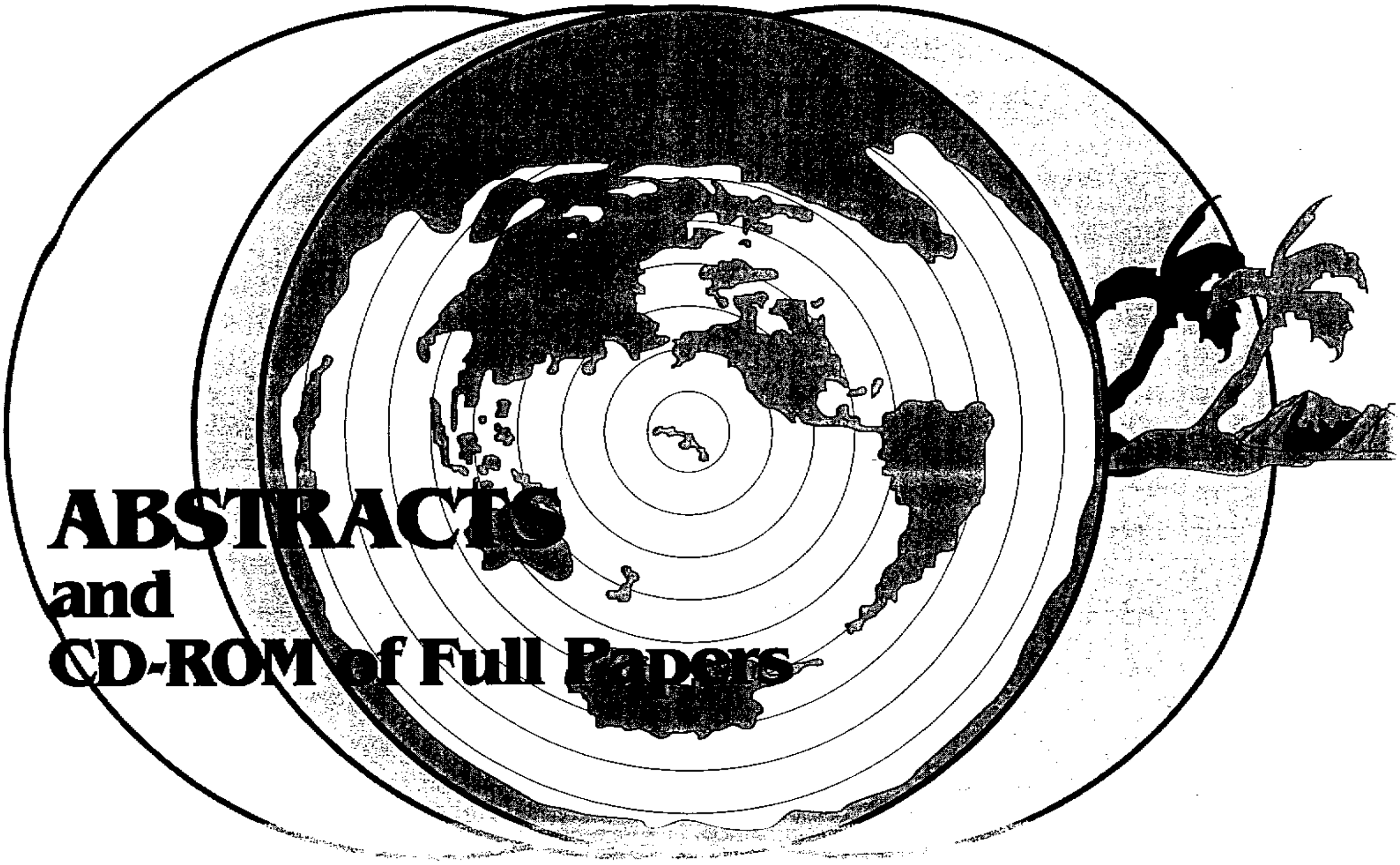
Proceedings of the Thirty-Sixth Annual Hawaii International Conference on System Sciences

2003



ABSTRACTS and CD-ROM of Full Papers

University of Hawaii

1907

University of Hawaii at Manoa
College of Business Administration

Published by the IEEE Computer Society
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314

IEEE Computer Society Order Number PR01874
Library of Congress Number 2002115204
ISBN 0-7695-1875-5

ISBN 0-7695-1875-5
90000
9 780769 518756

IEEE Computer Society Order Number PR01874
ISBN 0-7695-1874-5
Library of Congress Control Number 2002115204

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1 714 821 8380
Fax: + 1 714 821 4641
http://computer.org/
csbooks@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1 732 981 0060
Fax: + 1 732 981 9667
http://shop.ieee.org/store/
customer-service@ieee.org

IEEE Computer Society
Asia/Pacific Office
Watanabe Bldg., 1-4-2
Minami-Aoyama
Minato-ku, Tokyo 107-0062
JAPAN
Tel: + 81 3 3408 3118
Fax: + 81 3 3408 3553
tokyo.ofc@computer.org

HICSS Production Editor: Eileen Robichaud Dennis, CMP

Editorial production by Danielle C. Martin, Lorretta Palagi, and Randy Bilof

Cover art production by Joe Daigle/Studio Productions

Printed in the United States of America by The Printing House

CD-ROM produced by InControl Productions, Inc.

COMPUTER
SOCIETY

IEEE

Preface

The Thirty-Sixth Hawaii International Conference on System Sciences (HICSS) was held at the Hilton Waikoloa Village on the Island of Hawaii, on January 6 - 9, 2003. The proceedings consist of over 500 papers in nine major tracks. The topics are summarized on the Overview page, which shows the tracks and minitracks. These Proceedings consist of papers that have not previously been published. They have undergone a detailed peer review process and were selected based on rigorous standards. At the conference, these papers were presented by the authors and discussed in highly interactive sessions. HICSS provides a unique forum for the interchange of ideas, advances, and applications among academicians and practitioners in the information, computing, and system sciences. HICSS is sponsored by the College of Business Administration at the University of Hawaii.

The HICSS series of conferences is now in the 36th year. Very few conferences have been able to grow and develop as HICSS has over this period. The computer age is barely 50 years old, and HICSS has been an important event in the world of computer science and information technology during most of that time.

The conference continues to be one of the best working conferences in computer-related sciences, with a high level of interaction among the leading scientists, engineers, and professionals. Many of the papers and presentations at HICSS have become journal articles, have developed into monographs or books, or have appeared in special issues of journals of wide circulation such as Computer, IEEE Software, Decision Support Systems, and Journal of Management Information Systems. Some major topics such as Group Support Systems, Model Management Systems and Negotiation Support Systems have been largely initiated, developed and documented by presentations and papers at HICSS. Sincere thanks go to all the authors, attendees, coordinators, chairpersons, task force members, advisory committees, and administrative staff who make the conference a success.

The HICSS proceedings have made an impressive contribution to the literature base with almost 70,000 pages during the past 36 years. Since 1999, we have engaged in a process of expanding the Proceedings to include additional forms and media. The objective has been to capture the benefits of electronic media while preserving the benefits of printed Proceedings. This CD/ROM of full papers is included in a sleeve inside the back cover of the book of abstracts. Both are distributed at the Conference and sold afterwards by the IEEE Computer Society Press. You may peruse the abstracts to find papers of interest, then use the CD to read the full paper. Please notice that each abstract has a filename code identifying its Track, Minitrack, and sequence in the program. You may go directly to the file of that name on the CD to read the full paper. You may also use the excellent Electronic Guide or the search facility on the CD to find papers of interest. We have also provided an additional form of the Proceedings so that participants can have access to full papers during the conference. During the conference, a participant may request a printed version of the paper at a nominal cost from our "print-on-demand" facility. We think the resulting spectrum of Proceedings forms makes them a rich resource during as well as after the conference.

Ralph H. Sprague, Jr.
HICSS-36 Conference Chair
E-mail Address: sprague@hawaii.edu

Proceedings Overview for HICSS-36

Collaboration Systems and Technology Track
Asynchronous Learning Networks
Computer Supported Collaborative Learning Requiring Immersive Presence
Collaborative Environments for Value Creation
Collaborative Vision Development
Distributed Collaborative Project Management
Distributed Group Support Systems
Global Virtual Collaboration
Group Support Systems Patterns: ThinkLets and Methodologies
Measuring the Effectiveness of Collaboration Technology
Negotiation Support Systems
Next Generation of Learning Platforms
Technology Supported Learning
User Experiences: Collaboration and Knowledge Management
Wireless Mobile Collaboration
Complex Systems Track
Information Management and Visualization
Markets and Regulation
Robust and Resilient Critical Infrastructure
Security and Reliability
Self-Organized Criticality
Decision Technologies for Management Track
Data Mining and Information Retrieval
Enterprise Architecture, Implementation, and Infrastructure Management
E-Services
Intelligent Decision Support for e-Logistics and Supply Chain Management
Intelligent Systems and Soft Computing
Mobile Commerce
Modeling Nonlinear Natural & Human Systems
Modeling Knowledge-Intensive Processes
Digital Documents and Media Track
Creating The Experience of Media – From Media Design to Media
Digital Libraries
Enterprise Content Management
Information Retrieval and Applications
Knowledge Management, Organizational Memory, and Organizational Learning
XML and the Semantic Wed: Implication and Applications
Emerging Technologies Track
Augmented Cognition and Human-Robot Interaction
IT-Enabled Management of Government: eGovernment
Policy for IT-Enabled Public Administration: ePolicy
IT-Enabled Democracy: e-Democracy
IT-Enabled Government Services: e-Government Services
Telecommunications and Network Convergence: Theory and Practice
Web-Based Education (WBE) Diffusion Across Disciplines [illegible]
Information Technology in Health Care Track
Data & Knowledge Management in Health Care
E-Health Strategies
Healthcare Chain Workflow Management by Use of IT
IT-enabled Governance Structures in Health Care
Telemedicine

Internet and Digital Economy Track
Business-to-Business Electronic Commerce
Electronic Customer Relationship Management
E-Commerce Systems Development Methodologies
Economics and Electronic Commerce
Electronic Marketing
Financial Industry in the Digital Economy
Intellectual Property Rights in the Digital Economy: Issues, Economics, Law and Ethics
Internet & Workflow Automation: Technical & Managerial Issues
Internet Security
Managing Information on the Web
Measuring and Theorizing about Net-enabled Organizations
On-Line Communities in the Digital Economy
Peer-to-Peer (P2P) Computing
Personalization Management Systems
Organizational Systems and Technologies Track
Competitive Strategy and Information Systems
Creativity in Information Systems
Data Warehousing
ERP/ERP II Systems Issues and Answers
Information Systems in Global Business
IT Governance and its Mechanisms
Measuring IT Value
Project Management
Research Methods and Applications in Organizational Systems Research
Social Issues in Organizations
Technology Management in the Knowledge Based Economy
Topics in Organizational Systems and Technology
Trends in Outsourcing of Information Systems
Virtuality in Organizations and Communities
Software Technology Track
Cluster I: Emerging Fields:
Computational Issues in Bioinformatics
Formal Methods for Engineering Special-Purpose Parallel Systems
Peer-2-Peer E-Commerce Systems and Applications
Quality of Service in the Internet
Web-based Information Sharing: Technologies, Strategies, and Approaches
Cluster II: Mobile Computing and Wireless Networks
Mobile Distributed Information Systems
Routing in Wireless and Internet Networks
Quality of Service in Mobile and Wireless Networks
Wireless LANs and PANs
Cluster III: Software Development
Distributed Object and Component-based Software Systems
Domain-Specific Languages
[illegible]
Cluster IV: Trustworthy Computing
Framework and Methods for the Study and Analysis of Trust in Information Systems
Secure and Survivable Software Systems
Testing and Certification of Trustworthy Systems

Software Technology Track

Cluster I: Emerging Fields

Computational Issues in Bioinformatics (STCIB)

Formal Methods for Engineering Special-Purpose Parallel Systems (STFME)

Peer-to-Peer E-Commerce Systems and Applications (STPTP)

Quality of Service in the Internet (STQSI)

Web-Based Information Sharing: Technologies, Strategies and Approaches (STWIS)

Cluster II: Mobile Computing and Wireless Networks

Mobile Distributed Information Systems (STMDI)

Routing in Wireless and Internet Networks (STRWI)

Quality of Service in Mobile and Wireless Networks (STQSM)

Wireless LANs and PANs (STWLP)

Cluster III: Software Development

Distributed Object and Component-Based Software Systems (STDOC)

Domain-Specific Languages (STDSL)

Experimental Software Engineering (STESE)

Cluster IV: Trustworthy Computing

Frameworks and Methods for the Study and Analysis of Trust in Information Systems (STFMS)

Secure and Survivable Software Systems (STSSS)

Testing and Certification of Trustworthy Systems (STTCT)

Experimental Software Engineering (STESE)

Chairs: Karlheinz Kautz and Pekka Abrahamsson

Software engineering theory and practice is still to a large extent based more on faith than on science. Only by contributing to the scientific and empirically grounded body of knowledge within a specific area of application, theory and practice can develop. Experimentation is an important scientific approach to collect empirical data and to test theories as well as to bring light to new phenomena so that theories can be formulated and corrected. This is the background for the emerging field of experimental software engineering.

The focus of this minitrack is on experiments and experimental studies performed in academic or industrial settings where the aim is to study the software professionals' work practices related to the development of software. This minitrack is divided in two three-paper sessions. The papers are briefly introduced in the following.

The three papers in the first session are experiments performed in an academic setting. Syversen, Anda and Sjøberg report the results from an experiment with 26 subjects where they explore how a use case model can best be applied in an object-oriented development process. Serrano, Calero and Piattini describe how to apply the experimental method in metrics definition for multidimensional data models. Their paper gives an overview of the method including a description of how it was applied. The first session is concluded with a paper authored by Liu and Grandon where they empirically explore with 79 subjects how task performance and domain-specific self-efficacy influence the perceived ease of use of object-oriented analysis techniques.

The first two papers in the second session include a set of experiments and an empirical study performed in an industrial setting. Jokela describes five different experiments where the attempt is to assess the quality of the usability engineering processes of four different companies. Jokela explains how the assessment process is iteratively changed and improved based on the results of the earlier experiments. Börjesson and Mathiassen compare two software process improvement initiatives carried out in industry. They focus on factors affecting the implementation success. Dugan, Glinert and Rogers conclude the minitrack by introducing a technology-focused methodology called CAMELOT, which is intended for testing computer supported co-operative work software. They report results from an experiment where the proposed methodology was tried out.

An Evaluation of Applying Use Cases to Construct Design versus Validate Design

E. Syversen, B. Anda, and D.I.K. Sjøberg

(STESE01)

Use case models capture and describe the functional requirements of a software system. A use case driven development process, where a use case model is the principal basis for constructing an object-oriented design, is recommended when applying UML. There are, however, some problems with use case driven development processes and alternative ways of applying a use case model have been proposed. One alternative is to apply the use case model in a responsibility-driven process as a means to validate the design model. We wish to study how a use case model best can be applied in an object-oriented development process and have conducted a pilot experiment with 26 students as subjects to compare a use case driven process against a responsibility-driven process in which a use case model is applied to validate the design model. Each subject was given detailed guidelines on one of the two processes, and used those to construct design models consisting of class and sequence diagrams. The resulting class diagrams were evaluated with regards to realism, that is, how well they satisfied the requirements, size and number of errors. The results show that the validation process produced more realistic class diagrams, but with a larger variation in the number of classes. This indicates that the use case driven process gave more, but not always more appropriate, guidance on how to construct a class diagram The experiences from this pilot experiment were also used to improve the experimental design, and the design of a follow-up experiment is presented.

Experimental Validation of Multidimensional Data Models Metrics

M. Serrano, C. Calero, and M. Piattini

(STESE02)

Multidimensional data models are playing an increasingly prominent role in support of day-to-day business decisions. Due to their significance in taking strategic decisions it is fundamental to assure its quality. Although there are some useful guidelines proposals for designing multidimensional data models, objective indicators (metrics) are needed to help designers and managers to develop quality multidimensional data models. In this paper we present two metrics (Number of Fact Tables, NFT and Number of Dimensional Tables, NDT) we have defined for multidimensional data models and an experiment developed in order to validate them as quality indicators. As a result of this experiment it seems that the number of fact tables can be considered as a solid quality indicator of a multidimensional data model.

How Performance and Self-Efficacy Influence the Ease of Use of Object-Orientation: The Moderating Effect of Prior Training

L. Liu and E.E. Grandon

In this study, we empirically explore how task performance and domain-specific self-efficacy influence the perceived ease of use (PEU) of object-oriented techniques. We hypothesize that both self-efficacy and performance positively influence PEU according to existing literature. However, we postulate that the effect of self-efficacy on PEU, even though still positive, becomes weaker when subjects are given prior training in structured analysis. In contrast, the relationship between performance and PEU becomes stronger when subjects are given the same training. We conducted two tests and collected data from a group of 79 subjects. We found a strong support for most of the research hypotheses.

Assessments of Usability Engineering Processes: Experiences from Experiments

T. Jokela

(STESE04)

We carried out eight assessments of usability engineering processes in four industrial companies, for the purpose of providing a basis for process improvement. The research method was an empirical and iterative one: we carried out assessments, gathered data to understand the success criteria of assessments, and developed step by step an assessment method. The new assessment method has a new kind of process model and is implemented in workshops. The success criteria include issues such as: an assessment is expected to be an effective training occasion; the understandability of the concepts of an assessment approach and the results of an assessment are more important than a standard assessment approach; an assessment should focus on the usability engineering substance of processes rather than on the management of processes. Assessment maturity should be checked before a formal assessment is initiated.

Making SPI Happen: The IDEAL Distribution of Effort

A. Börjesson and L. Mathiassen

(STESE05)

Software Process Improvement (SPI) has become one of the most widely used approaches to increase the capability of software organisations. Many organisations experience a successful start on their SPI initiatives only to realise that the commitment to change weakens significantly after a first phase of initial excitement. This paper explores this problem based on experiences from Ericsson. Two quite similar SPI initiatives situated within the same organisational context are compared and contrasted. Both initiatives were carefully planned and managed following the IDEAL process; they got of to a successful start; and they both developed new or improved processes. But only one of the initiatives led to improvements of engineering practices while the other had little or no effect on the software operation at Ericsson. Our research shows that the effort of the two SPI initiatives is distributed quite differently between the phases of the IDEAL model and between generic actions and actions dedicated to particular software projects. The paper explores this phenomenon both as an indicator and possible explanation of differences in implementation success. Interpretations relating to key issues in SPI are offered together with a discussion of implications for research and practice.

CAMELOT: Technology Focused Testing of CSCW Applications

R.F. Dugan, Jr., E.P. Glinert, and E.H. Rogers

(STESE06)

In this paper we describe CAMELOT, a novel technology-focused methodology for testing Computer [illegible] social issues, CAMELOT is intended for use by application developers, user interface specialists, performance engineers, and quality assurance personnel. The method divides an evaluation into two stages: single user and multi-user. The single user stage is subdivided into general computing and human-computer interaction testing. The multi-user stage is decomposed into distributed computing and human-human interaction testing. The methodology provides a detailed, codified, checklist of testing techniques for each stage. CAMELOT was applied to a conventionally tested, mature CSCW application and uncovered over two dozen problems.

Experimental Validation of Multidimensional Data Models Metrics

Manuel Serrano, Coral Calero, Mario Piattini
ALARCOS Research Group
E.S. Informática. University of Castilla – La Mancha
Paseo de la Universidad, 4
13071 Ciudad Real (Spain)
{Manuel.Serrano, Coral.Calero, Mario.Piattini}@uclm.es

Abstract

Multidimensional data models are playing an increasingly prominent role in support of day-to-day business decisions. Due to their significance in taking strategic decisions it is fundamental to assure its quality. Although there are some useful guidelines proposals for designing multidimensional data models, objective indicators (metrics) are needed to help designers and managers to develop quality multidimensional data models. In this paper we present two metrics (Number of Fact Tables, NFT and Number of Dimensional Tables, NDT) we have defined for multidimensional data models and an experiment developed in order to validate them as quality indicators. As a result of this experiment it seems that the number of fact tables can be considered as a solid quality indicator of a multidimensional data model.

1. Introduction

Nowadays organizations can store vast amounts of data obtained at a relatively low cost, although these data may fail to provide information [11]. Organizations are adapting new technologies as:

- *Data warehouses,* large repositories that integrates data from several sources in an organization for analysis
- *Online analytical processing* (OLAP) systems, which provide fast answers for queries that aggregate large amounts of detail data to find overall trends.
- *Data mining* applications, that seek to discover knowledge by searching semiautomatically for previously unknown patterns and relationships in multidimensional databases

On the three cases, multidimensional data models can be used [19] enabling executives, managers and analysts to make better and faster decisions. Consequently organizations are adopting them to manage information efficiently as "the" main organizational asset. For all these reasons it is fundamental to assure the quality of multidimensional data models.

A multidimensional data model is a direct reflection of the manner in which a business process is viewed. It captures the measurements of importance to a business, and the parameters by which the measurements are broken out. The measurements are referred to as *fact* or *measures*. The parameters by which a fact can be viewed are referred to as *dimensions* [1]. Usually multidimensional data models are represented as star schemas, which consists of one central table and several dimension tables. The measures of interest are stored in the fact table (e.g., sales, inventory). For each dimension of the multidimensional model there exists a dimensional table (e.g., product, time) that stores the information about the dimensions [15].

In figure 1 we present an example of multidimensional data model design, in which we have two fact tables (Returns_Facts and Sales Facts) and four dimensional tables (Product, Store, Return_Reason and Time):

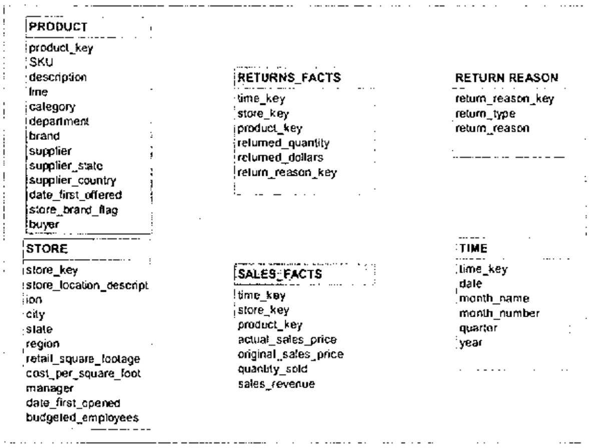


Figure 1. Example of a multidimensional data model design

In recent years different authors have proposed some useful guidelines for designing multidimensional data models. However, more objective indicators are needed to help designers and managers to develop quality multidimensional data models [12] [16] [17]. Also, interesting recommendations for achieving a "good" multidimensional data model have been suggested [17] [1] [13] but quality criteria are not enough on their own to ensure quality in practice, as different people will generally have different interpretations of the same concept. The objective should be to replace intuitive notions of design "quality" with formal, quantitative measures in order to reduce subjectivity and bias in the evaluation process.

Quality is a crucial issue during multidimensional design and there are several approaches focussing on the quality of the multidimensional models [24] [15] [6] but they lack of concrete and quantitative metrics for the measurement of multidimensional models quality. These objective metrics could be used by the designers, for example, for choosing among different designs semantically equivalents.

The final aim of our work is to define a set of metrics for assuring multidimensional data model quality. However, it is not enough with proposing metrics. It is fundamental to be sure that these metrics are really useful for the goal they were conceived. In this sense metrics must be defined with an specific goal and is necessary to prove that they achieve this goal. In this paper we present an experiment made with two metrics for multidimensional data models for knowing if they can be used as indicators of its understandability (one quality factor, as remarked on the standard ISO 9126). These metrics have been chosen in order to continue with the experimental work we are doing [22] with the metrics we have proposed for multidimensional data models [8]. The final goal of all the experimental work is the selection (among the complete set of metrics) of those metrics which have demonstrated to be solid indicators of the multidimensional data models quality.

The two metrics presented have been developed in a methodological way that is explained in the following paragraphs. By applying this method we obtain correct metrics and we know the main mathematical and statistical characteristics of the metrics defined.

On the next section, the method used will be presented. All the information related to the controlled experiment comes in section 3 and in section 4 the experiment planning is detailed. Analysis and interpretation of the results are presented in section 5 and in section 6 some conclusions are presented together with the future work. All the experiments presented have been prepared following the recommendations of [26].

2. Method for defining valid metrics

The method we follow consists of a number of steps which ensure the reliability of the obtained metrics (see figure 2).

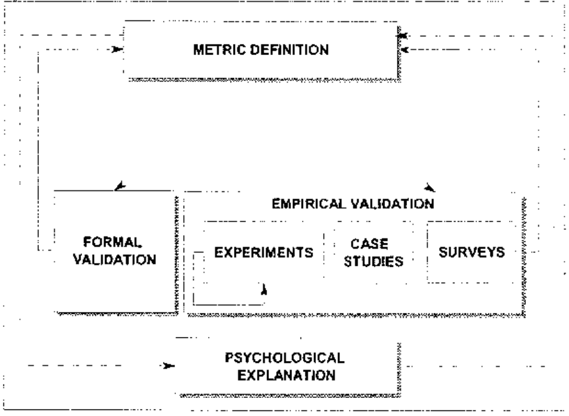


Figure 2. Steps followed in the definition and validation of metrics

In this figure four main activities are represented:

- **Metrics definition.** The definition must be made taking into account the specific characteristics of the multidimensional data model. Two of the metrics that capture one of the main characteristics of a multidimensaional data model (being S a schema of a multidimensional data model) are:

  NFT(S). Number of fact tables of the schema S.
  NDT(S). Number of dimension tables of the schema S.

  For the example given in figure 1, the metrics values are NFT = 2 and NDT = 4.

  Although these metrics could seem too simple, this is not a negative aspect because, as [10] remark this is a desirable property for software metrics in general.

- **Formal validation.** The second step is the formal validation of the metrics. Formal validation helps us to know when and how to apply metrics. There are two main tendencies in metrics validation: the property-based frameworks [7] [18] and those based on measurement theory [21] [27] [25]. Formal validation of the presented metrics in the Zuse's formal framework are ascertained in [8]. As a result of this formal validation we have obtained that NDT is above the ordinal scale and NFT is in the ratio scale, which means that are formally valid software metrics as remarked by [27].

- **Empirical validation.** Here, the objective is to prove the practical utility of the proposed metrics. Basically, we can divide the empirical validation into: experimentation, case studies and surveys [20].

- Experiments are launched when we have control over the situation and want to manipulate behavior directly, precisely and systematically.
- Case studies are used for monitoring projects, activities or assignments. The case study is normally aimed at tracking a specific attribute or establishing relationships between different attributes.
- A survey is often an investigation performed in retrospect, when for example, a tool or technique, has been in use for a while.

Replication of the experiments is also necessary because with only the isolated results of one experiment it is difficult to appreciate how widely applicable the results are and, thus, to assess to what extent they really contribute to the field [3]. In this paper, a controlled experiment conducted with the two metrics presented is explained in detail.

- **Psychological explanation**. Ideally, we should also be able to explain the influence of the values of the metrics from a psychological point of view. Some authors, such as [23], propose the use of cognitive psychology as a reference discipline in method engineering and the study of information modeling. In this way, cognitive psychology theories such as the Adaptive Control of Thought (ACT) [2] could justify the influence of certain metrics in multidimensional data model comprehension. Knowledge of the limitations of the human information processing capacity could also be helpful to establish metrics threshold values for assuring multidimensional data model quality.

As we can see in figure 2, the process of defining and validating metrics is evolutionary and iterative and as a result of the feedback, these metrics could be redefined or discarded depending on their formal and empirical validation or psychological explanation.

3. Controlled Experiment

The objective of our empirical study is to determine if the two metrics presented can be used as mechanisms for controlling the quality of multidimensional data models from a practical point of view (because metrics must be used in the real world and it is there, where they must be useful).

The experimentation presented in this paper is part of the experimental work we are developing with the metrics we have defined for multidimensional data models. On [22] we presented a set of metrics for multidimensional data models and a first study developed with the metrics in order to know if they were correlated with quality (concretely with understandability, one of the factors that influences on quality, [14]). As a result of this experiment we obtained that NFT was a good quality indicator and NDT seemed to be less correlated with quality. We also obtained that the number of tables (NT, defined as the sum of NFT and NDT) seemed to be correlated. Taking into account the definition of the NT metric, we decided to made another experiment in order to obtain a more conclusive results about the metrics used in the definition of NT and knowing if we can use only one of them instead the three (this is, if only NFT influences on the NT metric or if there is also influence of NDT not detected on the previous experiment). As quality is a multidimensional concept, also in this new experiment we work with understandability, one of the factors that influences quality [14].

Taking all these considerations into account and using the GQM template [4] [5] the goal definition of our experiment can be defined as follows:

Analyse	two multidimensional data model metrics
For the purpose of	evaluate
With respect to	the capability to be used as understandability indicators
From the point of view of	data model designers
In the context of	computer science students at the University of Castilla-La Mancha (Spain) enrolled in the final-year of their studies

4. Experiment Planning

The subjects were chosen by convenience, i.e. undergraduate computer science students at the University of Castilla-La Mancha, with knowledge about database and multidimensional data model design.

The experiment is specific since it is focused on two multidimensional data model metrics and addresses a real problem, i.e. can these indicators be used to assess the understandability of multidimensional data models? To this end it investigates the correlation between the two metrics and the understandability of the multidimensional data models.

4.1. Hypotheses

We wish to test the following hypotheses:

- Null hypothesis, H_0: There is no significant correlation between the structural complexity metrics and the understandability of the schemas.
- Alternative hypothesis, H_1: There is a significant correlation between the NFT metric and the understandability of the schemas.
- Alternative hypothesis, H_2: There is a significant correlation between the NDT metric and the understandability of the schemas.

- Alternative hypothesis, H_3: There is a significant correlation between the combination of the NFT and NDT metrics and the understandability of the schemas.

Hypotheses H1 and H2 are stated on the basis that when the number of tables (fact or dimensional) increases (or decreases) the understandability will be affected in some way. Hypothesis H3 is stated to determine if there is any kind of interaction between both metrics.

4.2. Variables

The independent variables are the variables for which the effects should be evaluated. In our experiment these variables correspond with the two metrics under study: NFT and NDT. Each of these metrics (factors) can take two different values (levels): one and two for the NFT metric and three and five for the NDT metric.

The dependent variable is the understandability of the multidimensional data model schema. The understandability of the tests was measured as the accuracy with which they completed the tasks included in each test (so, as the number of correct answers given to the exercises proposed) on a fixed time, this way, all the tests, not necessarily completed, would be taken as valid. The tasks they must perform were asked to answer a set of questions about each of the schemas, consulting the data tables associated with the schemas. We allowed 60 seconds to understand each of the schemas and 90 seconds to answer each of the questions (these times were calculated in a pre-experiment phase with a reduced control group of subjects). The experiment times were controlled by the experiment supervisor.

4.3. Design

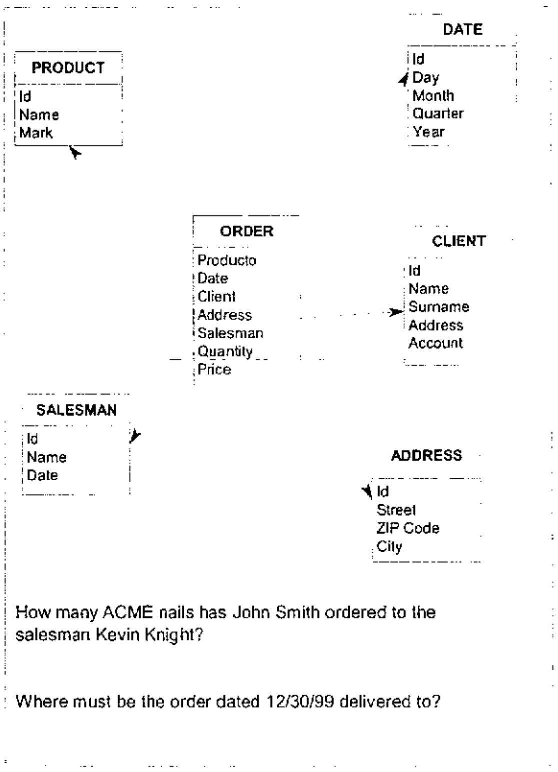
Taking the hypotheses into account, the experiment must consider two factors: NFT and NDT, so we have a factorial model, where all the values a factor can take are combined with all the values of the other factor (in our case, both metrics can take two values, 1 and 2 for the NFT metric and 3 and 5 for the NDT metric). A crossed design as the one described before produces the matrix shown in table 1 where each value of the matrix is a pair (NFT, NDT).

Table 1. Crossed design for the experiment

		Factor B (NDT)	
		LOW	**HIGH**
Factor A (NFT)	**LOW**	1,3	1,5
	HIGH	2,3	2,5

4.4. Data used in the study

The objects used in the experiment were four multidimensional data model schemas. We prepared the material to be handed to the subjects, consisting of the four multidimensional data models schemas, a set of pages with the data of each schema table, and an answer form (see figure 3). In this answer form two different questions related to the specific schema were made to the subjects. With these questions we tried to force the subjects to "navigate" among the dimensional tables thorugh the fact tables of the schema. In this way we can derive if the number of both types of tables has any influence on the number of correct answers given. All the schemas were related to the same universe of discourse that is general enough to be easily understood by each of the subjects.



How many ACME nails has John Smith ordered to the salesman Kevin Knight?

Where must be the order dated 12/30/99 delivered to?

Figure 3. Answer form of the experiment

Before the subjects took the test, the experiment was conducted with a small group of people in order to improve it and ensure that both the experiment and the documentation, were well designed. As a result of this pre-test no changes were necessary.

Experiments were made consecutively during an hour session. Before we started, 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.

The complete documentation was given to each subject. When the time of each test ended (60 seconds to understand and 90 seconds to answer each of the questions), the subjects were informed and, immediately, they could change to the next test.

Tests were performed in a different order by the different subjects in order to avoid learning effects. When the tests were marked, the right answers above the total answers were selected for obtaining the results of the experiment.

4.5. 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 [26] 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 the experiment, all the subjects had approximately, the same experience working with multidimensional data models.
- **Knowledge of the universe of discourse.** The multidimensional data models were from the same universe of discourse and they are 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 half an hour, 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.** The tests were presented as a rehearsal to exam questions, so students participation was overwhelming.
- **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 multidimensional data models and operations representative of real cases in the experiment 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, and with a greater difference between the values of each metric are necessary to obtain more conclusive results regarding the relationship between the metrics and the understandability of multidimensional data models.

5. Analysis and interpretation

To proceed to the analysis, we have first 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 $\alpha=0.05$ which means a 95% level of confidence.

The best test to apply to our data for obtaining results susceptible of being interpreted, taking into account that we have normal data, are parametric tests. The most appropriate one taking into account our hypothesis is the F-statistic which is extremely robust.

In table 2 we present the results obtained from this statistic with the values of the dependent variables in the four cases. In this table the first column represents the source of variation. The second column represents the degrees of freedom, the third one indicates the results obtained for our experiment, and these values must be compared to the table values. In each row of the table we have the two factors of the experiment, the interaction and the total.

The values which appear in the F-ratio column have to be compared with the appropriate value in the tables. In our case the value we need is F1,48 . The numerator is 1 (in the three cases, NFT, NDT and interaction, we have one degree of freedom) and the value forty-eight is calculated as T*I*(n-1) being T the amount of levels NFT can take, I the amount of levels NDT can take (in our case both factors can take two values, so T = I = 2) and n the number of subjects. Consulting the tables we find that $F_{1,48} = 4{,}032$.

The way to know if the hypothesis is true or not is by comparing the value in the tables with the value obtained in the experiment. If the value obtained is greater than the value in the tables, the hypothesis can be considered valid, if not it can be considered invalid.

Table 2. Results of the F-statistic for the experiment

Variation Source	Degree of freedom	F-Ratio
NFT	1	5,851
NDT	1	0,711
NFT * NDT	1	1,157
Total	48	

Analysing the table 2, we obtain that:

Alternative hypothesis, H_1: *There is a significant correlation between the NFT metric and the understandability of the models:* As 5,851 > 4,032, NFT influences the results of the experiment, hence the alternative hypothesis H_1 is valid.

Alternative hypothesis, H2: *There is a significant correlation between the NDT metric and the* understandability *of the models*: As 0,711 < 4,032, NDT does not influence the results of the experiment, hence the alternative hypothesis H_2 is not valid.

Alternative hypothesis, H3: *There is a significant correlation between* the *combination of the NFT and NDT metrics and the understandability of the schemas*: As 1,157 < 4,032, the combination of the NFT and NDT metrics does not influence the results of the experiment, hence the alternative hypothesis H_3 is not valid.

As a conclusion of the experiment we can deduce that the number of fact tables seems to be a solid indicator of the complexity of the multidimensional data models but the number of the dimensional tables is neither an indicator of this complexity nor can it modulate the complexity.

We cannot however consider the results obtained in this experiment as conclusive results. We are aware that it is necessary to replicate the experiment and to carry out new ones in order to confirm our results. It is also necessary to apply these measures to data obtained from "real projects".

6. Conclusions and future work

One of the primary duties of IT professionals must be to assure the quality of information, as it is one of the main assets of organizations. Quality considerations have accompanied multidimensional data model research from the beginning [15]

Although some interesting guidelines have been proposed for designing "good" multidimensional models, more objective indicators are needed. We are working on elaborating a set of valid metrics for measuring multidimensional data model quality, which can help designers in their daily work, for example on selecting between different designs semantically equivalents. In order to create a valid set of metrics, we are not only proposing the metrics, but we are also validating them formally and empirically.

In this paper we have presented an experiment we have undertaken in order to see if the number of fact tables and dimensional tables can have a practical effect on the complexity of a multidimensional schema. It seems that the number of fact tables is a solid indicator of the complexity of a multidimensional schema. However, these are only the first steps and we must carry on more experiments to obtain a set of valid metrics.

Following the MMLC (Measure Model Life Cycle) of [9] with this and other experiments, our research can be classified into the creation phase. At this moment we are going to start different collaborations with enterprises in order to go to the acceptance phase through the systematic experimentation in a context suitable to reproduce the characteristics of the application environment, with real business cases and real users.

Acknowledgement

This research is part of the CALDEA project (TIC 2000-0024-P4-02) supported by the Subdirección General de Proyectos de Investigación, Ministerio de Ciencia y Tecnologia of Spain.

References

[1] Adamson, C. and Venerable, M., *Data Warehouse Design Solutions* John Wiley and Sons, 1998

[2] Anderson, J.R., *The Architecture of Cognition*, Harvard Universitiy Press, Cambridge, 1983.

[3] Basili V., Shull F. and Lanubile F., "Building knowledge through families of experiments", *IEEE Transactions on Software Engineering*, 1999, 25(4), 435-437.

[4] Basili, V.R. and Weiss, D., "A methodology for Collecting Valid Software Engineering Data", *IEEE Transactions on Software Engineering*, 1984, SE-10. No. 6. pp. 728-738

[5] Basili, V. and Rombach, H., "The TAME project: towards improvement-oriented software environments", *IEEE Transactions on Software Engineering*, 1988, 14(6), pp. 728-738

[6] Bouzeghoub, M. and Kedad, Z. "Quality in Data Warehousing" in *Information and Database Quality* Kluwer Academic Publisher, pp 163-198, 2002.

[7] Briand, L.C., Morasca, S. and Basili, V., "Property-based software engineering measurement", *IEEE Transactions on Software Engineering.*, 1996, **22**(1) pp. 68-85

[8] Calero, C., Piattini, M., Pascual, C., Serrano, M., "Towards Datawarehouse Quality Metrics". International Workshop on Design and Management of Data Warehouses (DMDW'01), June, 2001

[9] Cantone, G. and Donzelli, P., "Production and Maintenance of software measurement models", *Journal of Software Engineering and Knowledge Engineering* (5), 2000, pp. 605-626.

[10] Fenton N. and Neil, M. "Software Metrics: a Roadmap", *Future of Software Engineering.* Anthony Finkelstein, ACM, 2000, pp. 359-370.

[11] Gardner, S.R.., "Building the data warehouse", *Communications of the ACM,* September, 1998, Vol. 41, Nr.9. pp. 52-60

[12] Hammergren, T.: *Data Warehousing Building the Corporate Knowledge Base*, International Thomson Computer Press, Milford, 1996.

[13] Inmon, W. H., *Building the Data Warehouse.* John Wiley and Sons, 1997, 2nd edn.

[14] ISO/IEC 9126-1. *Information Technology - Software product quality - Part 1: Quality Model*, 1999.

[15] Jarke, M., Lenzerini, M., Vassilou, Y. and Vassiliadis, P., *Fundamentals of Data Warehouses*, Ed. Springer, 2000.

[16] Kelly, S., *Data Warehousing in Action,* John Wiley & Sons, 1997.

[17] Kimball, R., Reeves, L., Ross, M. and Thornthwaite, W., *The Data Warehouse Lifecycle Toolkit,* John Wiley and Sons, 1998.

[18] Morasca S. and Briand L., "Towards a Theoretical Framework for measuring software attributes", *Fourth International.Software Metrics Symposium (METRICS 1997)*, 1997, 119-126.

[19] Petersen T.B. and Jensen, C.S., "Multidimensional Database Technology", *Computer*, December, 2001, Col 34, Nr. 12 pp 40-46

[20] Pflegeer S., "Experimental design and analysis in software engineering part 1-5", *Software Engineering Notes.* ACM Sigsoft 19(4), 16-20; 20(1), 22-26; 20(2), 14-16; 20(3), 13-15; 20(4), 14-17., 1994-1995.

[21] Poels, G. and Dedene, G., "Distance-based software measurement: necessary and sufficient properties for software measures", *Information and Software Technology*, 2000, 42(1), 35-46.

[22] Serrano, M. Calero, C., Piattini, M., "Validating Metrics for Data Warehouses", *Proceedings of the Conference on Empirical Assessment in Software Engineering (EASE 2002), Keele, UK, 8-10 April 2002*

[23] Siau K., "Information Modeling and Method Engineering: A Psychological Perspective", *Journal of Database Management, 10 (4),* 1999, pp. 44-50.

[24] Vassiliadis, P., *Data Warehouse Modeling and Quality Issues.* PhD thesis, Department of Electrical and Computer Engineering, 2000.

[25] Whitmire S. *Object Oriented Design Measurement.* John Wiley & Sons, Inc., 1997.

[26] Wohlin, C., Runeson, P., Höst, M., Ohlson, M., Regnell, B. and Wesslén, A., *Experimentation in Software Engineering: An Introduction*, Kluwer Academic Publishers, 2000

[27] Zuse, H., *A Framework of Software Measurement*, Walter de Gruyter, 1998.