Using Case-Based Methods in an Experimental Design: A Mixed-Method Approach for Evaluating Collaboration-Intensive Software Modeling Approaches

Anil Kumar Thurimella*

Abstract

Experiments are suitable for studying emerging approaches at Universities by using students as subjects. We identify a problem that the existing experimental designs are not well suited for studying qualitative properties and human factors which cannot be controlled and studied in isolation. On the other hand case studies are well suited for studying those situations. However, case studies are meant to be used to study real life situations and are not suitable for experimental studies with students. To address the problem, we propose an Experiment Embedded with Case-based Methods (EECM) by adding case-based methods to a traditional experimental design. EECM would provide flexibility to the researches for collecting and analyzing data with mixed-methods. We have applied the EECM design for the evaluation of a software reuse approach in a large quasi-experiment with students in the software engineering lecture environment.

Motivation

The case study methodologies have been considered important for the software engineering community [1]. Case studies [2] are mainly used when the extent of control over contemporary events is small. Furthermore, case studies are implemented in real-life situations where the aspects of an approach are difficult to separate and study. Let us consider a situation where an architect has to decide on a design issue. In the situation, there are several factors that influence the decisions which are difficult to control and study in isolation:

- Architectural decision-making is communication intensive - the architect needs to

* Anil Kumar Thurimella, Harman & Chair for Applied Software Engineering, Technical University of Munich Email: anilthurimella@gmail.com

collaborate with relevant stakeholders, collect arguments and evaluate alternatives.

- Human factors such as motivation, enthusiasm, leadership, creativity etc influence decision-making.
- The decision-making is also dependent on the availability of legacy documentation of design decisions and the quality of the available documentation.

Evaluating human factors has been considered significant. For example, human *motivation* is a key factor for obtaining high quality output (e.g. design decisions, code, software reuse etc) [3]. Several software engineering activities involves collaboration between stakeholders of various backgrounds. Studying *communication* is critical because communication breakdowns reduce productivity and increase failures in the system [4]. Another human factor is *creativity* [5]. Enabling creativity in the development process would aid stakeholders to introduce novel and alternative features in the system. *Usability* of processes and tools is another key human factor as well.

For evaluating such human factors, qualitative and quantitative data would be used. For example, qualitative data could be collected on how users of the architectural decision-making approach use creativity while qualitative data could be obtained by assessing the design developed. Case studies are well-suited for handling qualitative and quantitative data, that is, for supporting mixed-methods. Mixed-method research has been encouraged across various case study application domains (e.g. [11, 12, 13]). To get rich data form mixed-method analysis, case studies combine multiple-data collection techniques such as observations, interviews and analysis of design/code developed. Traditional mixed-method designs for studying human factors do not focus on case studies. For example, Dov et al. used mixed-methods to study human factors [15]. However, Dov's study is based on an ethnographic design.

Based on the above described arguments, we believe that case study methods are suitable for studying several software engineering approaches involving human-factors including collaboration. Cases studies are meant to be used in real life with rich context (e.g. industrial evaluation [14]). However, case studies are not suitable for implementing pilot/evaluation studies at Universities. Researchers from Universities contribute new modeling approaches and they are to be evaluated at Universities before their introduction to the industrial environments. Furthermore, it is often difficult for the researchers to evaluate the approaches directly at the industry because of issues such as confidentiality of industrial specifications.

To evaluate research contributions at Universities, experiments have been used with students as representatives for professionals with some restrictions [6]. Traditional experimental designs (e.g. controlled experiments) [7] are useful to evaluate casual relationships, that is, identify effect of independent variables on dependent variables. The experiments mainly yield quantitative evidences, which is not sufficient to study qualitative properties.

In order to make use of the strengths of case studies for the evaluations at academic settings, we included several case-based methods in a typical experimental design to contribute a new methodology called *Experiment with Embedded Case-based Methods (EECM)*. We illustrate EECM based on the quasi-experiment [8] that was recently reported. We describe the high-level concepts of EECM in section 2 based on an overview table. We conclude the paper in section 3 with an outlook for the future enhancements of the methodology.

EECM was tested in a quasi-experiment [8] which focuses on the evaluation of a decision-oriented software reuse approach called IVM (Issue-based variability modeling). However, EECM was not proposed in quasi-experimental report [8]. In this paper, we do not review the findings of IVM as well. Moreover, we use high-level concepts of the study for illustrations.

EECM Methodology

An overview of the EECM concepts and their application details in the quasi-experiment are presented in Table 1. We structure the table based on the following concepts:

- *ID* representing the identifier and hierarchical level of the property.
- *Property* of an empirical study relevant to EECM.
- *Concept* proposed in this paper on how the property is supported in EECM. These concepts are already established in the case study or experimentation community. For example, using students as subjects (row with ID = 2) is a common practice in the

experiments while using an embedded design is a common practice in the case study community (row 9). As EECM includes a blend of concepts form the case study and experimentation community, we represent the concepts on a uniform level.

- The *Reason* for proposing the concept.
- A concept of the table is used in the experimentation, case study or both these communities. For example, Aim/Purpose is used in both experiments and case studies whereas, multiple data collection methods (row 10) is popular in the case study community. We represented the original *Community* where the concept has been used to clarify where the concept is derived from. In particular, we use the selections *Experimentation*, *Case study* and *Both* (generic concept in case study and experimentation, e.g. aim of a study).
- The *Application* of the concept in the previously reported study [8].

Table 1. An overview of the empirical concepts of EECM methodology.

ID	Property	Concept	Reason	Community	Application
1	Aim/Purpose	Study qualitative and quantitative properties of an approach in an academic setting. Originally, the approach should have been developed for the industry.	Focus on qualitative and quantitative properties including human factors.	Both	Studied motivation, creativity, reuse and collaboration related issues of IVM.
2	Subjects	Students represent professionals at industry.	Availability for evaluation studies in a lecture environment.	Experiment ation	Around 250 students from multiple study programs.
3	Implementer	Use of neutral mentors between researcher and subjects.	Reducing researcher's bias. Aiding the collection of observational data by the mentors.	Case study	Four tutors were used as neutral mentors.

4	Environment	Lecture or a lab course environment.	Possibility to give tasks to the students.	Experiment ation	Software engineering lecture environment at TU Munich.
5	High-level design	• An experimental setting if groups would be created randomly or with systematic sampling (e.g. stratified random sampling). • A quasi-experimental setting if the lecture groups are used based on the restrictions from the study programs (reduced control on creating groups).	Imitate teams in a real project/environment at Universities.	Experiment ation	A quasi-experimental setting was used because it was restricted to use lecture groups as the study groups. Within the groups we separated roles similar to the industry.
6	Training	The subjects are trained by the mentors in the lecture sessions.	Subjects get required skills to perform tasks.	Both	Training sessions were implemented by mentors.
7	Tasks	Tasks are designed on the approach to be evaluated and have to be key elements of the course (e.g. assignments) itself. The tasks are likely to be performed in a team.	Basis for the students to perform activities of the approach.	Experiment ation	Three tasks were performed in all the groups of the quasi-experiment.
8	Low-level design	A team using the approach as a unit-of-analysis.	Focal point for collecting rich data.	Case study	Team performing a task was considered as a unit-of-analysis.
9	Low-level design	Embedded design with multiple instances of the unit of analysis with a variety.	Multiple replications and variations in the properties of the units-of-analysis improve the strength and the validity of the results.	Case study	Around 50 instances of unit of analysis with team size varying from 2-22. The team composition with respect to participants was also different across the teams.

10	Data collection	Multiple data collection methods such as observations, responses from subjects to a 1questionnaire and analyzing the records developed by subjects.	Rich data collection.	Case study	Mentors collected observational data, subjects responded to a questionnaire and the researchers analyzed the artifacts developed by students.
11	Data analysis	Mixed-methods.	Combining qualitative and quantitative data for strong conclusions.	Case study	The study yielded 9 results: • Three were based pure qualitative evidences, • Four were based on a combination of qualitative and quantitative data. • Two were based on pure qualitative evidences. Except for the two results based on pure qualitative evidences, the other results have strong evidences.
11.1	Quantitative analysis	Regression analysis	Effect of independent variable on a dependent variable.	Experimentation	Five variables were measured for checking causal relationships (e.g. team size, average motivation, quality of output, amount of reuse and quality of reuse).
11.2	Qualitative analysis	Data triangulation	Improving validity of qualitative data.	Case studies	Evidences on communication, creativity and usability were obtained based on data triangulations.
11.3	Combining qualitative and quantitative evidences	• Use of both open and closed questions in the questionnaire to subjects. • Collect qualitative and quantitative	Strong conclusions	Case studies	Combinations of qualitative and quantitative data used in the quasi experiment are represented in Table 2.

		data from observations. • Combine qualitative and quantitative data from all instruments using data triangulation.			

The initial versions of the concepts are identified and used during the designing the quasi-experiment to address our needs. For example, we wanted to reduce researcher's bias with students and by reviewing the literature we learned that case studies recommend neutral mentors for this purpose. Similarly, we had possibility to experiment with students based on the tasks, which lead to the concepts on subjects, environment and tasks. We did not want to overload the students with large questionnaires. As an alternative to collect sufficient data we designed the experiment with multiple data collection methods. To report the concepts in this paper, we structured them on a uniform level in the form of an overview table. All the rationales are documented in the Reason column.

We believe that the list of properties is not complete. Rather, we want to highlight some of the key properties of EECM design within this paper so that researchers can adopt the table based on their needs in order to design a new study. For example, a researcher can use an alternative statistical method (e.g. correlation) instead of the regression analysis (row 11.1). Furthermore, the researcher can introduce a new row (at level 5.5) to perform the variance analysis on the demographic information (e.g. culture, experience etc) of the experimental groups. Similarly, researchers interested in replicating EECM-based studies may extend the table with new properties such as meta-analysis (for quantitative data) and cross case analysis (for qualitative data). The properties are already classified based on case study/experimentation so that researchers can review literature and include alternative and additional techniques.

To summarize, the key concepts described in the table, EECM methodology has been designed to evaluate collaboration-intensive approaches where there is significant quantitative and qualitative data including data from observations (see Aim). We emphasize the need for observational data because of multiple reasons: Several human related aspects would be observed and noted. Furthermore, students in a lecture environment would not have much time for responding to large questionnaires from the researchers while performing the tasks would be an element of the course. Therefore, observations would provide an alternative source for collecting data.

The key rationales for the design are the following. Adding unit-of-analysis aids collecting rich data by defining the focus of observations while embedded design aids to study replications in the data from the units-of-analysis. Embedded designs were encouraged for improving replications within case design [7, 8] because of the possibility to study replications of unit-of-analysis within a case. Furthermore, neutral mentors reduce the researcher bias in collecting qualitative data. In addition to this, mentors would collect observational data as much as possible contributing to a light-weight empirical study from the subjects' point of view.

The researcher would be able to use multiple and flexible data collection techniques with mixed-methods like a case study. Moreover, we have displayed regression analysis and data triangulation only as instances (see 11.1 to 11.3 in Table 1). Data triangulations may be used to combine qualitative and quantitative data. For, example Table 2 presents the combinations of qualitative and quantitative data that we used in our EECM implementation. We marked the cells with "Yes" where qualitative and quantitative data was combined (see Table 2). From the Table 2, it can be observed that there were data triangulations between qualitative and quantitative data from all the different sources of data with an exception of the mentor data. As mentors noted different points, we did not have data triangulations within the data collected by mentors (see the cell marked with No).

Table 2 - Combinations of Qualitative and Quantitative Data in the Quasi-Experimental Illustration

		Qualitative	
		Observations from mentors	**Responses of subjects from open questions**
Quantitative	**Observations from mentors**	No	Yes
	Variables from the analysis of artifacts	Yes	Yes
	Responses of subjects from closed questions	Yes	Yes

Based on our experience in evaluating EECM in the quasi-experiment, we noticed following benefits:

- EECM provided dedicated methods such as modeling a team as a unit-of-analysis and using an embedded design with a variety of teams, to study collaboration and human factors with a focus on mixed-methods.
- Researchers would be able to plan and implement empirical studies in a lecture environment in an academic setting. As a replica of industrial settings would be created, we experienced that students are able to learn about industrial tasks directly from the courses at Universities.
- The methodology provides a flexible way to design experiments with large number of students by systematically decomposing the experimental setting into units-of-analysis and enabling a flexible data collection and analysis. We have already demonstrated an implementation of a large quasi-experiment with around 250 subjects organized in around 50 teams with various team sizes.

EECM improves traditional experimental designs in the software engineering context (e.g. [7]) by including several methods that have been used in the case study community such as unit-of-analysis, embedded design, multiple data collection methods, neutral mentors and mixed-methods (see Community column with — Case Study). For example, a traditional experimental design mainly focuses on quantitative data as well as evaluating causal relationships between variables and does not use unit-of-analysis or embedded design. EECM differs from case-based designs enabling experimentation evaluation of contemporary aspects mainly with students. Traditional case study designs are suitable to evaluate contemporary aspects in real life situations only.

Conclusion and outlook

We have reported the gap in the design of experiments to evaluate approaches with several properties, particularly human-factors that cannot be controlled. To address the gap we embed several case-based methods in an experimental design. We illustrate the concepts based on a quasi-experiment which was implemented to study software reuse in a software engineering lecture environment with more than 250 subjects.

We implemented the methodology only once. In future, we would like to implement new empirical studies to evaluate the EECM design concepts. This would aid improving the EECM concepts as well as improves the external validity of the approach.

We would like to research on planning and implementing replication studies for improving external validity of study conclusions. It should be researched how to combine evidences from multiple EECM studies. For example, theory building with cross-case analysis could be used for combining data based on various properties while meta-analysis could be used to combine statistical data. Similarly, we should research on strategies to implement multiple EECM interrelated studies with different focuses, in cases of practical difficulties to implement large empirical studies.

References

[1] Runeson, P. Host, M. Rainer, Regnell, A.B. (2012) Case Study Research in Software Engineering: Guidelines and Examples, WILEY.

[2] Yin, R.K. (2003). Case study Research: Design and Methods, 3rd ed. Sage Publications.

[3] Beecham, S., Baddoo, N., Hall, T., Robinson, H., Sharp, H. (2008). —Motivation in software engineering: a systematic literature review . Information and Software Technology 50 (10). 1820 860–878.

[4] Cataldo, M. Herbsleb, J. D. (2013). —Coordination Breakdowns and Their Impact on Development Productivity and Software Failures . *IEEE Transactions on Software Engineering* 39(3). 343-360.

[5] Ocker, R.J. (2010). —Promoting group creativity in upstream requirements engineering . An Interdisciplinary Journal on Humans in ICT Environments 6 (1). 55–70.

[6] Hoest, M., Regnell, B., Wohlin, C. (2000). —Using students as subjects—a comparative study of students and professionals in lead-time impact assessment . Empirical Software Engineering Journal 5 (3). 201–214.

[7] Wohlin, C. Runeson, P. Host, M. Ohlsson, M.C. Regnell, B.,Wesslen, A. (2000). —Experimentation in Software Engineering: An Introduction . Springer.

[8] Thurimella, A.K. Bruegge, B. (2013) —A mixed-method approach for the empirical evaluation of issue-based variability modeling . The Journal of Systems and Software 86 (7). 1831-1849.

[9] Meyer, C. B. (2001). —A case in case study methodology . *Field methods*, *13*(4), 329-352.

[10] Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry*, *12*(2). 219-245

[11] Dunning, H. Williams, A. Abonyi, S. Crooks, V. (2008). —A mixed method approach to quality of life research: a case study approach . In: From the Issue Entitled Quality of Life Studies in Saskatoon: Evidence from Community–University Research Partnerships. Springer, pp. 145–158.

[12] Kaplan, B. Duchon, D. (1988). —Combining qualitative and quantitative methods in information systems research: a case study . MIS Quarterly 12 (4), 571–586.

[13] Bryman, A. (2006) —Integrating quantitative and qualitative research: how is it done? Qualitative Research, 6(1). 97-113.

[14] Dhungana, D. Grünbacher, P. Rabiser, R. (2011) —The DOPLER meta-tool for decision-oriented variability modeling: a multiple case study . Automated Software Engineering 18(1). 77-114.

[15] Dow, S., MacIntyre, B. Mateas, M. (2008). —Understanding Engagement: A Mixed-Method Approach to Observing Game Play . In *Workshop on Evaluating User Experiences in Game in CHI '08* (pp. 5-10).

****