Workshop Series on Empirical Software Engineering

Vol. 1

Christian Bunse
Andreas Jedlitschka (Eds.)

Empirical Studies in Software Engineering

First International Workshop, WSESE 2002
Rovaniemi, Finland, December 2002
Proceedings

ESERNET

Many academics and practitioners believe that evaluation has to play a vital in software engineering. As well as evaluating both application level and component level products, software engineers need to be concerned with the evaluation of development processes, engineering methods, the people involved, and supplier organizations. However, both academics and practitioners are concerned about the cost and effectiveness of applying many of the existing assessment methods. One of the aims of this workshop was to address the question of justifying the use of empirical assessment and evaluation in software engineering, as well as to hear about practical experiences. Both this theme, together with the other issues involved in evaluation, were addressed through technical presentations and experience reports. Thus, the workshop was a unique forum dedicated to the presentation and discussion of research and practical experiences addressing all aspects of empirical assessment and evaluation in software engineering.

Christian Bunse
Andreas Jedlitschka

ISBN 3-8167-6374-X

Workshop Series on Empirical Software Engineering

Workshop Series on Empirical Software Engineering

Volume 1
Christian Bunse, Andreas Jedlitschka (Eds.)

En

“

Procee
Empiri

F

Editors:
Christian Bunse, Andreas Jedlitschka
Fraunhofer Institute Experimental Software Engineering
Sauerwiesen 6
67761 Kaiserslautern
Germany

Printing:
Satz- und Druckcenter des Fraunhofer-Informationszentrum Raum und Bau IRB, Stuttgart

Printed on acid-free and chlorine-free bleached paper.

Fraunhofer-Informationszentrum Raum und Bau IRB
Postfach 80 04 69, D-70504 Stuttgart
Nobelstraße 12, 70569 Stuttgart
Germany
Phone +49(0) 711 / 970-25 00
Fax +49(0) 711 / 970-25 08
e-mail irb@irb.fraunhofer.de
URL www.irb.fraunhofer.de

Christi

Workshop Series on Empirical Studies in Software Engineering

1st Workshop
"Empirical Studies in Software Engineering"

Proceedings of the PROFES 2002 Workshop on Empirical Studies in Software Engineering

Rovaniemi, Finland, December 9th 2002

Christian Bunse and Andreas Jedlitschka (Eds.)

Acknowlegdements

This workshop has been conducted in the context of ESERNET[1] (Experimental Software Engineering Network) funded by the Information Societies Technology Programme of the European Union under grant IST-2000-26754.
The organizers would also like to thank the PROFES organization committee for hosting the workshop at the PROFES 2002 conference.

[1] http://www.esernet.org

Workshop „Empirical Studies in Software Engineering"
in conjunction with the 4th International Conference on
Product Focused Software Process Improvement (PROFES)

Rovaniemi, Finland
9.-11. December 2002

Outline & Goals

Many academics and practitioners believe that evaluation has a vital role to play in software engineering. As well as evaluating both application level and component level products, software engineers need to be concerned with the evaluation of development processes, engineering methods and supplier organizations. However, both academics and practitioners are concerned about the cost and effectiveness of applying many of the existing assessment methods. One of the aims of the workshop is to address the question of justifying the use of empirical assessment and evaluation in software engineering, as well as to hear about practical experiences. Both this theme, together with the other issues involved in evaluation, will be addressed through technical presentations and experience reports. Thus, the workshop is a unique forum dedicated to the presentation and discussion of research and practical experiences addressing all aspects of empirical assessment and evaluation in software engineering.

Topics of Interest

Papers related, but not limited to the following topics are invited:

- Justifying the use of empirical assessment and evaluation in an industrial environment
- Product and component evaluation
- Process and tool evaluation
- Quality assessment
- Software Process Improvement
- Experience reports from commercial or industrial environments
- Reports on concrete empirical studies
- Management strategies for a set of empirical studies
- Success stories

Target Groups

The workshop addresses practitioners and/or researchers who are interested in empirical software engineering, software process improvement, and quality management.

Practitioners are encouraged to submit papers on successful as well as unsuccessful projects. Researchers are asked to submit papers of innovative approaches/empirical studies on implementing new technologies, methodologies, or tools in industrial settings.

Workshop Organizers

Christian Bunse – Fraunhofer Institute Experimental Software Engineering (IESE), Sauerwiesen 6, D-67661 Kaiserslautern, Germany; email Christian.Bunse@iese.fraunhofer.de
Andreas Jedlitschka – Fraunhofer Institute Experimental Software Engineering (IESE), Sauerwiesen 6, D-67661 Kaiserslautern, Germany; email Andreas.Jedlitschka@iese.fraunhofer.de

Programme Committee

Christian Bunse – Fraunhofer IESE, Kaiserslautern, Germany
Andreas Jedlitschka – Fraunhofer IESE, Kaiserslautern, Germany
Elixabete Ostolaza – European Software Institute (ESI), Bilbao, Spain
Toni Sandelin – VTT, Oulu, Finland
Claes Wohlin – Blekinge Insitute of Technology (BTH), Ronneby, Sweden
Marcus Ciolkowski – Technical University Kaiserslautern, Germany

Table of Contents

An Empirical Study with Datawarehouse Design Methods[1]

Manuel Serrano, Coral Calero, Mario Piattini

Alarcos Research Group
E. S. Informática de Ciudad Real
University of Castilla – La Mancha
{Manuel.Serrano, Coral.Calero, Mario.Piattini}@uclm.es
http://alarcos.inf-cr.uclm.es/

Abstract. Datawarehouses have become the key trend in corporate computing, since they provide managers with the most accurate and relevant information to improve strategic decisions. Most of the datawarehouse design methods are based on multidimensional data models using star schemas. Although it is supposed that star designs are the most appropriate for datawarehouses, there are not studies that demonstrate this hypothesis. In this paper, we present an empirical study trying to assess if star schemas based diagrams are easier to understand and manage than those created by the traditional design of relational databases. As a conclusion of this study we can deduce that it seems that there is not difference in the understandability of the diagrams made with the two design methods.

1 Introduction

Nowadays, organizations can store vast amounts of data obtained at a relatively low cost, however these data fail to provide information [4]. Datawarehouses have appeared as a solution to this problem supporting decision making processes.

A datawarehouse is defined as a "collection of subject-oriented, integrated, non-volatile data that supports the management decision process" [7]. Datawarehouses have become the key trend in corporate computing in the last years, since they provide managers with the most accurate and relevant information to improve strategic decisions. Also the future for datawarehouse is promising, Jarke et al. [8] forecast a market of 12 Millions American dollars for the datawarehouse market for the next years.

Several methodologies for datawarehouse design have been proposed recently [1] [5] [7] [8] [9] [11], all of them based on the well known star schema design techniques, leaving aside the traditional design of relational databases. These authors state that star schema designs improve the understandability and use of

[1] This research is part of the CALDEA project (TIC 2000-0024-P4-02) supported by the Subdireccion General de Proyectos de Investigación, Ministerio de Educación y Cultura of Spain.

datawarehouses. However, common wisdom, intuition, speculation, and proofs of concepts are not reliable sources of credible knowledge, being necessary the experimentation [2]. As it is known, progress in any discipline involves building models that can be tested, through empirical study, to check whether the current understanding of the field is correct.

In this paper we present and empirical study we have researched in order to assess if star diagrams are easier to understand that those made using the traditional design of relational databases.

Next section present the goal of the experiment, section 3 shows the experiment and in section 4 the analysis of the results of the experiment is discussed. Last section reports the conclusions we can obtain of this paper.

2 Controlled Experiment

Although in the design, execution and analysis of the controlled experiment we have followed some suggestions provided by [14] [10] [12] and [3], for describing the experiment we use (with only minor changes) the format proposed by Wohlin et al. [14].

The goal of this empirical study is to determine if datawarehouse diagrams based on star schemas design are easier to understand than those based on the traditional design of relational databases, from the point of view of the user. Using the GQM paradigm [13] for organising the experiment, we can say that the goal of our experiment is "To analyse *two datawarehouse design methods,* in order to *evaluate* their *semantic comprehension* from the point of view of the *user* in the context of *university students*".

3 Experiment Planning

Our study is composed by a controlled experiment carried out by Ph. D. students of the High School of Computer Science of Ciudad Real (University of Castilla – La Mancha, UCLM) and two replications of the first experiment made by third course students of Computer Science of UCLM and Ph. D. students of the University of Pinar del Rio (Cuba). This study was carried out with students due to the difficulty of experimenting with professionals and taking into account that differences between professional and students are small and experimentation with students is feasible under certain circumstances [6], also the tasks the subjects had to perform in the experiment, were similar to those made in real environments.

All the subjects have knowledge in designing and using relational databases, also Ph. D. students have knowledge in designing and using datawarehouses.

The experiment is composed by three diagrams designed by the traditional design of relational databases and three semantically equivalent diagrams designed using star schema design. All The diagrams represented real problems are were simple enough to be easily understood. The subjects had to create several SQL sentences using these diagrams.

As the goal of the experiment was to determine if there was difference in the comprehension between the diagrams semantically equivalents designed with a different method, we checked if there was difference in the answer times to some questions.

Hypotheses. Our main hypothesis is:

H_{01}: There is not difference between the subjects using the two types of design (traditional relational and star schemas)

As our main goal is to determine if the source of differences is the design type, we formulate the next hypotheses in order to avoid that the domain of the schemas and its interaction with the design type affect the study with our principal hypothesis:

H_{02}: There is no difference between the subjects using the three types of domains with respect to time.

H_{03}: There is no difference between the subjects using the different treatments (interaction design type x domain) with respect to time.

Variables. The independent variables are the variables for which the effects should be evaluated. In our experiment these variables correspond with the domain of the diagrams (DOMAIN) and the design type used to create the diagrams (DESIGN_TYPE). The dependent variable is the time used to answer the question formulated to the subjects.

Design. Taking the hypotheses into account, we created a experimental design with three diagrams designed with the traditional relational method and three diagrams semantically equivalents designed with star schemas. The subjects had to build SQL sentences and write down the time (in seconds) used to answer the question.

Experimental Objects. The objects were, as we previously said, six diagrams. There was three pairs of semantically equivalent diagrams, each one designed by a method. The subjects were given the six diagrams in different order with an answer form enclosed where they had to write the SQL sentences and the time they used to build these sentences.

The experiment were made in a single session. Subjects were given an intensive explanation session, what kind of exercises may be developed, the material given, how response the questions made, and how much time they had to take every test. However, the subjects were not aware of the aspects we intended to study. Neither were they informed of the actual hypotheses stated.

Validity of results. As we know different threats to the validity of the results of an experiment exist. In this section we are going to discuss threats to construct, internal and external validity. First we will give the definition of each threat [14] and after we will study each of them for our experiment.

Construct validity.Construct validity is concerned with the relationship between theory and observation. We propose, as a reasonable measure of understandability the accuracy with which the subjects completed the tasks included in each test (so, as the number of correct answers given to the exercises proposed) on a fixed time. To assure construct validity it would be necessary to perform more experiments, varying the operations to be developed.

Internal validity. Internal validity is related to the assurance that the relationship observed between the treatment and the outcome is a causal relationship, and that it is not a result of a factor over which we have no control or have not measured. Regarding internal validity the following issues must be considered:

- **Differences among subjects.** Within-subject experiments reduce variability among subjects. In each experiment, all the subjects had, approximately, the same experience working with data models.
- **Knowledge of the universe of discourse.** The data models were general enough to be easily understood by each of the subjects. As a result the knowledge of the domain does not affect the internal validity.
- **Learning effects.** The subjects were given the tests in different order, to cancel out learning effects. Subjects were supervised and required to answer in the order in which the tests appeared.
- **Fatigue effects.** On average the experiment lasted for less than two hours, so fatigue was not very relevant. Also, the different order in the tests helped to cancel out these effects.
- **Persistence effects.** We must be sure that, when a set of subjects perform an experiment, the effects of previous similar experiments do not persist. In our case persistence effects are not present because the subjects had never performed a similar experiment.
- **Subject motivation.** Subjects were motivated because the exercises of the experiment was a part of the knowledge they must acquire in their formation.
- **Other factors**. Plagiarism and influence among subjects were controlled. Students were told that chatting was forbidden and lecturers were in the class during the experiment assuring this aspect.

External validity. External validity is concerned with generalization of the results. Regarding external validity the following issues must be considered:

- Materials and tasks used. We tried to use relational and multidimensional data models and operations representative of real cases in the experiments although more experiments with larger and more complex multidimensional data models are necessary.
- Subjects. Due to the difficulty of getting professionals to perform the experiment, the experiment was done by students. In general, more experiments with a larger number of subjects, students and professionals are necessary to obtain more conclusive results.

4 Analysis and Interpretation

To proceed with the analysis, we have first to preset a level of significance. Taking into account the characteristics of our experiment we have selected $\alpha = 0{,}1$ because raising the level of significance we can raise the power of the statistical test (probability of reject our hypotheses when are false).

Due to the goals and design of our experiment we must use an univariate ANOVA test of repeated measures. In tables 1, 2 and 3 the obtained results from the apply of the statistic to the experiment data are shown. Analysing the significance value (Sig. column of the tables) we can see that all the values are greater than $\alpha = 0{,}1$ and thus we cannot reject the hypotheses of there is not difference between the answer time of the experiment with respect to the design type (DESIGN_TYPE), the domain of the diagrams (DOMAIN) and the interaction of both factors (DESIGN_TYPE X DOMAIN)

As conclusion of the experiments we can deduce that it does not seem to be difference in the comprehension of the diagrams due to the design method used in the construction of the diagrams. This conclusion can be influenced by the diagram size of the experiment, that maybe are not big enough, and by the fact that the subjects have more experience working with relational databases so the results can be skewed. We must replicate the experiment using bigger and real diagrams in order to determine if there is any difference in the comprehension of the diagrams due to the design type choosen.

Table 1. ANOVA for Ph D. Students of UCLM

Source		SS	df	MS	F	Sig.	Power
DESIGN_TYPE	Hypothesis	10730,09	1	10730,09	0,01	0,95	0,10
	Error	1680826,46	1	1680826,46			
DOMAIN	Hypothesis	14558,91	2	7279,45	0,44	0,69	0,14
	Error	33043,27	2	16521,63			
DESIGN_TYPE X DOMAIN	Hypothesis	8045,24	2	4022,62	2,73	0,27	0,32
	Error	2945,24	2	1472,62			

Table 2. ANOVA for third course students of UCLM

Source		SS	df	MS	F	Sig.	Power
DESIGN_TYPE	Hypothesis	8557,55	1	8557,55	0,09	0,82	0,10
	Error	98157,55	1	98157,55			
DOMAIN	Hypothesis	4643,79	2	2321,90	0,19	0,84	0,12
	Error	24705,57	2	12352,78			
DESIGN_TYPE X DOMAIN	Hypothesis	9252,44	2	4626,22	0,33	0,75	0,13
	Error	27950,91	2	13975,45			

Table 3. ANOVA for Ph D. Students of University of Pinar del Rio (Cuba)

Source		SS	df	MS	F	Sig.	Power
DESIGN_TYPE	Hypothesis	101117,21	1	101117,21	0,02	0,90	0,10
	Error	4183548,98	1	4183548,98			
DOMAIN	Hypothesis	4633505,47	2	2316752,73	4,29	0,19	0,41
	Error	1080115,07	2	540057,53			
DESIGN_TYPE X DOMAIN	Hypothesis	1278292,95	2	639146,47	2,14	0,32	0,27
	Error	597794,23	2	298897,12			

5 Conclusions

Datawarehouses have become the key trend in corporate computing, since they provide managers with the most accurate and relevant information to improve strategic decisions.

Several datawarehouse design methods have been proposed based on star schemas as it is supposed that this type of schemas raise the comprehension and efficiency of datawarehouses. Although this idea is widely accepted, there is not empirical studies supporting it. In this paper we have presented the experiment we have carried out to corroborate this assertion.

As a conclusion of our experiment, it does not seem to be difference between the two design type with respect to comprehension and we cannot assure that star schemas are easier to understand. This conclusion can be affected by the experience of the subjects, which is wider in relational databases, and by the diagrams size. In order to have more conclusive results we must replicate the experiment with bigger diagrams in order to observe if star schemas have a better comprehension. Also, it would be desirable to replicate the experiment with real data and diagrams and with professionals.

References

1. Adamson, C. and Venerable, M.: Data Warehouse Design Solutions. John Wiley and Sons. (1998)
2. Basili, V. and Lanubile, F.: Building Knowledge through Families of Experiments. IEEE Transactions on Software Engineering. 25(4), (1999) 456-473.
3. Briand, L., Arisholm, S., Counsell, F., Houdek, F. and Thévenod-Fosse, P.: Empirical Studies of Object-Oriented Artefacts, Methods and Processes: State of the Art and Future Directions, Empirical Software Engineering, 4(4). (1999) 387-404.
4. Gardner, S.R.: Building the data warehouse, Communications of the ACM, 41(9). (1998) 52-60.
5. Hammergren, T.: Data Warehousing Building the Corporate Knowledge Base International Thomson Computer Press, Milford . (1996)
6. Hörst, M., Regnell, B. and Wohlin, C.: Using Students as Subjects – A Comparative Study of Students & Profesionals in Lead-Time Impact Assessment. 4th Conference on Empirical Assessment & Evaluation in Software Engineering, EASE, Keele University, UK. (2000)
7. Inmon, W. H.: Building the Data Warehouse. John Wiley and Sons, 2nd edn. (1997)
8. Jarke, M., Lenzerini, M., Vassilou, Y. and Vassiliadis, P.: Fundamentals of Data Warehouses. Springer. (2000)
9. Kelly, S.: Data Warehousing in Action. John Wiley & Sons. (1997)
10. Kitchenham, B., Pfleeger, S., Hoaglin, D., EL Emam, K. and Rosenberg, J.: Preliminary Guidelines for Empirical Research in Software Engineering. IEEE Transactions on Software Engineering, 28(8). (2002) 721-734.
11. Kimball, R., Reeves, L., Ross, M. and Thornthwaite, W.: The Data Warehouse Lifecycle Toolkit. John Wiley and Sons. (1998)
12. Perry, D., Porter, A. and Votta, L.: Empirical Studies of Software Engineering: A roadmap, Future of Software Engineering, ACM. Ed. Anthony Finkelstein, (2000) 345-355.
13. Van Solingen, R. and Berhghout, E.: The Goal/Question/Metric Method, McGraw-Hill. (1999)
14. Wohlin, C., Runeson, P., Hörst, M., Ohlson, M., Regnell, B. and Wesslén, A.: Experimentation in Software Engineering: An Introduction, Kluwer Academic Publishers. (2000)