Empirical Software Engineering

Report 3: Analyzing and Comparing Case Studies

Group 15 **Search Year** 18
181250060 蒋祚竑 181250066 李镔达
181250090 刘育麟 181870108 李泳劭

1. Introduction

We collected totally 43 experimental studies papers published in 2018 while 6 of them from EASE, 25 from EMSE and 12 from ESEM. In this period we found some errors, finally we fix the result and select some papers.

1.1 Rectification

After re-judging the papers we selected in previous work, we find that there are some mistake of judgement about case study papers and update the involved result as followed:

1.1.1 EASE

EMPIRICAL METHOD	EMPIRICAL STUDIES	COUNT
Experiments	[1],[5],[10],[14],[22],[23]	6
Case studies	[2],[6],[9],[24],[25]	5
Survey	[11],[12],*[15]*,[16],[17],[20],[21],[26]	8*
Action research	[3],[13]	2
Systematic literature review	[4],[18]	2
Ethnographies	[7],[8]	2
Expert opinion		0
other/can not judge	[19]	1

* From case studies to survey

1.1.2 EMSE

EMPIRICAL METHOD	EMPIRICAL STUDIES	COUNT
Experiments	[1],*[3],[5],[6]*,[12],*[15],[17],[44]*,[18],[24],[25],[26],[27],[29],[38],[40],[43],[50],[51],[52],[57],[67],[78]	23*
Case studies	[8],[10],[13],[23],[35],[54],[58],[59],[63],[64],[65],[66],[70],[71],[72],[75],[77],[80]	18

EMPIRICAL METHOD	EMPIRICAL STUDIES	COUNT
Survey	[9],[11],[16],[19],[33],[37],[39],[41],*[46]*,[68],[69],[73],[74], [79]	14[†]
Action research	[14],[30],[36],[49],[60],[61]	6
Systematic literature review	[7],[21],[28],[32],[34],[42],[45],[47],[53],[55],[56]	11
Ethnographies	[4],[22],[62]	3
Expert opinion	[2],[20],[31]	3
other/can not judge	[76]	1

[*] From case studies to experiments

[†] From case studies to survey

1.2 Chosen

We choose 5 papers for further analysis and comparison. They are EASE[17], EMSE[13] [71], ESEM[18][5].

1.2.1 EASE[17]

Power and Politics of User Involvement in Software Development

Abstract

[CONTEXT] Involving users in software development is a complex and multi-faceted concept. Empirical research that studies power and politics of user involvement in software development is scarce.

[OBJECTIVE] In this paper, we present the results from a case study of a software development project, where organizational politics was explored in context of user involvement in software development.

[METHOD] We collected data through 30 interviews with 20 participants, attending workshops, observing project meetings, and analysing projects documents. The qualitative data was rigorously and iteratively analyzed.

[RESULTS] The results indicate that the politics was a significant factor used to exert power and influence in decision-making processes. Communication channels were exploited for political purposes. These contributed to the users' dissatisfaction with their involvement thus impacting on the project outcome.

[CONCLUSION] Having multiple teams of stakeholders with different levels of power in decision-making, the politics is inevitable and inescapable. Without careful attention, the political aspect of user involvement in software development can contribute to unsuccessful project.

Citation

> *Power and Politics of User Involvement in Software Development*
> *Muneera Bano, Didar Zowghi, Francesca da Rimini*
> *10.1145/3210459.3210477*

1.2.2 EMSE[13]

An industrial case study on the use of UML in software maintenance and its perceived benefits and hurdles

Abstract

UML is a commonly-used graphical language for the modelling of software. Works regarding UML's effectiveness have studied projects that develop software systems from scratch. Yet the maintenance of software consumes a large share of the overall time and effort required to develop software systems. This study, therefore, focuses on the use of UML in software maintenance. We wish to elicit the practices of the software modelling used during maintenance in industry and understand what are perceived as hurdles and benefits when using modelling. In order to achieve a high level of realism, we performed a case study in a multinational company's ICT department. The analysis is based on 31 interviews with employees who work on software maintenance projects. The interviewees played different roles and provided complementary views about the use, hurdles and benefits of software modelling and the use of UML. Our study uncovered a broad range of modelling-related practices, which are presented in a theoretical framework that illustrates how these practices are linked to the specific goals and context of software engineering projects. We present a list of recommended practices that contribute to the increased effectiveness of software modelling. The use of software modelling notations (like UML) is considered beneficial for software maintenance, but needs to be tailored to its context. Various practices that contribute to the effective use of modelling are commonly overlooked, suggesting that a more conscious holistic approach with which to integrate modelling practices into the overall software engineering approach is required.

1.2.3 EMSE[71]

The impact of human factors on the participation decision of reviewers in modern code review

Abstract

Modern Code Review (MCR) plays a key role in software quality practices. In MCR pro☐cess, a new patch (i.e., a set of code changes) is encouraged to be examined by reviewers in order to identify weaknesses in source code prior to an integration into main software repos☐itories. To mitigate the risk of having future defects, prior work suggests that MCR should be performed with sufficient review participation. Indeed, recent work shows that a low number of participated reviewers is associated with poor software quality. However, there is a likely case that a new patch still suffers from poor review participation even though reviewers were invited. Hence, in this paper, we set out to investigate the factors that are associated with the participation decision of an invited reviewer. Through a case study of 230,090 patches spread across the Android, LibreOffice, OpenStack and Qt systems, we find that (1) 16%-66% of patches have at least one invited reviewer who did not respond to the review invitation; (2) human factors play an important role in predicting whether or not an invited reviewer will participate in a review; (3) a review participation rate of an invited reviewers and code authoring experience of an invited reviewer are highly associated with the participation decision of an invited reviewer. These results can help practitioners better understand about how human factors associate with the participation decision of reviewers and serve as guidelines for inviting reviewers, leading to a better inviting decision and a better reviewer participation.

1.2.4 ESEM[18]

Abstract

Background: Testing is one of the main methods for quality as- surance in the development of embedded software, as well as in software engineering in general. Consequently, test results (and how they are reported and visualized) may substantially influence business decisions in software-intensive organizations. Aims: This case study examines the role of test results from automated nightly software testing and the visualizations for decision making they enable at an embedded systems company in Sweden. In particular, we want to identify the use of the visualizations for supporting decisions from three aspects: in daily work, at feature branch merge, and at release time. Method: We conducted an embedded case study with multiple units of analysis by conducting interviews, questionnaires, using archival data and participant observations. Results: Several visualizations and reports built on top of the test results database are utilized in supporting daily work, merging a feature branch to the master and at release time. Some important visualizations are: lists of failing test cases, easy access to log files, and heatmap trend plots. The industrial practitioners perceived the visualizations and reporting as valuable, however they also mentioned several areas of improvement such as better ways of visualizing test coverage in a functional area as well as better navi- gation between different views. Conclusions: We conclude that visualizations of test results are a vital decision making tool for a variety of roles and tasks in embedded software development, however the visualizations need to be continuously improved to keep their value for its stakeholders.

Citation

> *Decision making and visualizations based on test results*
> *Per Erik Strandberg, Wasif Afzal, Daniel Sundmark*
> *10.1145/3239235.3268921*

1.2.5 ESEM[5]

Abstract

Background: Requirement engineering is often considered a critical activity in system development projects. The increasing complexity of software as well as number and heterogeneity of stakeholders motivate the development of methods and tools for improving large-scale requirement engineering. Aims: The empirical study presented in this paper aim to identify and understand the char- acteristics and challenges of a platform, as desired by experts, to support requirement engineering for individual stakeholders, based on the current pain-points of their organizations when dealing with a

large number requirements. Method: We conducted a multiple case study with three companies in different domains. We collected data through ten semi-structured interviews with experts from these companies. Results: The main pain-point for stakeholders is handling the vast amount of data from different sources. The foreseen platform should leverage such data to manage changes in requirements according to customers' and users' preferences. It should also offer stakeholders an estimation of how long a re- quirements engineering task will take to complete, along with an easier requirements dependency identification and requirements reuse strategy. Conclusions: The findings provide empirical evidence about how practitioners wish to improve their requirement engi- neering processes and tools. The insights are a starting point for in-depth investigations into the problems and solutions presented. Practitioners can use the results to improve existing or design new practices and tools.

Citation

> *Needs and challenges for a platform to support large-scale requirements engineering: a multiple-case study*
> *Davide Fucci, Cristina Palomares, Xavier Franch, Dolors Costal et al.*
> *10.1145/3239235.3240498*

1.3 Reason

- EASE[17]：This report is about an issue user involvement in Software Development which is easy and attractive for us because we will usually face this problem in the work that many system requirements are not match user requirement. User involvement in Software Development maybe solve this problem.
- EMSE[13][71]：*1.* These 2 papers has a clear article structures so that I can easily find their case study design and analysis procedure. *2.* The EMSE[13] is an interview-evidence case study and the researchers generate a series of theories, which make it very special. As for EMSE[71], its research questions are interesting and realistic.
- ESEM[18][5]：*1.* Structure and description is clear. *2.* The design and procedure of the case study is rigorous. *3.* One is single case study and the other is multiple case study

1.4 Problems and solutions

1.4.1 Problems

- EASE[17]：because case study is a study method which many reports categorize inaccurately to this category, we find many wrong reports in case study
- EMSE[13][71]：The EMSE papers are all long papers that have more than 30 pages. It really takes time to build an understanding towards the paper and extract the information we need.
- ESEM[18][5]：too much reading work to do

1.4.2 Solutions

- Read slides provided in course to comprehend method of case study
- We find characteristics with slides aside so that the knowledge of papers can be identified
- We had some discussion when divergence appeared

2. Study Analysis and Comparison

2.1 Study Analysis

2.1.1 EASE[17]

Table 1: Methodological characteristics of EASE[17]

CHARACTERISTICS	EASE[17]
Objective	In this paper, we present the results from a case study of a software development project, where organizational politics was explored in context of user involvement in software development.
Participants	Our case study was conducted in a large financial institution of one of the State Government Organizations (SGO) in Australia. Our research participants were public servants working for SGO, who in many cases had been employed for anywhere between 1-4 decades.
Case amount	Single
Organization	Embedded
Triangulation	Data (source) triangulation

CHARACTERISTICS	EASE[17]
Data collection techniques	From an organization + About an individual + direct methods
Interview	Semi-structured interview
Interview mode	Funnel mode
Source	Interview
Research question	1. How does organizational power and politics influence user involvement in software development?
Quantitative Analysis	First, the audio interviews were professionally transcribed after which we used NVivo for coding the data against the subjects of the interview questions and thematic analysis. As more themes emerged during the coding process, more nodes and sub-nodes were added to the analytical schema in NVivo. These thematic categories highlighted the complex nature of software development and implementation, and the many technical and social interdependencies at play. In total, we coded the data against a possible 86 themes on which to perform the next stage of analysis, comparing and contrasting individuals' perceptions of the software project. Such a fine-grained analysis provided us with an extremely rich data set from which to identify patterns, trends, points of comparison and areas of potential conflict.

CHARACTERISTICS	EASE[17]
Qualitative Analysis	In order to understand how politics manifested in UI-SS relationship, we were focusing on the power held by different people within organization. To determine in whom power resided and how they exercised their power, we analyzed participants' comments about who made crucial decisions about project personnel, budgets and time-frames. To better understand the power dynamics operating between various parties, we looked at if and how users and stakeholders had challenged contentious decisions and processes, and how those in authority had responded. We noted when people expressed feeling frustrated and/or powerless to draw attention to, or mitigate, software-related technical problems affecting business processes; such accounts indicated that these individuals held a relatively lesser degree of agency in the SDP. Finally, we revisited what people had said about their own and others' roles in the project in terms of decision-making, delegation and authorization powers. In some cases, perceptions made by one respondent were echoed by a number of others. In other cases, experiences of involvement or the project outcome would dramatically differ amongst subjects, often it would seem according to the user's role in the project. We used such points of comparison and contrast to develop insights.

Table 2: Validity of EASE[17]

VALIDITIES	EASE[17]
Conclusion Validity	In our case study, we observed the similar phenomenon; the users were dissatisfied with their involvement and this contributed to their dissatisfaction about the project outcome (product/system).
Internal Validity	Our case study was exploratory and interpretive in nature [29]. In the interpretive research paradigm the role of researcher is challenging in order to observe, understand and analyze the interpretations of empirical data as reported based on the experiences of the people.
Construct Validity	User involvement (UI) in software development (SD) has been studied extensively for more than four decades. Many early researchers considered it axiomatic that user involvement has direct positive influence on successful outcome of the system.
External Validity	Our analysis has revealed that organizational politics can contribute to user dissatisfaction about their involvement, thus negatively influencing the project outcomes.

2.1.2 EMSE[13]

Table 3: Methodological characteristics of EMSE[13]

CHARACTERISTICS	EMSE[13]
Objective	Discover what practices industrial software professionals use when using UML, as well as to find out how they perceive the effectiveness of software modelling, paying particular attention to software maintenance tasks.
Participants	Employees from the ICT Department of a multinational transport company in Western Europe
Case amount	Single
Organization	Embedded
Triangulation	Data triangulation & Investigator triangulation & Methodological triangulation
Data collection techniques	From an organization(department of a multinational transport company) about an individual(employees)
Interview	Semi-structured interviews
Interview mode	Funnel mode
Source	Interview
Research question	RQ1: What practices are involved in using UML in software maintenance projects? RQ2: What are the costs-factors and benefit-factors of using UML in software maintenance projects? RQ3: What are the factual and perceived hurdles when maintaining documentation, and UML models as part of that documentation? RQ4: What are best practices when using diagramming and modelling in documentation?

CHARACTERISTICS	EMSE[13]
Qualitative Analysis	Theory generation "We analysed each transcription, highlighting the important and surprising statements. This was done through the simultaneous use of NVivo 10 (Richards 1999) and Word **by means of open thematic coding.** This coding was carried out independently by two researchers, and then discussed. **We then coded the statements and grouped them under more general categories or factors** (Seaman 1999). We also used NVivo 10 and Word to perform this step." We **decided to generate a theory as the final result** of the grounded-theory process

Table 4: Validity of EMSE[13]

VALIDITIES	EMSE[13]
Conclusion Validity	" this relates to the ability to draw a correct conclusion from a study. The chain of evidence from the interviews and documentation analysed through to the synthesized evidence was maintained using a word-for-word transcription. This analysis took a long time to carry out, but this was due in part to our desire to ensure that we did not make mistakes in the interpretation while the analysis was being under□taken. We therefore asked the interviewees to give feedback to the researchers on the transcripts of the interviews. This practice is known as 'member checking,' and it was used continuously to obtain feedback on both the transcripts and the analyses. Tools were also used during the analysis of the data. Furthermore, the individual coding performed by each researcher was discussed by them, so that they could verify and reach an agreement on them. "
Internal Validity	" The main threat to the internal validity of this study concerns our ability to control influences from other factors beyond those which have been accounted for in this study. For example, the age of the interviewees, the relationship of the interviewees with their team members, or their motivation, might be influential factors as regards being for, or against, the use of UML. Also the consideration of the term UML as a synonym of Rational Unified Process or even Object Orientation might be an internal threat to validity to this study."

VALIDITIES	EMSE[13]
Construct Validity	in the data collection multiple sources of evidence were used. Also the transcript of interviews and observations were sent back to the interviewees to enable correction of raw data. In addition, analyses were presented to them and to the internal research supervisor, in order to maintain their trust in the research. The validity of the developed theory would need to be tested in other case studies
External Validity	": this concerns limitations as regards generalising the results of a study to a broader industrial practice. The sample of the case study and interviewees might be a threat to the validity of this study, although the sampling process was as randomised as possible. We acknowledge the fact that using only one case study may limit the generalisability of the results of this study. However, we believe that reporting these early findings is necessary, as it serves as an encouragement for other researchers to replicate our study using different case studies. The generalisation of the results might be extended to cases which have common characteristics. On the other hand, an interview provides Bspontaneous recall^ of an answer if it lists a concrete example/instance on an open question: BQ: For example, which diagrams do you use?^ BA: Class diagrams.^ If in a subsequent question, we ask a closed question: BQ: Do you also use sequence diagrams?^ BA: sometimes, but not always.^ Then this illustrates that the person actually does also use some other diagram, but needs to be triggered/queued to say so. In interviews it is not always possible to get the interviewee to recall the exact information that is relevant to the question/research. From this perspective, an interview study should not be considered to be complete or accurate in a quantitative sense. Nevertheless, quantitative analyses sometimes provide clear indications of trends"

2.1.3 EMSE[71]

Table 5: Methodological characteristics of EMSE[71]

CHARACTERISTICS	EMSE[71]
Objective	To investigate the factors that are associated with the participation decision of an invited reviewer in modern code review
Participants	230,090 patches spread across the Android, LibreOffice, OpenStack and Qt systems
Case amount	Multiple

CHARACTERISTICS	EMSE[71]
Organization	Holistic
Triangulation	Data (source) triangulation
Data collection techniques	From an organization(large software systems such as Android, LibreOffice, OpenStack etc) about an individual(patch)
Interview	/
Interview mode	/
Source	Documentation
Research question	RQ1: How often do patches suffer from the unresponded review invitations? RQ2: Can human factors help determining the likelihood of the participation decision of reviewers? RQ3: What are the factors mostly associated with participation decision?
Qualitative Analysis	**RQ1:Quantitative Analysis- Descriptive statistics**: "To address the RQ1, we analyze descriptive statistics of the number of reviewers who did not respond to the review invitation of patches. In particular, we count how many patches that have reviewers who did not respond to the review invitation. Furthermore, we investigate whether or not inviting many reviewers can decrease a chance of having an invited reviewer Empirical Software Engineering who did not respond to the review invitation." **RQ2:Quantitative Analysis-predictive modelling**: "To address our RQ2, for each dataset, we construct two nonlinear logistic regression models to predict whether or not an invited reviewer will participate in the review. **RQ3:Quantitative Analysis-predictive modelling**: "We analyze nonlinear logistic regression models to determine the performance of the models, and to quantitatively understand the relationship between the independent variables and the participation decision of an invited reviewer."

Table 6: Validity of EMSE[71]

VALIDITIES	EMSE[71]

VALIDITIES	EMSE[71]
Internal Validity	"We identify whether or not an invited reviewer participated in a review of a patch using a review score and comments that are posted in the patch. However, it is possible that the invited reviewers perform code review through other communication media such as in person discussion (Bacchelli and Bird 2013; Guzzi et al. 2013), a group IRC (Shihab et al. 2009) or a mailing list (Rigby et al. 2008). Since we identify the participation decision based on comments and review score, performing code review outside of the platform may lead to an inaccurate participation decision. "
Construct Validity	"We compute our studied metrics at the creation time of patches. Unfortunately, the Gerrit code review tool does not record when the author invites a reviewer. Hence, we must rely on this heuristic and assume that all reviewers are invited at the same time as the creation time of patches."
External Validity	"We perform a study on four open source software systems that use the Gerrit code review tool, which may limit the generalizability of our results. Additionally, we find that there is a possibility that the same metric performs differently for different systems (e.g., the reviewing experience of an invited reviewer)."

2.1.4 ESEM[18]

Table 7: Methodological characteristics of ESEM[18]

CHARACTERISTICS	ESEM[18]
Objective	The objective of the case study is to explore the use of the TRDB and the visualizations it supports for the three critical decision points of: daily work, merging a code branch to the master branch, and at release time.
Participants	We do this by conducting an industrial case study at Westermo Research and Development AB (Westermo), which designs and manufactures robust data communication products for mission-critical systems. the units of analysis in our case are the different roles we interviewed and sent questionnaire to. These roles are a project manager, a test manager, two developers and one test environment developer.
Case amount	Single

CHARACTERISTICS	ESEM[18]
Organization	Embedded
Triangulation	Methodological triangulation(interviews, questionnaires (as follow-up to interviews), archival data and participant observations.)
Data collection techniques	From an individual + About an individual + direct methods
Interview	Semi-structured interview
Interview mode	Funnel mode
Source	Archival records Interviews Participant-observation
Research question	RQ1: How are visualizations used for making decisions in daily work, at merging a code branch to a master branch and at release time? RQ2: How do industrial practitioners perceive the value of visualizations, both in general and with respect to decision making in daily work, at merging a code branch to a master branch and at release time?

CHARACTERISTICS	ESEM[18]
Qualitative Analysis	transcript would be forgotten in a printer, etc. We identified 17 themes, and categorized interview snippets related to a theme as positive, neutral, or negative. For example, in one of the interviews the interviewee said "[Looking at log files] is very valuable, the part where you can see the log. I'd like to have the functionality to see the kernel log, and those kinds of things. But there is a penalty to collect these." This was clearly a discussion on the theme logs, that was both positive (it is good to have easy access to logs) and negative (there could be more logs stored). The participant observations at sync meetings were summarized on a higher level than the interview transcripts, and were categorized in the same way as interview snippets. For example, during one sync meeting we noticed that the participants did an in-browser zoom of the web page and reacted when the font appeared to be broken. This was noted as "poor font zoomability" and categorized as negative under the theme of UI. The data from the questionnaires were on a Likert scale, which was used to show respondents perception of the usefulness, ease of navigation, learnability, understandability and time consumption of visualizations. The archival data available through, e.g., the doc- umentation on internal wiki pages, was used to explain the existing visualization support and was primarily used by the authors to both get familiarized with context and to understand the meaning of interview/questionnaire data.

Table 8: Validity of ESEM[18]

VALIDITIES	ESEM[18]
Conclusion Validity	In order to improve validity of the results we collected data from several sources with different techniques (method triangulation) and none of the interviews were conducted by only one researcher (observer triangulation). Further, by describing the case and provid- ing both the interview questions and the questionnaire, we have simplified the possibility for a replication study.

VALIDITIES	ESEM[18]
Construct Validity	Threats to the construct validity include the limits of scope: decision making was limited to three key points in the development process; we limited ourselves to interview five individuals with the four roles of developer, tester, project manager and test manager; and we limited the visualizations to already existing visualizations. However, we opted for interviewing key individuals belonging to different groups, and used the knowledge of the industrial co-author and the test manager as a criterion for interviewee selection.
External Validity	The study was conducted at Westermo, an embedded systems company in Sweden, doing agile software development, with a strong focus on test automation, using an internally developed test results database. Results on the available visualizations at Westermo are of course not generalizable, but we provide many generalizable observations on how an individual having a role can use a visualization of test results in order to enhance decision making.

2.1.5 ESEM[5]

Table 9: Methodological characteristics of ESEM[5]

CHARACTERISTICS	ESEM[5]
Objective	The objective of this study is to understand what companies consider an innovative RE platform.
Participants	Data were directly collected using semi-structured interviews with ten company representatives
Case amount	Multiple
Organization	Holistic
Triangulation	Data (source) triangulation
Data collection techniques	From an individual +About an organization+direct methods
Interview	Semi-structured interview
Interview mode	Funnel mode
Source	Interviews

CHARACTERISTICS	ESEM[5]
Research question	RQ1: What are the pain-points faced by the companies in the way they currently deal with large-scale RE? RQ2: What are the needs that a platform supporting large- scale RE should address? RQ3: What are the challenges of introducing such a platform in the case companies?
Quantitative Analysis	We performed qualitative data analysis based on the transcripts obtained as described in [20]. We coded the transcripts using a line- by-line approach, in a semi-exploratory fashion. The initial codes were based on the main pain-points in RE activities, articipants wish for solutions, and foreseen challenges. For each transcript relative to a company, each relevant statement in the transcribed interview was assigned codes by two researchers following an open coding approach

Table 10: Validity of ESEM[5]

VALIDITIES	ESEM[5]
Conclusion Validity	As it is usual for case studies, we do not claim strong generalization of our findings. We showed that the results apply to different extent within the case companies. We make this explicit when reporting the findings by indicating the context in which they are more sensible. Only some of our results (e.g., the pain-point of information overload) generalize over the three cases. For these results, we do not claim statistical generalization to a pre-defined population of companies due to the low number of cases and the non-probabilistic sampling approach driven by the project settings. However, they are important in the context of the project to, for example, prioritize what needs should be fulfilled when developing a solution. We claim analytical generalization of our findings by expanding the theory current on RE-related pains and needs presented in Méndez et al. [7] when considering companies dealing with large-scale requirements in a bespoke or market-driven domain.

VALIDITIES	ESEM[5]
Construct Validity	This threat can arise once there is not a shared understanding of the terms used in the interviews between inter- viewers and interviewees. In our study, the interviews were based on concepts from the domain of the stakeholders who participated in the study. The company representatives were involved in the creation of the interview script. At the beginning of each interview session, we dedicated time to go through the status quo of the company with the participants. In addition, some of the interviews were carried out with managers and some of their employees. In our study, we believe this is no big issue since: 1) since the com- panies where the interviews where this happened do not follow a hierarchical employee structure (i.e., there is no fear to managers), and 2) The employees were really participative in the interviews. However, we cannot control some information was missed out be- cause of that. Triangulation with other data collection methods (e.g., observation) could have improved construct validity.
External Validity	As it is usual for case studies, we do not claim strong generalization of our findings. We showed that the results apply to different extent within the case companies. We make this explicit when reporting the findings by indicating the context in which they are more sensible. Only some of our results (e.g., the pain-point of information overload) generalize over the three cases. For these results, we do not claim statistical generalization to a pre-defined population of companies due to the low number of cases and the non-probabilistic sampling approach driven by the project settings. However, they are important in the context of the project to, for example, prioritize what needs should be fulfilled when developing a solution. We claim analytical generalization of our findings by expanding the theory current on RE-related pains and needs presented in Méndez et al. [7] when considering companies dealing with large-scale requirements in a bespoke or market-driven domain.
Internal Validity	This study does not involve the assessment of causal relationships using statistical methods, as the case study is purely descriptive. The recommendation presented in Section 5.1 are based on our interpretation of the codes, their co-occurrence, and their importance as perceived by the participants. The themes emerged after at least two researchers reviewed the codes.

VALIDITIES	ESEM[5]
Reliability	We strengthened the reliability of the interview scripts by running an internal pilot; moreover, we used established cod- ing techniques and tools to code the collected data. On the other hand, we acknowledge that the researchers, as well as some of the interviewees, had in-depth knowledge on the context in which the case studies were carried out—e.g., the OpenReq project. Therefore, other researchers wanting to carry-out the same study need to familiarize with the background and objectives of the project.

2.2 Comparison

1. Data Source & Participants

   In the 5 papers, only data source of EMSE[71] comes from documentation while others' from interviews. So there's only EMSE[71] conducting with non-human participants.

2. Case Study Design

   There are 2 different designs in these 5 papers, they are single-embedded study of EASE[17], EMSE[13] and ESEM[18] and multiple-holistic of EMSE[71] and ESEM[5]. Normally, there would be too much work as parts of the whole study if the number of unit of analysis in each case exceed 1 thus no multiple-embedded study included, in contrast, single-holistic study would lack reliability so it's not included as well.