Jyrki Kontio
Reidar Conradi (Eds.)

LNCS 2349

Software Quality – ECSQ 2002

Quality Connection – 7th European Conference on Software Quality
Helsinki, Finland, June 2002
Proceedings

Springer

Jyrki Kontio   Reidar Conradi (Eds.)

Software Quality – ECSQ 2002

Quality Connection – 7th European Conference on Software Quality
Helsinki, Finland, June 9-13, 2002
Proceedings

Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Jyrki Kontio
Nokia Research Center, Nokia Group
P. O. Box 407, 00045 Nokia Group, Finland
E-mail: jyrki.kontio@nokia.com

Reidar Conradi
Norwegian University of Science and Technology (NTNU)
Department of Computer and Information Science (IDI)
Gloeshaugen, 7491 Trondheim, Norway
E-mail: conradi@idi.ntnu.no

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Software quality : quality connection - 7th European conference on software quality; proceedings / ECSQ 2002, Helsinki, Finland, June 9 - 13, 2002. Jyrki Kontio ; Reidar Conradi (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Tokyo : Springer, 2002
(Lecture notes in computer science ; Vol. 2349)
ISBN 3-540-43749-5

CR Subject Classification (1998): D.2, K.6.3, K.4.3, K.6, J.1, H.5.3

ISSN 0302-9743
ISBN 3-540-43749-5 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Steingräber Satztechnik GmbH, Heidelberg
Printed on acid-free paper SPIN 10870017 06/3142 5 4 3 2 1 0

Preface

Software professionals and companies live in a new world today. Increasingly complex systems need to be built faster and cheaper. While many of the established approaches in software quality are still valid, the software quality community is going through a paradigm shift that requires a re-assessment of our current method and tool portfolio, as well as creating new and more effective solutions.

We have selected two themes for this conference to highlight this paradigm shift. Our first theme, "*production of attractive and reliable software at Internet speed*" sums up the dilemma many software organisations face. In order to be competitive, software should contain advanced features and run reliably – yet it should be developed quickly and cost effectively for the right market window. Finding the right balance between these objectives is a critical question that will determine business success in the years to come.

Our second theme, "*production of software with a dynamic partnership network*" highlights the current trend of using partnerships and subcontractors as integral players in the software development process. Partnerships sometimes need to be created quickly to respond to a market opportunity, yet the costs and speed of cooperation must be competitive. Different companies have different processes, quality tools and cultures, yet they should cooperate seamlessly for the best result.

The 7^{th} European Conference on Software Quality – Quality Connection – addresses these challenges as the papers in these proceedings show. We received a total of 78 technical and experience-based papers and two to three referees reviewed each paper. The papers were selected based on how well they satisfied six evaluation criteria: relevance to the conference themes; novelty of contribution; industrial significance; empirical validation; positioning with other work; and writing style and correctness. After a rigorous review process, 31 papers were accepted and are printed in these conference proceedings. These papers provide a solid technical foundation for the conference and offer novel contributions to the community. In addition, the proceedings include keynote and invited papers that provide timely perspectives in this transition.

In addition to the material included in the proceedings, the programme committee selected noteworthy contributions to be presented at the Quality Forum during the conference. These contributions are published by the conference organiser as *Quality Connection - 7^{th} European Conference on Software Quality 2002 - Conference Notes* (ISBN 952-5136-24-8). These contributions present interesting new issues and ideas, as well as practical experiences, on key approaches in achieving software quality.

One of our targets was to establish this conference as the main European forum for providing and sharing the latest and most reliable information on software quality. Thus, we aimed at improving the scientific quality of the accepted papers, while maintaining the practical orientation of the conference. Having completed the review process, we feel that this objective has been reached.

The greatest thanks belong to the authors who have conducted the research and are willing to share their results and insights. In addition, the programme committee did a very thorough and objective job in reviewing and discussing each paper and by providing detailed feedback to the authors.

This conference series is supervised by the European Organization for Quality and its Software Group, currently chaired by Mr. Finn Svendsen. We are grateful to them for providing this forum for these contributions. The General Chair of the conference, professor H. Dieter Rombach, has also been an excellent source of advice and guidance in making this conference happen. We would also express our gratitude to Center for Excellence Finland, and especially to the General Secretary of the conference, Ms. Maija Uusisuo, who has done an excellent job in hosting and organising the conference.

These proceedings are published as the conference takes place. At the same time, the software quality community is entering an era of new challenges. The selected presentations give indications of what the new software quality paradigm will look like. We would like to welcome you and your partners to peruse and apply the knowledge and insight contained in these papers to develop attractive and reliable software even faster.

Helsinki and Trondheim, March 15, 2002

Dr. Jyrki Kontio & Prof. Reidar Conradi
ECSQ2002 Programme Committee Co-Chairs

Organisation

The 7th European Conference on Software Quality (ECSQ2002) is organised by the Center for Excellence Finland in co-operation with the European Organization for Quality – Software Group (EOQ-SG).

General Chair

Professor H. Dieter Rombach, Fraunhofer IESE, Germany

Programme Committee Co-Chairs

Dr. Jyrki Kontio, Nokia Reseach Center, Finland
Professor Reidar Conradi, Norwegian University of Science and Technology, Trondheim, Norway

General Secretary of the Conference

Maija Uusisuo, Center for Excellence Finland, Finland

EOQ-SG Executive Committee

President:	Finn N. Svendsen, Grundfos A/S, Denmark
Member:	Walter Wintersteiger, Management and Informatik, Austria
Member:	Francois de Nazelle, QUAL-AS, France
Member:	Karol Frühauf, Infogem AG, Switzerland
Member:	Mika Heikinheimo, Flander Oy, Finland

Organising Committee

Chair:	Jyrki Kontio, Nokia Research Center, Finland
Coordinator:	Maija Uusisuo, Center for Excellence Finland, Finland
Member:	Casper Lassenius, Software Business and Engineering Institute (SoberIT), Helsinki Univ. of Tech., Finland
Member:	Petri Lehtipuu, Center for Excellence Finland, Finland
Member:	Markku Oivo, University of Oulu, Finland

Programme Committee

Member:	Juha-Markus Aalto, Nokia Mobile Phones, Finland
Member:	Adriana Bicego, Etnoteam, Italy
Member:	Lionel C. Briand, Carleton University, Canada
Member:	Chang Wen Kui, Tung Hai University, Taiwan
Member:	Ton Dekkers, IQUIP Informatica B.V., The Netherlands
Member:	Alec Dorling, QAI Europe, U.K.
Member:	Khaled El Emam, National Research Council, Canada
Member:	Christer Fernström, Xerox Labs, France
Member:	Norbert Fuchs, Alcatel Austria, Austria
Member:	Alfonso Fuggetta, Politecnico di Milano, Italy
Member:	Paul Gemoets, Oracle, Belgium
Member:	Januzs Górski, Technical University of Gdańsk, Poland
Member:	Ilkka Haikala, Tampere University of Technology, Finland
Member:	Shlomo Harlev, Tamam, Israel Aircraft Industries, Israel
Member:	Bernd Hindel, Method Park Software AG i.G., Germany
Member:	Yoshinori Iizuka, University of Tokyo, Japan
Member:	Ross Jeffery, University of New South Wales, Australia
Member:	Karlheinz Kautz, Copenhagen Business School, Danmark
Member:	Günter Koch, Austrian Research Centers, Austria
Member:	Pasi Kuvaja, University of Oulu, Finland
Member:	Kari Känsälä, Nokia Research Center, Finland
Member:	Dieter Landes, Coburg Univ. of Applied Sciences, Germany
Member:	Tim Lister, Atlantic Systems Guild Inc., U.S.A.
Member:	Patricia A. McQuaid, California Polytechnic State Univ., U.S.A
Member:	Winifred Menezes, Q-Labs Inc., U.S.A.
Member:	Sandro Morasca, Politecnico di Milano, Italy
Member:	Tony Moynihan, Dublin City University, Ireland
Member:	Paolo Nesi, University of Florence, Italy
Member:	Risto Nevalainen, STTF, Finland
Member:	Folke Nilsson, Quality IT, Sweden
Member:	Toshiro Ohno, Tsukuba International University, Japan
Member:	Elixabete Ostolaza, ESI / Strategic SPI, Spain
Member:	Shari Lawrence Pfleeger, University of Maryland, U.S.A.
Member:	Gopal Raghavan, Nokia Research Center, U.S.A.
Member:	Terry Rout, Griffith University, Australia
Member:	Kevin Ryan, University of Limerick, Ireland
Member:	Dag Sjøberg, University of Oslo, Norway
Member:	Torbjørn Skramstad, NTNU, Trondheim, Norway
Member:	Colin Tully, Colin Tully Associates, U.K.
Member:	Otto Vinter, Delta, Denmark
Member:	Lawrence Votta, Motorola Inc., U.S.A.
Member:	Claes Wohlin, Blekinge Institute of Technology, Sweden

Main Sponsor

Nokia Oyj, Finland

Other Sponsors

American Society for Quality
Finnair Oyj
Helsinki University of Technology
QPR Software Oyj Plc.
SecGo Group Oy
Siemens Oy
Solid Information Technology Oy
Sonera Oyj
Stonesoft Oyj
TietoEnator Oyj

Table of Contents

An Empirical Study with Metrics for Object-Relational Databases

Coral Calero[1], Houari Sahraoui[2], and Mario Piattini[1]

[1] Departamento de Informática University of Castilla-La Mancha
Paseo de la Universidad, 4
13071, Ciudad Real (Spain)
{Coral.Calero,Mario.Piattin}@uclm.es
[2] DIRO, Université de Montréal, CP 6128 succ Centre-Ville,
Montréal QC H3C 3J7, Canada
sahraouh@iro.umontreal.ca

Abstract. Object-relational databases are supposed to be the substitutes of relational ones because they are a good mixture between the relational model and object-oriented principles. In this paper we present the empirical work we have developed with four metrics for object-relational databases (Percentage of Complex Columns (PCC), Number of Shared Classes (NSC), Number of Involved Classes (NIC) and Table Size (TS)) defined at different granularity levels (attribute, class, table and schema). The empirical work presented is the validation made with the aim of proving the usefulness of the four metrics in estimating the complexity of an object-relational schema . This study can be considered to be a replica of another one we made with the same purpose but with two main differences: the dependent variable and the way we analyze the results. The results obtained from the empirical work seem to prove the usefulness of the TS metric in estimating the complexity of an object-relational schema however, conclusions about the other metrics are difficult to extract.

1 Introduction

Relational databases are the most important ones in the database world. This success can be explained because they are not too difficult to understand and also because there is a standard (SQL) for them. However, relational databases are not capable of working with the new kind of data (as complex data, image, sound, GIS, CAD) modern applications require.

In response to these new necessities, object oriented databases were created as the successors of the relational ones, but they have not been successful. Perhaps because, unlike the relational model, OO databases have not had a widely accepted standard until the appearance of the ODMG, but also because it is very difficult to convince 85% of the market of this revolutionary change: another technology, another way of thinking for their workers, another kind of data types, etc.

In order to allow the possibility of working with new data and applications without a revolutionary change in the market, the relational industry has reacted and has created object-relational databases that can be considered to be an evolutionary

J. Kontio and R. Conradi (Eds.): ECSQ 2002, LNCS 2349, pp. 298–309, 2002.

solution. The idea is to allow the adaptation of the relational market by supporting some object orientation capabilities. So, these databases have all the elements of the relational model (relations connected by referential integrity relationships) but with the particularity that the columns of a relation can be defined over a class. In fact, some studies predict that object-relational databases will substitute the relational ones by the year 2003 ([4]). Perhaps the data is not so accurate, but the change seems to be possible.

Taking into account the brilliant future predicted for the object-relational databases in the database market and to ensure that it becomes a reality, it is essential to have quality designs. One widely accepted mechanism for assuring the quality of a software product in general and of object-relational database designs in particular, is the use of metrics specifically designed for this kind of databases.

Metrics must be defined for capturing a specific attribute of a product (in our case object-relational databases). We think one of the most important internal factors to be captured is complexity. If we have a set of metrics for capturing complexity, we are also capturing understandability ([5]) and hence maintainability, one important dimension in software product quality ([3]).

In this paper we present the empirical work we have developed in order to know if the metrics defined for object-relational databases are good mechanisms for measuring their complexity and, hence, for assuring the quality of the schema design of this kind of databases. The controlled experiment presented is a replica of a previous one. However, there are two main differences between the study presented here and the previous one: the dependent variable and the way we analyze the results. We decided to change these two factors to see if the results obtained on the first study were independent of these factors and hence could be generalized.

A deeper explanation about object-relational databases and the metrics we have designed for them will be presented in the next section. Section Three will present the problem statement and the experiment planning will appear in Section Four. Analysis and interpretation will be discussed in Section Five and conclusions and future work will appear on the last section.

2 Metrics for Object-Relational Databases

As we have said in the previous section, object-relational databases can be considered a mixture of relational databases with object oriented characteristics.

So, an object-relational database schema is composed by a number of tables related by referential integrity, which have columns that can be defined over simple or complex (user-defined) data types. Simple data types may be one of the classic data types as integer, number or character (correspond to the relational columns). A complex or user-defined data type (also called classes), can be related with other data types by generalization associations.

In Table 1 we present an example of an object-relational database schema whose definition is in SQL:1999. The example defines two tables (agency and houses). Table agency has 2 simple columns (idAgency and name) and 1 complex column (situation) and table houses has 7 simple columns (idHouse, idAgency, price, rooms, size, description and photo) and 1 complex column (situation). We can also see that in table houses we have two columns defined as LOB type, this data type is used to store multimedia data (even GB).

Table 1. Example of an object-relational schema

create **type** address as(street varchar(15), city varchar(10), county varchar(10), zip varchar(8));	create table **agency**(idAgency integer, name varchar(10), situation address primary key (idAgency)); create table **houses**(idHouse integer, idAgency integer, price double, rooms integer, size float(4), situation address, description clob(100K), photo blob(100K) primary key idHouse; foreign key (idAgency) references agency);

Considering all the characteristics explained, we present the next metrics (the detailed explanation of these metrics can be found in [1]:

- **TS** (Table Size). This metric is calculated by adding the size of the simple columns (in our case we consider that a simple column has a size equal to one) of the table (TSSC) plus the size of the complex columns of the table (TSCC). The size of a complex column is defined as the size of the class hierarchy above which the column is defined divided by the number of complex columns that are defined over this hierarchy (SHC/NHC). The size of the hierarchy is defined as the sum of the classes of the generalization hierarchy. The size of a class is calculated as the sum of its attributes plus the size of its methods. It is necessary to take into account that a class can also have simple attributes (with a size equal to one) and complex attributes (attributes related to other classes by an aggregation relationship and its size are calculated as the size of a class). It will also be necessary to divide the size of a class by the number of classes which are related to the class with an aggregation relationship.
- **NIC** (Number of Involved Classes). Number of classes needed for defining all the columns of a table.
- **NSC** (Number of Shared Classes). Number of classes that are used by more than one table.
- **PCC** (Percentage of Complex Columns). Number of the columns of a table that are complex related with the total number of columns of the table.

Table 2. Relationships among the characteristics of the object-relational database levels and the metrics

	TABLE	CLASS	SCHEMA	ATTRIBUTE
PCC	X			
NIC	X	X		
NSC	X	X	X	
TS	X	X	X	X

Table 3. Relationship between our metrics and coupling and size

	SIZE	COUPLING
TS	X	x
NIC	X	x
NSC		x
PCC	X	

Table 4. Metrics values for the example presented in Table 1

	TS	NIC	NSC	PCC
Agency	4	1	1	1
Houses	9	1	1	1

We have selected these metrics because they cover all the possible levels of an object-relational database (attribute, class, table and schema). In Table 2, we reflect the different levels involved on the definition of each of the metrics. For calculating the PCC metric, it is necessary to look at the information of the table. For the NIC metric, it is necessary to look at the information of a table and the classes used by this table. For the NSC metric, we work at a table level. We also need to consider all the classes involved in the definition of this table and the rest of the schema tables that also use these classes. Finally, the calculation of the TS metric requires the consideration of a table: its simple attributes, the classes the table uses for the definition of its complex columns and whether or not these classes are used by other tables

We can also relate the metrics with two important factors related to complexity, size and coupling (see Table 3).

Applying these metrics to the example presented in Table 1, we obtain the values shown in Table 4.

The next step is to try to prove if our metrics can be considered complexity metrics.

3 Problem Statement

As we indicated previously, the main goal of this paper is to prove that the presented metrics can be used as complexity metrics.

The experimentation presented in this paper can be considered a replica of another controlled experiment carried out with the same objective. However, with the aim of being able to generalize the results, some changes were included in this second controlled experiment with respect to the first one:

- the dependent variable is now measured as the time needed by the subjects to answer a set of questions about the object-relational schemas, unlike the dependent variable in the previous study that was measured as the number of correct answers obtained in a fixed time.
- the way we analyze the results. In this study we use traditional statistics and in the previous one we used advanced techniques (RoC and C4.5).

In the next section, we will briefly present the results obtained from the original experiment. At the end of this paper we will discuss the conclusions we can draw from the experimentation process as a whole.

3.1 Previous Work

In [2] the previous experimental work made with these and other metrics is presented. In that case we also sought to prove the utility of our metrics as complexity indicators. Our hypothesis was that *complexity is itself impacted by the size and the coupling between the elements of an object-relational database schema (tables and classes).* We worked with six object-relational schemas and the dependent variable was measured as the time needed by the subjects to make some operations with the schemas. To analyze the usefulness of the metrics, we used two machine learning (ML) techniques: C4.5 ([6]), a Top Down Induction of Decision Trees algorithm, and RoC ([7]), a robust Bayesian classifier. As for subjects, we worked with people from both Canada and Spain. As a conclusion, both techniques indicated that the table size metric (TS) is a good indicator for the complexity of a table. The rest of the metrics do not seem to have a real impact on the complexity of a table.

3.1 Current Work

As previously indicated, the main goal of this paper is to explain the empirical work we developed in order to know if the results obtained with the previous study about our metrics were repeated and, hence, could be generalized. The hypothesis did not vary in the replication of the experiment. However, we did change the way in which the dependent variable was measured (we wanted to capture complexity in another way in order to confirm if the previous results were independent of the way it was captured) and the way we analyzed the results (to know if the results are independent of the technique used). By carrying out this kind of replication in which the same hypothesis is studied but some details of the experiment are changed, our aim is to make the results of the experiment more reliable.

The goal definition of our experiment can be summarized as:

- *To analyze* the metrics for object-relational databases
- *for the purpose* of evaluating if they can be used as a useful mechanism

- *with respect to* the object-relational databases complexity
- from the designer *point of view*
- *in the context of* students of a database course with knowledge of object-relational databases, relational databases and object-oriented programming.

4 Experiment Planning

Nine students from the University of Montreal (Canada) and six students from the University of Castilla-La Mancha (Spain) participated in the experiment. All of them were finishing a course on advanced databases and knew all the concepts related with object-relational databases. They were also familiar with relational databases and object-oriented programming.

We tried to define the tests involved in the experiment in such a way that they would be representative of real cases. The use of a simple experiment allows other experimenters to replicate it because the number of examples is not too large. All the experimental materials (laboratory package) of this experiment can be found in http://alarcos.inf-cr.uclm.es.

4.1 Hypotheses Formulation

Our hypothesis is that complexity is itself impacted by the size and the coupling between the elements of an object-relational database schema (tables and classes).

4.2 Variables in the Study

Independent variables. In our experiment these variables correspond with the metrics under study: TS, NIC, NSC and PCC. Each of these metrics (factors) can take different values (levels) .

Table 5. Descriptive statistics of each metric

	Minimum	Maximum	Average	Stan.Dev.
NIC	,00	7,00	2,6410	1,9532
NSC	,00	5,00	1,0513	1,1459
PCC	,00	,66	,2982	,1736
TS	2,00	25,00	9,9474	6,4840

Dependent variables. As we have previously stated, the complexity of the tests was measured as the time each subject used to perform all the tasks of each one of the 21 object-relational database schemas. Taking into account the experience of the subjects in relational and object-relational database design and object-oriented programming, and the fact that the tasks were not too difficult, we thought that all of them would give the right answers. We were proved right as all the subjects (except one whose results were discarded) answered the tests correctly. Therefore, we were able to work with the results of fourteen subjects.

Regarding the time, it is necessary to point out that it included the time to analyze the schema and the time to answer the questions about it.

4.3 Analysis

Taking the hypotheses into account, the experiments must consider the relationship between the metrics and the time. So, we decided to made a correlation among these factors. For the correlation test we used two different correlation tests: the Pearson correlation, and the Spearman's non-parametric correlation to identify potential relationships between the time needed by the subjects and the metrics defined.

4.4 Data Used in the Study

Twenty-one object-relational database schemas were used for performing the experiment. The documentation was approximately twenty-four pages long and included, in addition to the database schemas, a general description of the metrics and a requirements document.

For each design, four operations were performed. The subjects had to calculate the values of four structural metrics for each table of each one of the 21 schemas. With the results they had to complete a questionnaire. Part of this questionnaire is shown in Figure 1. In this questionnaire, the result of each metric in each table of each schema had to be recorded along with the time needed to calculate these values.

Table Name	**Schema Name**	**Begin (Hour :Minute)**	**O1**	**O2**	**O3**	**O4**	**End (Hour :Minute)**
Laboratories	Laboratory_Reports						
Books	Library						
Activities	Sport_2						
Houses	Housing						
Categories	Directory						
Cities	Regions						

Figure 1. Question/Answer Paper

Before the subjects took the test, the experiment was conducted with a small group of people in order to improve it, if necessary and ensure that both the experiment and the documentation were well designed. No changes were necessary.

Tests were given to the subjects and they had a week to answer and return the results. Before giving them the material, the complete experiment was explained to the subjects, what kind of exercises they had to do, the material they would be given, what kind of answers they had to provide and how they had to record the time spent solving the problem. The metrics with an example was also explained.

Before starting each test, the subjects had to record the start time, and when they had completed the test, they had to record the end time. As the annotation of the time was the responsibility of the subjects and in order to avoid fatigue effects in the subjects (because the experiment had a lot of schemas) we decided to give the subjects one week to carry out the experiment.

In order to avoid learning effects we required subjects to calculate the metrics of the tables in a determined order, in this way they never worked with two tables of the same schema consecutively. As all the metrics were correctly answered, to obtain the results of the experiment we used the number of minutes needed by each subject. We may indicate here that one test were discarded because we detected very strange results on the time annotated by a subject and he told us that he had not respected the order given in the question/answer paper.

4.5 Validity of the Results

We are aware that different threats to the validity of the results of an experiment exist. In this section we are going to discuss threats to internal and external validity and to construct.

Construct Validity

Construct validity is concerned with the relationship between theory and observation ([8]). We propose, as a reasonable measure of complexity, the time for calculating the values of the metrics taking into account that this time comprises not only the time needed to calculate the metrics but also the time needed by the subject to analyze and to understand the object-relational database schema.

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 which we have not measured ([8]). Regarding internal validity the following issues must be considered:

- Differences among subjects. Although we worked with subjects of different countries, we think there were no significant differences among them because they had approximately the same experience on relational and object-relational databases and on object-oriented programming. In fact all of them (except for one) calculated the metrics correctly.
- Differences among schemas. We tried to define the tests involved in the experiment in such a way that they were representative of real cases. However, the domain of the schemas was different and this could influence the results obtained in some way.
- Precision in the time values. The subjects were responsible for recording the start and finish times of each test. We think this method is more effective than having

a supervisor who records the time of each subject. However, we are aware that the subject could introduce some imprecision.

- Learning effects. We required subjects to calculate the metrics of the tables in a determined order, in this way they never worked with two tables of the same schema consecutively. In this way, we think we prevent learning effects.
- Fatigue effects. In order to avoid fatigue effects in the subjects (because the experiment had a lot of schemas) we decided to give the subjects one week to carry out the experiment. We think this could be dangerous (from the plagiarism perspective, for example) if the subjects were be forced to made the experiment but, as we used volunteers, we think this is not a problem.
- Persistence effects. Not present because the subjects had never performed a similar experiment.
- Subject motivation. The experiment was made by volunteers so we think they were sufficiently motivated.
- Other factors. Subjects were informed they should not share answers with other subjects. Nevertheless, the subjects were not watched or controlled during the experiment so we cannot assure that the mentioned effects do not appear, although we think it is improbable.

External Validity

External validity is concerned with generalization of the results ([8]). Regarding external validity the following issues must be considered:

- Materials and tasks used. We tried to use schemas and operations representative of real cases in the experiments although more experiments with larger and more complex schemas are necessary.
- Subjects. In general, more experiments with a larger number of subjects, students and professionals, and with different kinds of operations are necessary to obtain more conclusive results about the metrics.
- We tried to increase external validity by performing the replica with some changes (in the dependent variable and in the technique used to analyze the results), so the results can be more general.

4 Analysis and Interpretation

As we have explained previously, we had fourteen data sets for each table of each schema, one of which corresponded to the answer of each subject of the experiment. For applying the tests of correlation we decided to work, for each table, with the time average for calculating each metric.

First we had to check the normality of the data obtained after the calculation of the average. If the data were normal, the best option in our case was to use parametric tests because they are more efficient than non-parametric tests and because they require fewer data points (this point is very important in our case because we only have eleven data points), and therefore smaller experiments than do non-parametric tests.

Among the tests we can apply to ascertain if a distribution is normal, we have the Shapiro-Wilk test and the Kolmogorov-Smirnov test. Both tests were applied to our data and we found that they were normal. In Figure 2 the normality test graph is shown. As we can see, the points are near the line, so the data can be considered normal and the parametric tests can be applied.

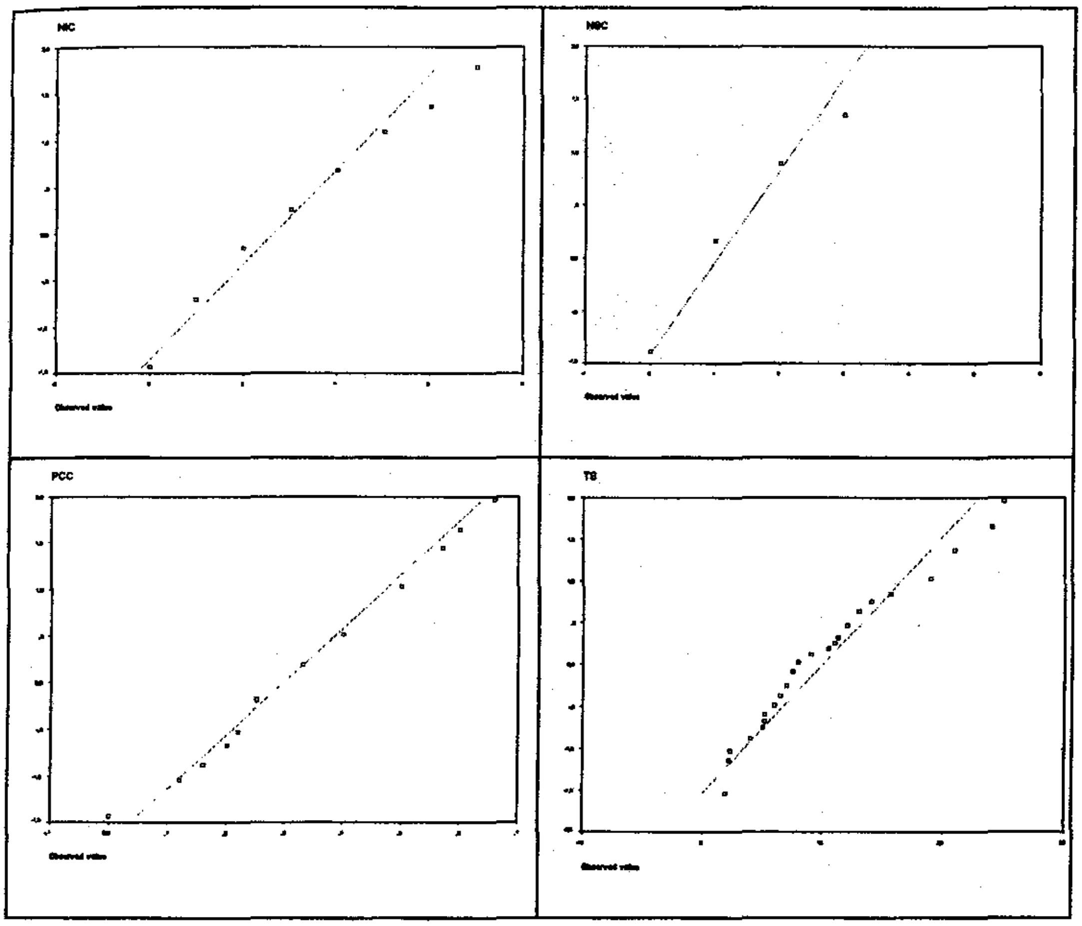

Figure 2. Normality tests for the data of the four cases

To proceed to the analysis, we first have to preset a level of significance. Several factors have to be considered when setting α because we can commit a Type I error (probability of incorrectly rejecting the null hypothesis). We decided to select α=0.05 which means a 95% level of confidence.

Although we knew our data were normal and the parametric tests could be applied, we decided to apply both, parametric and non-parametric tests in order to check if we would obtain the same results. So, we applied both Pearson and Spearman correlation tests to our data. We present the results obtained from the Pearson (Table 6) and Spearman (Table 7) correlations below.

From the Pearson correlation we can conclude that the number of involved classes, the percentage of complex columns and the table size seem to be significant metrics as complexity indicators. The number of shared classes does not seem to be significant. This can be explained because the NSC is easily calculable (it is only the number of classes that are shared by more than one table) and does not add complexity to the schema.

Table 6. Pearson correlation results

		NIC	NSC	PCC	TS	TIME AVG
Pearson Correlation	**NIC**	1,000	,267	,826	,829	,607
	NSC	,267	1,000	,130	-,172	,164
	PCC	,826	,130	1,000	,694	,575
	TS	,829	-,172	,694	1,000	,513
	TIME AVG	,607	,164	,575	,513	1,000

Table 7. Spearman correlation results

		NIC	NSC	PCC	TS	TIME AVG
Rho de Spearman	**NIC**	1,000	,299	,805	,807	,733
	NSC	,299	1,000	,188	-,082	,331
	PCC	,805	,188	1,000	,605	,650
	TS	,807	-,082	,605	1,000	,648
	TIME AVG	,733	,331	,650	,648	1,000

From the Spearman correlation we can conclude the same as from the Pearson's one, the number of involved classes, the percentage of complex columns and the table size seem to be significant metrics as complexity indicators, unlike the number of shared classes which does not seem to be significant.

From both tests we can conclude that the NIC, PCC and TS seem to be good indicators of the complexity of an object-relational database schema. Also, we can see that there is a correlation among TS and the rest of the metrics (except NSC), so perhaps it would be enough to use the TS metric.

From the experiments as a whole, the previous one and the one presented in this paper, we can conclude that TS seems to be a good indicator of the object-relational databases complexity and it is not possible to obtain a conclusive result for the other metrics.

It would be necessary to carry out more experiments with the set of metrics to determine if the PCC, the NIC and the NSC metrics are useful complexity mechanisms of the object-relational database schemas or not. The new experiments can also help us to know if the statistic technique applied also has impact on the results obtained.

It would be also interesting to carry out new experiments with the participation of professionals instead of students, and make not only controlled experiments but also case studies with real data. In this way, the experimental work would be complete and the results would be more conclusive.

5 Conclusions and Future Work

Object-relational database are considered to be the substitutes of the relational ones in the database market. For assuring that the designs obtained have quality, it is fundamental to have metrics for the designs of these databases. However, it is more important to be sure that the metrics proposed are useful for the aim they seek to achieve. Performing empirical validation with the metrics is fundamental in order to demonstrate their practical utility. In this paper we have presented part of the empirical work we are developing with a set of metrics for object-relational database schema designs (TS, NIC, NSC and PCC).

The controlled experiment presented here can be considered as the replica of a previous one. The objective of this empirical validation is to know if the metrics can be considered indicators of the complexity of the database design.

As a result of the experiment we have concluded that the TS, NIC and PCC metrics seem to be good complexity indicators. Also, we have indicated that there is a correlation among TS and the rest of the metrics (except NSC), so perhaps, it would be enough to use the TS metric.

As a result from the whole empirical work, we can only obtain conclusive results for the TS metric.

Looking at these results it is necessary to follow up to learn more about the metrics. Perhaps, the experimental process would be improved by changing the subjects (working with practitioners), and developing real cases (working with real data). However, we think this study was necessary in order to have preliminary results about the metrics.

References

1. Calero, C., Piattini, M., Ruiz, F. and Polo, M. Validation of metrics for Object-Relational Databases, *International Workshop on Quantitative Approaches in Object-Oriented Software Engineering (ECOOP99)*, (Lisbon ,Portugal. June 1999), 14-18
2. Calero. C., Sahraoui, H, Piattini, M., Lounis, H. Estimating Object-Relational Database Understandability Using Structural Metrics, *12th International Conference and Workshop on Database and Expert Systems Applications, (DEXA 2001)*, (Munich, Germany, 3-7 September 2001).
3. ISO (1999). ISO 9126. Software Product Evaluation-Quality Characteristics and Guidelines for their Use. *ISO/IEC Standard 9126*. Geneva.
4. Leavitt, N. (2000). Whatever happened to Object-Oriented Databases?. Industry Trends, *IEEE Computer Society*. pp. 16-19. August
5. Li, H.F. and Cheng, W.K. An empirical study of software metrics. *IEEE Trans. on Software Engineering*, (1987), 13 (6): 679-708.
6. Quinlan, J.R., *C4.5: Programs for Machine Learning, (1993),* Morgan Kaufmann Publishers.
7. Ramoni, M. and Sebastiani, P. Bayesian methods for intelligent data analysis. In M. Berthold and D.J. Hand, editors, *An Introduction to Intelligent Data Analysis*, (New York, 1999). Springer.
8. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B. and Wesslén, A. *(2000). Experimentation in Software Engineering. An Introduction.* Ed. Kluwer Academic Publishers