the qualitative variable *Did subjects' and expert's selection match?* This variable takes YES/NO values, depending on whether the subjects and experts choose exactly the same techniques. As Table 24 shows, eight out of 11 of the subjects using our method (framework) selected the same technique as the expert using the framework, whereas only three selections made by subjects using their own or learned methods matched.

Table 24 is similar to Table 19, and can be analyzed statistically in exactly the same way (using a chi-squared test). The results of the chi-squared test for the selection methods ($\chi^2 = 6.203$, d.f. = 2) are shown in Table 25. The result is significant (p-value = 0.045 < 0.05). This means that the own and learned methods are similar, but differ significantly from the framework method. In other words, the subjects using the framework mostly agree with an expert using the method, which suggests that our proposed approach is more systematic and therefore produces similar outputs independently of the person using the method.

Again, we ran a similar analysis to the above for case statements. As shown in Table 25, the chi-squared test found that there is no significant difference between *Did subjects' and expert's selection match?* and the case statements (p-value = 0.883 > 0.05) for the case statements used in the experiment. Again, we can confirm that it is the selection methods and not the case statements that cause the subjective selection effects.

8.1.4. Discussion of the results of Experiment 1

As Fig. 8 shows, this experiment output the following findings:

Table 25
Chi-squared test for fit between techniques selected by subject and expert.

Experimental conditions (independent variables)	Chi-squared value	d.f.	p-Value
Selection methods	6.203	2	.045
Case statements	.248	2	.883

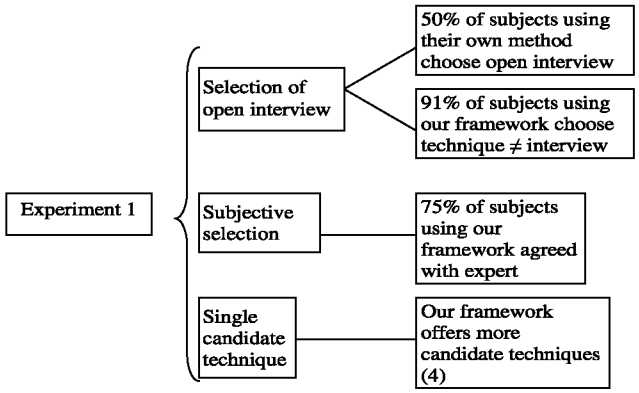

Fig. 8. Results of Experiment 1.

- Half of the subjects using their own knowledge for selection select the open interview. This result confirms that, in the absence of systematic methods, requirements engineers (especially novices) tend to opt for the most popular technique (open interview), irrespective of the project and session in which it is to be used.
- The solution we propose selected techniques other than open interview in 91% of the cases. This result suggests that novice requirements engineers using more systematic selection methods are able to make more varied decisions than their peers using selection based on either their own knowledge or what they learned on a training course about elicitation techniques. In other words, the method proposed here manages to encapsulate more knowledge about elicitation technique adequacy providing a wider range of options.
- The proposed solution makes significantly more elicitation techniques (4) available to requirements engineers than the usual methods. Again, this result suggests that the proposed method encapsulates knowledge about elicitation technique adequacy providing a wider range of options.
- Most of the subjects that used the proposed framework agreed (75%) with the expert using the same method, that is, a systematic method that provides the knowledge required to make

Table 24
Summary of results on fit between subjects' and expert's selection.

Did subjects' and expert's selection match?	Selection methods			Case statement		
	Own	Learned	Framework	Financial	Technological	Industrial
Yes	3	3	8	5	4	5
No	8	8	3	6	7	6

good choices on elicitation techniques tends to level out decisions between novices and experts. Remember that the use of systematic methods that manage to guarantee the output and are less sensitive (optimally, insensitive) to the user is a characteristic of engineering disciplines.

8.2. Experiment 2

The *main goal* of this experiment was to examine the influence of the technique selection method used on the efficiency of the elicitation process. In this case, interest focused on studying how much information was gathered in each session, represented by concepts, relationships and requirements. Favorable results would suggest that our approach helps to improve the elicitation process. Finally, we asked subjects to give their opinion on the proposed framework.

The *experimental subjects* of this second experiment were the same SE master students as in Experiment 1. In this case, not only did they have to select techniques, but they also had to apply the techniques in two elicitation sessions.

The *factor* under study was the method used to select techniques: their own (like the learned method in Experiment 1), and the proposed framework (the solution presented here).

In the *experimental design* the eleven subjects were randomly assigned to two groups representing the selection method factor levels. A paper was drawn from each of two bags at the same time: one contained the subject names and the other held the available groups (the paper with groups was replaced after each extraction).

The *experiment procedure* was as follows: subjects dealt with only one problem domain (battery recycling process). The purpose of this battery recycling system is to control the operation of the machines that perform the recycling process. The system starts up, monitors and shuts down the machines, controlling possible errors. Additionally, the system takes stock of the quantity of batteries received, quantity of recycled product and destination of the waste products.

They had to elicit information from a real informant (we used two randomly assigned informants with similar domain knowledge) in two 30-min sessions (Session 1 and Session 2) in order to specify a system to automatically control the recycling process. Subjects were given a day to analyze the information elicited during each session and prepare a report before the next session. We collected the data from these reports.

The *response variables* considered in this experiment were the *number of concepts*, *number of relationships* and *number of requirements* in each session. The data collected from this second experiment are summarized in Table 26.

Below, we present the data analysis, separately for Sessions 1 and 2. We used the SPSS v.15 statistical tool in all cases.

8.2.1. Session 1

Table 27 summarizes the results of Session 1. Note that the subjects that used our framework to select which elicitation technique to apply in Session 1 captured, on average, more concepts (10.17 vs. 7.80) and relationships (7.33 vs. 6) but fewer requirements than those using their own method (6.83 vs. 7.2). In order to determine whether these differences were statistically significant, we used tests for the comparison of means of the number of concepts, relationships, and requirements identified using the subjects' *own* and *framework* methods.

Tests for the comparison of means like Student's t and Mann–Whitney U are applicable when the response variable is qualitative and the factor for comparison (selection method) has only two levels (in this case, *own* and *framework*). We have mentioned two similar tests (Student's t and Mann–Whitney U), because samples are unlikely to be normal when the experiments have few subjects (as in our case). Under conditions of non-normality, parametric tests

Table 26
Experiment 2 collected data.

Subject	Selection method (group)	Informant	Session 1			Session 2			Total		
			Concepts	Relationships	Requirements	Concepts	Relationships	Requirements	Concepts	Relationships	Requirements
10	Own	1	18	14	5	3	6	10	21	20	15
2	Own	2	10	7	3	2	3	14	12	10	17
6	Own	1	8	1	7	2	8	12	10	9	19
3	Own	2	6	5	11	5	4	7	11	9	18
9	Own	1	12	10	11	2	3	6	14	13	17
7	Own	2	7	7	4	4	3	10	11	10	14
5	Framework	1	7	4	4	2	3	5	9	7	9
1	Framework	2	10	8	7	2	1	7	12	9	14
11	Framework	1	5	4	10	1	2	6	6	6	16
8	Framework	2	10	7	6	2	2	8	12	9	14
4	Framework	1	7	7	9	1	3	5	8	10	14

Table 27
Comparison of the mean number of concepts, relationships and requirements acquired by subjects during session 1 using the elicitation techniques recommended by the *own* and *framework* methods.

Response variable	Method	Means	Normality test (Kolmogorov–Smirnov)			Homogeneity of variances test (Levene)		Equality of means test (Students's t for equal variances)		
			Value	Degrees of freedom (d.f.)	p-Value	Value	p-Value	t Value	d.f.	p-Value
Number of concepts	Own	7.80	.245	5	.200	1.39	.269	1.09	9	.304
	Framework	10.17	.189	6	.200					
Number of relationships	Own	6.00	.304	5	.149	1.39	.269	.626	9	.547
	Framework	7.33	.197	6	.200					
Number of requirements	Own	7.20	.175	5	.200	1.30	.283	−.199	9	.847
	Framework	6.83	.217	6	.200					

Table 28
Comparison of the meaHomogeneity of variances test (Levene)n number of concepts, relationships and requirements acquired by subjects during session 2 using the elicitation techniques recommended by the *own* and *framework* methods.

Response variable	Method	Means	Normality test (Kolmogorov–Smirnov)			Homogeneity of variances test (Levene)		Equality of means test (Students's t for equal variances)			Equality of means test (Mann–Whitley U)	
			Value	d.f.	p-Value	Value	p-Value	t Value	d.f.	p-Value	U value	p-Value
Number of concepts	Own	1.60	.367	5	.026	3.2	.105	–	–	–	4.5	.036
	Framework	3.00	.285	6	.138							
Number of relationships	Own	2.20	.231	5	.200	4.63	.060	2.31	9	.046	–	–
	Framework	4.50	.265	6	.200							
Number of requirements	Own	6.20	.221	5	.200	2.1	.184	2.50	9	.034	–	–
	Framework	9.83	.189	6	.200							

like Student's t are not applicable and should be replaced by non-parametric tests like the Mann–Whitney U.

We have used popular statistical tests (Kolmogorov–Smirnov and Shapiro–Wilk) to check sample normality. Both tests test the null hypothesis that the samples are normal. Table 27 illustrates the results of the Kolmogorov–Smirnov test. Note that the test was not significant in either case (the same applies for the Shapiro–Wilk test). Therefore, the samples can be considered normal, and the Student's t test is applicable.

It is good practice to determine whether the compared samples have homogeneous or heterogeneous variances, because there is more than one version of Student's t (called Student's t for equal or unequal variances [40, p. 369], respectively). The Levene test is usually used to determine the homogeneity or heterogeneity of variances. This test tests the null hypothesis that two groups (in this case, the groups applying the *own* and *framework* methods) have homogeneous variances. Table 27 illustrates the results of applying this test. Note that the Levene test was not significant in either case, and the Student's t test for equal variances is the best option.

Table 27 illustrates the results of applying Student's t test. The results for the three variables (*number of concepts*, *relationships* and *requirements*) were not significant. Consequently, the *framework* method cannot be said to be more effective than the *own* method during Session 1.

8.2.2. Session 2

Table 28 summarizes the results of Session 2. The subjects that used our framework to select an elicitation technique to apply captured more concepts (3.00 vs. 1.60), relationships (4.50 vs. 2.20) and requirements (9.83 vs. 6.2) than those using their own method. We use the same procedure as in Section 8.2.1 to determine whether or not the differences between these measures are statistically significant.

As shown in Table 28, the Kolmogorov–Smirnov normality test turned out to be significant (value = .367, d.f. = 5, p-value = .026) for the *number of concepts* response variable for one of the samples (*own* method). This means that this sample is not normal, and a non-parametric alternative (the abovementioned Mann–Whitney U) will have to be applied in place of the Student's t test. For the other two variables (*number of relationships* and *number of requirements*), the Kolmogorov–Smirnov normality test results were not significant. The samples are, therefore, normal and can be analyzed using theStudent's t test.

As regards homogeneity or heterogeneity of variances, the Levene test was not significant in any case, as shown in Table 28. This is of no consequence whatsoever for the *number of concepts* response variable, as non-parametric tests are not affected by heterogeneity of variances. For the other response variables (*number of relationships* and *number of requirements*), however, the homogeneity of variances indicates that Student's t for equal variances is the correct test to use.

We then examined whether or not the differences between means were significant. The results of the Mann–Whitney test are significant for the *number of concepts* ($U = 4.5$, p-value = $0.036 < 0.05$). This means that subjects that use the proposed framework are more effective at acquiring concepts than subjects using their own technique selection method during Session 2.

The results were also significant for the *number of relationships* ($t = 2.31$, d.f. = 9, p-value = $0.048 < 0.05$) and *number of requirements* ($t = 2.50$, d.f. = 9, p-value = $0.034 < 0.05$) variables, analyzed in this case by means of the Student's t test. This means that subjects that use the proposed framework are also more effective at acquiring relationships and requirements than the subjects using their own technique selection method during the second elicitation session.

8.2.3. Discussion of the results of Experiment 2

As shown in Fig. 9, the main result of this experiment was that the proposed framework made it significantly easier to capture more concepts, relationships and requirements from the domain in Session 2. Although it is striking that there were no significant differences in Session 1, it actually makes sense: all the subjects, irrespective of whether or not they used the framework, chose the open interview for Session 1 (which, incidentally, appears to be an adequate technique for an early elicitation session).

Finally, the survey administered to subjects after the experiments indicated that most respondents found the proposed solution easy to use (100%) and apply (70%), useful (60%) and correct (60%).

8.3. Threats to validity

We have tried to control possible threats to the conclusion, internal, construct and external validity [41] of the experiments.

Statistical conclusion validity, also called inferential validity, determines how confident we can be that the results of the statistical tests are correct. Regarding conclusion validity, neither experiment had many subjects (11). On this ground, appropriate non-parametric methods were applied in some cases where the normality and homoscedasticity conditions were hard to meet. Internal replications of the first experiment were run to help to

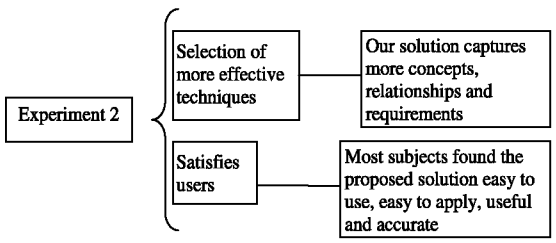

Fig. 9. Results of Experiment 2.

achieve statistical significance. We considered a significance value of 0.05 to avoid false positive conclusions (type I error).

Internal validity determines whether the relationship between the independent and dependent variables is causal (i.e., the proposed framework improves elicitation effectiveness) or due to moderator variables (e.g., learning, fatigue, etc.). We took the following precautions to guard against internal validity threats:

- Subjects did not apply the same method twice to prevent learning from mistakes.
- Methods were applied in a set order—from the least formal to the most formal method (i.e., the subject's own method, followed by the learned method and, finally, the proposed framework)—to assure that what subjects learned from one method did not benefit the next and prevent unconscious formalization.
- The only procedure (our proposed procedure) was applied in situ and supervised by a researcher using the same necessary elements to prevent experiment execution bias.
- Each subject applied different methods to different contextual situations to prevent experimental object learning.
- Sessions were short (30 min) to prevent subject boredom and fatigue.
- The same researcher reviewed the output reports and collected the experimental data to prevent instrumentation bias.
- Two informants with similar domain knowledge were used to prevent fatigue.
- Subject personality, skill and experience were offset by random assignment and using a control group.

Additionally, the subjects applied the methods in order, starting with their own method and ending with our framework, at set times during the course and were unaware that they were participating in an experiment. This should prevent multiple treatment interference. Also, in the second experiment, we used the simple blinding technique for informants and they were not informed of the selection method that the subjects had used for each elicitation session.

Construct validity determines how confident we can be that the implementation of the treatments (e.g., the use to which the subjects put the proposed framework) accurately represents the abstract construct (i.e., the theoretical framework). As regards construct validity, we used different response variables to confirm the cause-effect relationship of the studied variables. Thus, two response variables (Selected techniques (OI/OT) and number of selected techniques) confirmed the selection effectiveness construct in the first experiment, and number of concepts and number of relationships were used to verify the consistency of the selection procedure efficiency construct represented by the number of requirements variable in the second experiment.

The elicitation effectiveness construct is hard to operationalize in an experiment. In this context, operationalization means the process of identifying one or more measurable variables that truly represent the underlying construct (that is, what we really want to measure). We opted to operationalize elicitation effectiveness as the number of elicited requirements, as this is an easy element to measure and also signifies the ultimate aim of requirements elicitation. There are certainly other aspects related to elicitation effectiveness, such as requirements quality or client/user satisfaction. However, requirements quality and client satisfaction are much more subjective measures than quantifying the number of requirements. In other words, although other alternative measures of the effectiveness construct are conceivable, they appear to be less objective and, therefore, more controversial than the selected *number of requirements* measure.

External validity refers to the generalizability of the results to populations, different contexts and different time points. Regarding external validity, we used subjects of similar experience and ability to potential users of the proposed framework. The subjects were master's students, many of whom have professional experience and were working at the time they took the master's course. The experimental results can be assumed to be valid for novice software developers with little experience in requirements elicitation and reasonably accurate for slightly higher levels of experience. We would have to conduct further studies with more diversified populations to be able to claim that the results are applicable to all requirements engineers, especially engineers with a lot of requirements elicitation experience.

Finally, the problem types used also place a constraint on the experiments. Toy problems generally facilitate laboratory experimentation, but they pose a threat to the extrapolation of the findings to real environments. Besides, the relative simplicity of toy problems (compared with real-world problems) is likely to bias technique effectiveness.

9. Conclusion

Requirements engineers come across two major stumbling blocks when they have to choose the most adequate elicitation techniques for use: unfamiliarity with a wide range of elicitation techniques and when they are best used. This might be because the theoretical groundwork for forming a body of mature knowledge that identifies adequate application contexts for elicitation techniques is missing. Pragmatic information on techniques is heterogeneous, has not been compiled and is scattered across multiples sources.

This paper proposes a framework to help requirements engineers to select the most adequate elicitation techniques at any time. To do this, we have determined which attributes are relevant to the context of the elicitation process and influence the selection of one or other technique. Then, we have established the adequacy values of each technique for each contextual attribute value. Finally, we have created a procedure for selecting the most adequate elicitation techniques based on a particular contextual situation.

To build the framework we have explored both the existing underlying theory and the results of empirical research. Based on this, we have deduced and put together justified proposals about the framework components. Additionally, we have had to add information not found in theoretical or empirical sources. In these cases, we drew on our own experience and expertise.

The framework is at an early stage: it contains relevant attributes and adequacy values for techniques established on the basis of information available at the time of research. The elicitation techniques considered were selected subjectively, although we tried to assure that they matched up with the commonly used or best known techniques. The framework is easily extensible and changeable. Whenever any theoretical or empirical evidence is unearthed for an attribute, technique or adequacy value, the information can be easily added to framework.

We validated the proposed framework by running two experiments on software engineering master students. Both experiments have shown that our proposal helps to choose techniques other than open interview, provides access to more possible techniques and captures more requirements information.

We are aware that these results are not generalizable and we have to replicate the experiments with other larger samples. Although the results are favorable to our proposed framework, it is difficult to establish how much of this is due to the selected technique and how much is due to the actual selection procedure. Also on the future research agenda is the study of the construct that really represents technique adequacy. We are not sure that the number of requirements or session productivity really expresses

elicitation technique efficiency. Another aspect of this research requiring improvement in future research is the adequacy table. The shortage of evidence (both empirical and expert opinion) led us to have to identify the most likely values of the elicitation techniques in different contexts. This may bias the recommendations. Fortunately, the proposed method is designed to be easily corrected and expanded as new evidence is reported in the literature.

Finally, we have built a tool to support the use of the framework by interested users in practice (http://www.grise.upm.es/demos/faset/).

Acknowledgments

This work has been funded by the Spanish Ministry of Science and Innovation under Research Grant TIN2008-00555 and University of Atacama under Research Grant DIUDA 221201.

Appendix A.

Technique	Description	Refs.
Open-ended interview	In-person interaction between informant and elicitor where goals and contents are neither structured or completely defined	[42,43]
Structured interview	In-person interaction between informant and elicitor where session goals and contents have been prepared	[44,37]
Task observation	Observation by the elicitor of people at work	[45,46]
Concept ranking/ laddering	Elicitor gives informant or informants a set of domain concepts written on cards and asks them to do different things with them	[45,47,48]
Surveys/ questionnaires	Set of questions on paper or another medium presented or sent to one or more informants	[49,50]
Protocol analysis	Informants relate aloud what they do when they perform specific tasks	[41,51,52]
Repertory grid	Informants have to evaluate a set of domain elements based on constructs (element characteristics)	[53,54]
Brainstorming	Informants, individually or as a group, come up with ideas, deliberatively and in no particular order, which are not immediately assessed	[55,56]
Nominal group technique	Informants come up with ideas that are formally voted on for prioritization	[57,58]
Focus group	Semi-structured group interviews encouraging open discussion in the presence of the elicitor	[59,60]
Delphi method	Feedback of responses to a questionnaire with the aim of outputting an outcome representative of group opinion	[61,62]
Participant observation	The elicitor actively participates in some tasks to find out what skills and knowledge are required to effectively perform tasks or get acquainted with a domain	[63,64]
Prototyping	Develop a simplified version of the system under construction to help capture requirements	[65,66]
JAD workshop	System stakeholders work together at facilitated meetings	[67,68]
Use cases/ scenarios	Informants describe a set of possible action scenarios and events describing part of system behavior	[69,70]

References

[1] A. Abran, J.W. Moore, SWEBOK: Guide to the Software Engineering Body of Knowledge, 2004 ed. <http://www.computer.org/portal/web/swebok> (2004).
[2] D. Zowghi, C. Coulin, Requirements elicitation: a survey of techniques, approaches, and tools, in: A. Aurum, C. Wohlin (Eds.), Engineering and Managing Software Requirements, Springer-Verlag, New York, 2005, pp. 19–46 (book chapter).
[3] H.F. Hofmann, F. Lehner, Requirements engineering as a success factor in software projects, IEEE Softw. 18 (4) (2001) 58–66.
[4] P. Chatzoglou, L. Macaulay, Requirements capture and IS methodologies, Inf. Syst. J. 6 (3) (July 1996) 209–225.
[5] J. Goguen, C. Linde, Techniques for requirements elicitation, in: International Symposium on Requirements Engineering, Los Alamitos, IEEE Computer Society Press, California, January 1993, pp. 152–164.
[6] B. Nuseibeh, S.M. Easterbrook, Requirements engineering: a roadmap, the future of software engineering, in: A.C.W. Finkelstein (Ed.), Companion Volume to the Proceedings of the 22nd International Conference on Software Engineering, ICSE'00, IEEE Computer Society Press, 2000.
[7] N. Cooke, Varieties of knowledge elicitation techniques, Int. J. Hum. Comp. Stud. 41 (6) (Dec. 1994) 801–849.
[8] A. Davis, P. Hsia, Giving voice to requirements engineering: guest editors' introduction, IEEE Softw. 11 (2) (March 1994) 12–16.
[9] A. Davis, A. Hickey, The Role of Requirements Elicitation Techniques in Achieving Software Quality, Requirements Engineering: Foundations for Software Quality Workshop, Essen, Germany, 2002.
[10] G. Playle, C. Schroeder, Software requirements elicitation: problems, tools, and techniques, Crosstalk: J. Defense Softw. Eng. 9 (12) (1996) 19–24.
[11] H. Saiedian, R. Dale, Requirements engineering: making the connection between the software developer and customer, Inf. Softw. Technol. 42 (6) (2000) 419–428.
[12] A. Davis, O. Dieste, A. Hickey, N. Juristo, A.M Moreno, Effectiveness of requirements elicitation techniques: empirical results derived from a systematic review, in: Proc. of 14th IEEE International Conference on Requirements Engineering, 11–15 September 2006.
[13] J. Kim, J. Courtney, A survey of knowledge acquisition techniques and their relevance to managerial problem domains, Decis. Support Syst. 4 (1988) 269–284.
[14] J.S. Dhaliwal, I. Benbazat, A framework for the comparative evaluation of knowledge acquisition tools and techniques, Know. Acquisit. 2 (2) (June 1990) 145–166.
[15] T.A. Byrd, K.L. Cossick, R.W. Zmud, A synthesis of research on requirements analysis and knowledge acquisition techniques, MIS Quart. 16 (1992) 117–138.
[16] N. Maiden, G. Rugg, ACRE: selecting methods for requirements acquisition, Softw. Eng. J. 11 (3) (May 1996) 183–192.
[17] S. Robertson, J. Robertson, Mastering the Requirements Process, Addison-Wesley, 1999.
[18] A. Davis, A. Hickey, A tale of two ontologies: the basis for systems analysis technique selection, in: Proc. 9th Annual American Conference on Information, System, 2003.
[19] S. Lauesen, Software Requirements: Styles and Techniques, Addison-Wesley, 2002.
[20] Li. Jiang, A. Eberlein, B.H. Far, A case study validation of a knowledge-based approach for the selection of requirements engineering techniques, Requir. Eng. 13 (2) (2008) 117–146.

[21] E. Kheirkhah, A. Deraman, Important factors in selecting requirements engineering techniques, in: International Symposium on Information Technology, 2008.
[22] A. Davis, A. Hickey, Requirements elicitation and requirements elicitation technique selection: a model of two knowledge-intensive software development processes, in: Proc. of the 36th Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, California, 2003.
[23] N. Juristo, A.M. Moreno, Basics of Software Engineering Experimentation, second printing, Kluwer Academic Publishers, Boston, 2003.
[24] W.J. Lloyd, Tools and Methods for Effective Distributed Requirements Engineering: An Empirical Study, Master Thesis Dissertation, Virginia Tech, 2001. <http://scholar.lib.vt.edu/theses/available/etd-07262001-110924/> (2001).
[25] R. Agarwal, M. Tanniru, Knowledge acquisition using structured interviewing: an empirical investigation, J. Manage. Inf. Syst. 7 (1) (1990) 123–140.
[26] M. Roth, C. Wood, Knowledge acquisition from single versus multiple experts: a field study comparison using the Delphi technique, J. Know. Eng. 6 (3) (1993). Fall.
[27] A. Burton, N. Shadbolt, G. Rugg, A. Hedgecock, The efficacy of knowledge elicitation techniques: a comparison across domains and levels of expertise, Know. Acquisit. 2 (2) (June 1990) 167–178.
[28] C.J. Chao, G. Salvendy, Impact of cognitive abilities of experts on the effectiveness of elicited knowledge, Behav. Inf. Technol. 14 (3) (1995) 174–182.
[29] G.J. Browne, M.B. Rogich, An empirical investigation of user requirements elicitation: comparing the effectiveness of prompting techniques, J. Manage. Inf. Syst. 17 (4) (2001) 223–249.
[30] B.P. McCloskey, J. Geiwitz, J. Kornell, Empirical comparisons of knowledge acquisition techniques, in: Proc. of the Human Factors Society 35th Annual Meeting Human Factors Soc., vol. 1, Santa Monica, CA, USA, 1991, pp. 268–272.
[31] M. Grabowski, Knowledge acquisition methodologies: survey and empirical assessment, in: Proc. of the Ninth International Conference on Information Systems, 1988, pp. 47–54.
[32] N. Shadbolt, A. Burton, Empirical Studies in Knowledge Elicitation, ACM-SIGART Special Issue on Knowledge Acquisition, 1989.
[33] B. Fazlollahi, M. Tanniru, Selecting a requirements determination methodology-contingency approach revisited, Inf. Manage. 21 (5) (1991) 291–303.
[34] W. Wagner, Q. Chung, M. Najdawi, The impact of problem domains and knowledge acquisitions techniques: a content analysis of P/OM expert system case studies, Expert Syst. Appl. 24 (2003) 79–86.
[35] C. Holsapple, V. Raj, An exploratory study of two KA methods, Exp. Syst. 11 (2) (1994) 77–87.
[36] M. Keil, E. Carmel, Customer–developer links, Commun. ACM 38 (5) (1995) 33–44.
[37] D. Jitnah, J. Han, P. Steele, Software Requirements Engineering: An Overview, Technical Report 95-04, Peninsula School of Computing and Information Technology, Monash University, Melbourne, Australia, September 1995.
[38] H.R. Beyer, K. Holtzblatt, Apprenticing with the customer, Commun. ACM 38 (5) (1995) 45–52.
[39] O. Dieste, N. Juristo, Systematic review and aggregation of empirical studies on elicitation techniques, IEEE Trans. Software Eng. 37 (2) (2011) 283–304.
[40] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, fifth ed., John Wiley & Sons, 2011.
[41] C. Wohlin, P. Runeson, M. Höst, M.C. Ohlsson, B. Regnell, A. Wesslén, Experimentation in Software Engineering, second ed., Springer, 2012.
[42] L.A. Macaulay, Requirements Engineering, Springer Verlag Series on Applied Computing, Berlin, 1996.
[43] D. Pan, D. Zhu, J. Kim, Requirements Engineering Techniques. <http://sern.ucalgary.ca/~springl/Seng611/Seng611.html> (1997).
[44] R. Hoffman, The problem of extracting the knowledge of experts from the perspective of experimental psychology, AI Magaz. 8 (1987) 53–66.
[45] M. Christel, K. Kang, Issues in Requirements Elicitation, SEI Technical Report No.: SEI-92-TR-012, Carnegie Mellon Software Engineering Institute, September 1992.
[46] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, T. Carey, Human–Computer Interaction, Addison-Wesley, Reading MA, 1994.
[47] D. Maurer, T. Warfel, Card Sorting: A Definitive Guide, Boxes and Arrows, 2004.
[48] G. Rugg, P. McGeorge, Laddering knowledge elicitation, Exp.-Syst. 12 (4) (1995) 339–346.
[49] F.K. Shuttleworth, A study of questionnaire technique, J. Educ. Psychol. 22 (1981) 652–658.
[50] D.A. Dillman, Mail and Internet Surveys: The Tailored Design Method, second ed., Wiley, 2000.
[51] K.A. Ericsson, H.A. Simon, Protocol Analysis: Verbal Reports as Data, MIT Press, Cambridge, MA, 1984.
[52] C. Lewis, J. Reiman, Task-Centered User Interface Design: A Practical Introduction, University of Colorado, Boulder, CO, 1993.
[53] M. Easterby-Smith, The design, analysis and interpretation of repertory grid, in: M.L.G. ShawRecent (Ed.), Recent Advances in Personal Construct Psychology, Academic Press, London, 1983.
[54] B. Gonzalez-Baixauli, M. Laguna, J.C.S. do Prado Leite, Applying personal construct theory to requirements elicitation, IEEE Latin Am. Trans. 3 (1) (2005).
[55] A.F. Osborn, Applied Imagination, Charles Scribner's Sons, New York, 1953.
[56] E. de Bono, Serious Creativity, Harper Business, New York, US, 1992.
[57] A.L. Delbecq, A.H. Van de Ven, Nominal group techniques for involving clients and resource experts in program planning, in: Proc. Academy of Management, 1970, pp. 208–227.
[58] Y. Liou, Collaborative knowledge acquisition, Expert Systems with Applications, vol. 5, Pergamon Press Ltd., 1992, pp. 1–13.
[59] N.K. Denzin, Y.S. Lincoln, Handbook of Qualitative Research, Sage, London, 1994.
[60] D.L. Morgan, Successful Focus Groups: Advancing the State of the Art, Newbury Park, Sage, Calif., 1993.
[61] O. Helmer, N. Rescher, On the epistemology of the inexact sciences, Manage. Sci. 6 (1) (1959).
[62] H.A. Linstone, M. Turoff, The Delphi Method, Addison-Wesley Publishing Co., Inc., London, 1977.
[63] F. Hamilton, Cushing, University of Arizona Press, Zuni Folk Tales, 1999.
[64] J. Hughes, V. King, T. Rodden, H. Anderson, The Role of Ethnography in Interactive Systems Design, ACM Interactions, April 1995, pp. 57–65.
[65] S. Andriole, Fast, cheap requirements: prototype, or else! Manager column, IEEE Softw. 11 (2) (1994) 85–87.
[66] Luqi, How to Use Prototyping for Requirements Engineering, International Symposium on Requirements Engineering, Los Alamitos, IEEE Computer Society Press, CA, January 1993, p. 229.
[67] D.C. Andrews, JAD: a crucial dimension for rapid applications development, J. Syst. Manage. (1991) 23–31.
[68] J. Wood, D. Silver, Joint Application Design: How to Design Quality Systems in 40% Less Time, Wiley, NY, 1989.
[69] I. Jacobson, Object Oriented Development in an Industrial Environment, OOPSLA, 1987.
[70] M. Erdmann, R. Studer, Use-cases and scenarios for development knowledge-based systems, in: J. Cuena (Ed.), Proc. Of the 15th IFIP World Computer Congress, WCC'98, Conference on Information Technologies and Knowledge Systems, IT and KNOWS'98, 1998, pp. 259–272.