

*INFORMATION SYSTEMS –  
RESEARCH, TEACHING AND  
PRACTICE*

*Proceedings of the 5th UKAIS Conference  
University of Wales Institute, Cardiff  
26 – 28 April 2000*

**Mc  
Graw  
Hill**

Published by McGRAW-HILL Publishing Company  
Shoppenhangers Road, Maidenhead, Berkshire, SL6 2QL, England  
Tel 01628 502500; Fax 01628 770224

---

**British Library Cataloguing in Publication Data**

The CIP data of this title is available from the British Library

**Library of Congress Cataloging-in-Publication Data**

The CIP data of this title is available from the Library of Congress, Washington DC, USA

ISBN 0 07 709755 6

Copyright © 2000 McGraw-Hill International (UK) Limited. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of McGraw-Hill International (UK) Limited.

Printed and bound in Great Britain at the Basingstoke Press (75) Ltd, Basingstoke

# ***PREFACE***

**Paul Beynon-Davies  
Michael D. Williams  
Ian Beeson**

## **1. UKAIS2000**

*CROESO I GYMRU, CROESO I'R CYNHADLEDD UKAIS.  
WELCOME TO WALES, WELCOME TO THE UKAIS CONFERENCE.*

This years' UKAIS conference is held at the Cyncoed campus of the University of Wales Institute, Cardiff. The site was chosen in a selection process conducted by the Wales and West Regional Group of the UKAIS. The group is hosting the conference this year.

We feel that Cardiff offers an outstanding venue for the conference and welcome the fact that it is the first time the conference has been held outside England. We have attempted to formulate a varied and interesting conference programme that we hope everyone will enjoy.

## **2. REVIEW PROCESS**

A grand total of 95 papers were submitted to the conference this year. Each paper was sent for blind review to two referees. The programme committee of the conference this year decided to introduce three classes of papers with the aim of encouraging inclusion amongst IS academics. In total 64 full papers, 9 short papers and 12 discussion papers were accepted.

## **3. PLENARY SPEAKERS, PANELS AND TUTORIALS**

We have been lucky in getting a number of eminent persons from IS academia to speak at this years conference, indicating the increasing role that the conference plays in the IS calendar. The conference also includes a number of panel sessions and one tutorial session on the interesting subject of open source software.

## **4. PAPER STREAMS**

The papers were assigned to streams on the basis of the UKAIS sub-areas contained in the definition for information systems published by the Academy. Papers were also categorised in terms of being either research, practice or teaching papers.

## **5. LIST OF REVIEWERS**

Thanks to the following persons who acted as reviewers of the papers submitted this year.

Alison Adam, UMIST  
 David Avison, University of Southampton  
 Yehuda Baruch, University of East Anglia  
 Ian Beeson, University of the West of  
 England  
 Frances Bell, University of Salford  
 Paul Beynon-Davies, University of  
 Glamorgan  
 Michael Brittain, University of Wellington  
 (NZ)  
 Laurence Brooks, University of York  
 Tom Gough, University of Leeds  
 Ray Hackney, Manchester Metropolitan  
 University  
 Guy Fitzgerald, Brunel University  
 Donal Flynn, UMIST  
 Sue Fowell, University of New South Wales  
 (Australia)  
 Matthew Jones, University of Cambridge  
 Chris Kimble, University of York  
 Lydia Lau, University of Leeds  
 Margi Levy, University of Warwick  
 Feng Li, University of Strathclyde

Michael Lloyd-Williams, Manchester  
 Metropolitan  
 Nathalie Mitev, London School of  
 Economics  
 Philip Powell, University of London  
 Hugh Preston, University of Wales,  
 Aberystwyth  
 Steve Probert, Cranfield University  
 Nancy Russo, Northern Illinois University  
 (USA)  
 Carsten Sorensen, London School of  
 Economics  
 Frank Stowell, De Montfort University  
 Richard Vidgen, University of Bath  
 David Wainwright, University of Durham  
 Edgar Whitley, London School of  
 Economics  
 Bernard Williams, University of East Anglia  
 David Wilson, University of London  
 Brent Work, University of Surrey  
 Gareth Griffiths (Manchester Metropolitan  
 University)

## 6. MEMBERS OF THE UKAIS BOARD

Our thanks also to the UKAIS board for their support:

Laurence Brooks (York) - laurence.brooks@cs.york.ac.uk  
 Ray Hackney (Manchester Metropolitan) - r.hackney@mmu.ac.uk  
 Philip Powell (Newsletter Editor: Bath) - mnspp@management.bath.ac.uk  
 Feroz Siddiqui (Student Representative) - feroz.siddiqui@brunel.ac.uk  
 Steve Smithson (Treasurer:LSE) - s.smithson@lse.ac.uk  
 Frank Stowell (President: De Montfort) - fstowell@dmu.ac.uk  
 John Ward (Membership Secretary: Cranfield): j.m.ward@cranfield.ac.uk  
 Trevor Wood-Harper (Salford) - A.T.Wood-Harper@salford.ac.uk

# TABLE OF CONTENTS

	<i>Table of Contents</i>		v
	<i>Preface</i>		xi
	<i>Full Papers</i>		
1	Gottschalk Peter	Business Benefits from Data Warehousing: the Case of Oil Rig Åsgard B in the North Sea ... ..	1
2	Mutch Alistair	Lessons from history; or, reflections on the field of information systems following a reading of Checkland and Holwell ... ..	10
3	Chester Myrvin F.	Formal Methods and the Scientific Approach to Software Development ... ..	19
4	Rana Omer.F.	Beyond Components: a flexible approach to information systems design ... ..	28
5	McIntosh Steve.B.	Barriers to Virtual Teams ... ..	38
6	Feng Li	IS Success: an organisational Learning Perspective ... ..	49
7	Bandula Jayatilaka	Turning non-cooperative applications into cooperative ones by adding dynamics to the information system ... ..	57
	Philippe Roose		
	Marc Dalmau		
	Philippe Aniorité		
8	Owens Ian	Aligning e-Commerce with Business Strategy: the case of the Bank of Scotland ... ..	67
	Robertson David		
9	Grzeszczyk		
	Tadeusz Adam	Rough Sales Forecasting System ... ..	76
10	Preston Hugh	A framework of intent for re-engineering the medical profession ... ..	84
	Jackson Fiona		
11	Daniel Elizabeth	Factors Determining the Successful Development of Marketing Information Systems ... ..	92
	Wilson Hugh		
	Sutherland Frances		
12	Genero Marcelo	Formalisation of Metrics for Conceptual Data Models ... ..	99
	Piattini Mario		
	Calero Coral		
13	Calero Coral	Metrics for Relational Database Maintainability ... ..	109
	Piattini Mario		
	Genero Marcela		
14	Butler Tom	Beyond Simple Causal Models: Capturing the Context and Process of User Participatio ... ..	120
15	Shi J		
	Huang G Q	Distributed and Synchronized Process Planning on the Web .	130

16	Standing C. Benson S.	An Effective Framework for Policy and Infrastructure Issues for E-Commerce: Implications for Business and University Education and Research ... ..	139
17	Work Brent Gay Virginia	Using the Case Method in Teaching Information Systems to Undergraduates ... ..	148
18	Warren M. Hutchinson W.	Terrorism on the Web ... ..	154
19	Waring Teresa Wainwright David	Integrated Information Systems: Rhetoric and Reality ... ..	165
20	Watson Ray Hackney Ray McBride Neil	Out of Tune Out of Time: hermeneutics of strategic information systems planning ... ..	175
21	Rupino da Cunha Paulo Dias de Figueiredo Antonio	Interaction in Organisations: a missing element in information systems planning ... ..	184
22	Hlupic Vlatka	Improving the Success of Business Process Change: The REBUS Strategy ... ..	192
23	Hobbs Mike	An Agent-Based Information System for the NHS Electronic Health Record ... ..	202
24	Kelly Jacquie	Information Systems Design: enabling students to be more creative ... ..	211
25	Bergvall-Kåreborn Birgitta	Using the Concept Qualifying Function as a Conceptual Tool for Modelling – A case Study ... ..	219
26	Saleh Yasser M Taylor W Andrew	Information Systems Success and Effectiveness: under the microscope one more time ... ..	228
27	Estay-Niculcar Christian A.	'La Estructura Ausente': the social construction of information systems in projects ...	237
28	Vasconcelos José Kimble Chris Gouveia Feliz Ribeiro	A design for a Group Memory System using Ontologies ... ..	246
29	Kimble Chris Selby William	An Interdisciplinary study of information systems: Christopher Alexander and IS Failure ... ..	256
30	Bowers David S.	Database schema improvement techniques for CASE tools ..	266
31	Champion Donna Stowell Frank	Interpretivist Discovery for Information System Design .....	276
32	Choudhury Islam Patel Dilip	Generic Reusable Business Object Modelling for Strategic Information Systems Development ... ..	283
33	Mann Joan	The Dynamics of Runaway Information Systems Development Projects ... ..	293
34	Benson Steve Standing Craig	Synergising The Personal-Corporate Knowledge Nexus .....	303

35	Guo Minghong Wu Zimin Stowell Frank Cowell John	AM/OO Modelling: Achievements and Problems ... ..	312
36	Patel Nayna Hlupic Vlatka	Business Process Modelling of a Telephony System ... ..	322
37	Page Steve Orange Graham	A Knowledge Management Meta-Model for Systems Development: A Cultural Perspective ... ..	332
38	Stephens Robert	Authenticity and Ethics in IS: the limits of theoretical thinking ... ..	341
39	Kamm Richard Plant Nick Moggridge Anne	Practice in IS Education: live projects for livelier students ...	349
40	Nevill Nuala Wood-Harper Trevor	People Doing Information Systems Research: A view from leaders in the UK ... ..	356
41	de Cesare Sergio Capozza Francesco	Modelling Issues in a Transport Information System ... ..	364
42	Patel Dilip Doherty Neil Banerjee Shantanu	Factors Affecting the Successful Outcome of Systems Development Projects ... ..	375
43	Parry Mark McAulay Laurie Smith Matt Doherty Neil	Managers' Use of New Technologies: managing the barriers to the use of Videoconferencing and the Net ... ..	384
44	Oriogun Peter K Gilchrist Robert	The Effect of Studying an Information Systems Module on Students' Results in a Software Engineering Module ... ..	394
45	Singh Gurmak	The Changing Role of IT departments in Large Organisations ... ..	402
46	King Malcolm Hussin Husnayati	Identifying IT Alignment in Small Firms ... ..	409
47	Cragg Paul Plant Nick	Rural Regeneration and Community Information Systems in Context: "joined up thinking" in practice? ... ..	419
48	Davis Chris	Flow and Convergence in Information Systems Evaluation ..	429
49	Beeson Ian Venters WJ Wood JRG	Knowledge Tools within a Knowledge Organisation: How to support communication in a multinational organisation.	437
50	Beynon-Davies Paul Blyth Andrew Owens Ian Lloyd-Williams Michael	IS failure, Evaluation and Organisational learning ... ..	444
51	Feng Junkang	The notion of 'information bearing capacity' of a conceptual data schema ... ..	453

52	Flynn Donal Hussain Zahid	Metaphorical Assumptions Of Stakeholders Involved In Constructing IS Requirements ... ..	464
53	Baldwin Lynne P. Eldabi Tillal Paul Ray J.	Simulation Modelling as an Informative Tool in Healthcare Management Decision-Making ... ..	473
54	Stocking Sue Lloyd-Williams Michael Beynon-Davies Paul Collins Janet	Preparing for the Millennium: A Study of Public Sector Organisations ... ..	481
55	Nicholson Isobel A.	A Longitudinal Study of Software Developers' Views about How to Talk with Stakeholders ... ..	490
56	Galal Galal H	Grounding Systems Requirements	496
57	Birchall David W. Tovstiga George	Information Systems Requirements in Support of the Firm's Portfolio of Knowledge-Driven Capabilities ... ..	506
58	Brooks Laurence Kiddle Laurence	Critical Success Factors (CSF's) and Information Systems Development ... ..	516
59	Leong Ang Cheng	Systematic Derivation of IDEF1x Models from IDEF0 MODELS ... ..	528
60	Khalifa Ghassan Irani Zahir Baldwin Lynne P	Factors Impacting The Selection Of IT/IS Evaluation Methods ... ..	539
61	Farthing Dave W Beynon-Davies Paul	A survey of SSADM usage: the case for teaching design techniques ... ..	555
62	Collins Janet Lloyd-Williams Michael Beynon-Davies Paul Stocking Sue	Intra-Vendor System Migration in a Publicly Funded Organization ... ..	564
63	Shah Mahmood Hussain	Evaluating ERP Systems: a framework for decision making . Towards the integrated monitoring and evaluation system	574
64	Kostas Metaxiotis	IMES ... ..	582
65	Jacobs Gabriel Dowland William	The Dot-Com Economy in Wales: a long road ahead ... ..	590



*Short Papers*

66	Blackwell Geoff	Information Systems Research : A new researcher's dilemma and a tentative solution ... ..	597
67	Casimir Rommert J.	A logistic View of Data Warehousing ... ..	603
68	Maritz Mark	A Model to Support South African University Faculty in their use of the Internet ... ..	610
69	Green Patrick Dingley Sharon	Meta-framework for multiple technique component-based development ... ..	615
70	Perkins John Dingley Sharon	Skills Requirement Analysis for E-business: Fleshing out the technology ... ..	621
71	Wilson David W.	Maturity Models and Information Systems: from s-Curves to e-Commerce ... ..	628
72	Makewita Sumedha M.	Potential Knowledge Vs Realised Knowledge: a theory for the effectiveness of knowledge networks ... ..	634
73	Rawlings Peter	Information systems and the changing nature of international zoological field research ... ..	641
74	Howcroft Debra Wilson Melanie	Rescuing reality in information systems research ... ..	645

*Discussion Papers*

75	Chatham Robina Keith Patching	The IT Stereotype and its Organisational Implications ... ..	652
76	Hughes Robert T.	Computer-supported dispersed working: a case study ... ..	655
77	Blyth Andrew Beynon-Davies Paul	Trust and Electronic Commerce ... ..	657
78	Ramanath Ana Marie Macredie Robert	Software Life Cycle Based Interorganisational Systems Development Projects ... ..	660
79	Latham Ann	Matching Graduates with Organisational Needs ... ..	663
80	McClelland Robert James	Campus Wide Information Systems: an integration development for higher education ... ..	665
81	Choudrie Jyoti Hlupic Vlatka	Identifying and Understanding the Human and Organisational Aspects that Affect the Teams Undertaking Business Process Engineering ... ..	668
82	Doherty Neil Robinson Phil	The Impact of Organisational Issues on the Implementation of a Community Information System: A Case Study Approach ... ..	671
83	Thomas Rod	On the Relationship of Structuration Theory to Interpretive Information Systems Research ... ..	673
84	Korn J	Analysis of Information Systems ... ..	675

85	Walters David Toase Catherine Greenwood Anthony	The decision making behaviour of small businesses acquiring computer systems ... ..	677
86	Sa Jin Green Stewart Beeson Ian	Modelling Multiple Software Processes in the Business Context ... ..	680

*Panels & Tutorials*

Chapter	Author(s)	Panel/Tutorial Title	Page
87	Fitzgerald Guy Russo Nancy Fitzgerald Brian Beynon-Davies Paul	Information Systems Development ... ..	682
88	Gouveia Feliz Ribeiro Jerrams-Smith Jenniffer Badii Atta	Content Management and Personalisation ... ..	694
89	Feller Joseph	Understanding Open Source Software ... ..	698

# CHAPTER 13

## METRICS FOR RELATIONAL DATABASES MAINTAINABILITY

Coral Calero, Mario Piattini, Marcela Genero, Manuel Serrano,  
Ismael Caballero  
University of Castilla-La Mancha

### **Abstract.**

*Software measurement is widely recognised as an effective means to understand, control, predict and improve software development and maintenance projects. In the last decades a huge amount of software metrics has been proposed, but they focused primarily on programs. Metrics for databases have been neglected, mainly because databases have developed a secondary role in Information Systems (IS) infrastructure until a few years ago. But nowadays, databases are the core of IS, influencing considerably their maintenance. Maintainability is achieved by three factors: understandability, modifiability and testability, which in turn are influenced by complexity. This paper proposes different metrics for measuring relational database complexity. These measures are characterised using the measurement theory, particularly the formal framework put forward by Zuse (Zuse, 1998). Some empirical experiments were carried out in order to validate the proposed metrics.*

### 1. INTRODUCTION

Metrics are useful mechanisms for the improvement of the quality of software products, specially maintenance which is the most important problem of software development, ranging between 60 and 90 percent of life-cycle costs (Card and Glass, 1990) (Kim and March, 1995). Software measurement is widely recognised as an effective means to understand, monitor, control, predict and improve software development and maintenance projects (Briand et al, 1996). Measurement is used not only for understanding, controlling, and improving development, but also for determining the best ways to help practitioners and researchers (Pfleeger, 1997).

Software engineers have been putting forward huge quantities of metrics for software products, processes and resources (Melton, 1996) (Fenton and Pfleeger, 1997). Unfortunately, almost all the metrics proposed since McCabe's cyclomatic number (McCabe, 1976) until nowadays, have focused on program characteristics disregarding databases. As far as databases are concerned, metrics have been used for comparing data models rather than the schemata themselves. Several authors (Juhn and Naumann, 1985) (Jarvenpaa and Machesky 1986) (Shoval and Even-Chaime, 1987) (Batra et al, 1990) [20], (Rossi and Brinkkemper, 1996) have compared the best well-known models such as E/R, NIAM and the relational ones using different metrics. Although we think this work is interesting, metrics for comparing schemata are needed mainly for practical purposes, such as choosing between different design alternatives or giving designers limit values for certain characteristics (analogously to value 10 for McCabe's complexity of programs).

Design metrics for data aspects are important because the size of data and their nature contribute to many aspects of a system like the amount of effort to develop. Then measure data can potentially help control and predict aspects of the data model during the software development process (MacDonell, Shepperd and Sallis, 1997).

However, as Sneed and Foshag remark: "metrics for databases have been neglected in the metric community" (Sneed and Foshag, 1998). This neglect could be explained as, until recently, databases have developed just a secondary role with minor contributions to the complexity of the overall system. Nowadays, databases have been introduced in most of the important IS, becoming their essential core. For this reason, we think it is very important to measure databases and understand their contribution to the overall IS maintainability.

Maintainability is achieved by three factors: understandability, modifiability and testability, which in turn are influenced by complexity (Li and Chen, 1987). However, a general complexity measure is "the impossible holy grail" (Fenton, 1994). Henderson-Sellers distinguishes three types of complexity: computational, psychological and representational, and for psychological complexity he considers three components: problem complexity, human cognitive factors and product complexity (Henderson-Sellers, 1996). The last one is our focus.

Two different measures related with referential integrity are proposed in this paper for relational database schemata complexity: Depth of Referential Tree (DRT) and Referential degree (RD). These measures are characterised using measurement theory, particularly the formal framework proposed by Zuse [37].

Section 2 presents the different measures proposed. We give a brief introduction to the Zuse's framework in section 3, using it to characterise the metrics. Section 4 presents the experiments performed with the referential integrity related metrics. The conclusions and suggestions for future work are included in the last section.

## 2. MEASURES PROPOSED FOR RELATIONAL DATABASES

Since the Dr. Codd's proposal of the relational model in the late sixties (Codd, 1970) the database field has been extensively researched and relational database products have developed a very important industry.

The only indicator used to measure the quality of relational databases has been the normalisation theory, upon which (Gray, Carey, McGlynn, Pengelly, 1991) propose to obtain a normalisation ratio. However, foreign key is one of the main concepts of the relational model. It can be defined as follows (Date, 1995): Let R2 be a base relation. Then a foreign key in R2 is a subset of the set of attributes of R2, say FK, such that:

1. There exists a base relation R1 (R1 and R2 not necessarily distinct) with a candidate key CK, and
2. For all time, each value of FK in the current value of R2 is identical to the value of CK in some tuple in the current value of R1.

Related with the foreign key concept, the relational model includes the referential integrity rule: the database must not contain any unmatched foreign key values (Date, 1995).

### 2.1 Measures definition

Based on the referential integrity, we propose two different metrics:

1. Depth Referential Tree (DRT). DRT is defined as the length of the longest referential path in the database schema and  $DRT(A)$  is the length of the longest referential path in the table A. The cycles are considered only once.
  2. Referential Degree (RD). The RD metric is defined as the number of foreign keys in the schema and  $RD(A)$  is the number of foreign keys in the table A.
- As a example, if we consider the following schema:

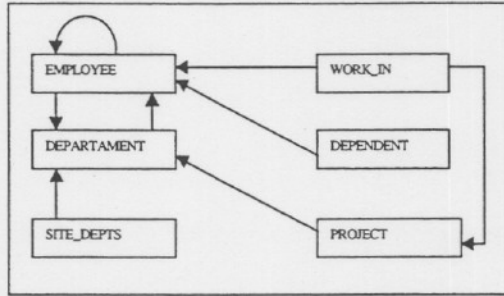


Figure 1. Schema for the example

Derived from the next example taken from (Elmasri and Navathe, 1997):

```

CREATE TABLE EMPLOYEE(
    NAMEP    VARCHAR(15) NOT NULL,
    INIC     CHAR,
    SURNAME  VARCHAR(15) NOT NULL,
    NSS      CHAR(9) NOT NULL,
    DATEN    DATE,
    ADDRESS  VARCHAR(30),
    SEX      CHAR,
    SALARY   DECIMAL(10,2),
    NSSSUPER CHAR(9),
    ND       INT NOT NULL,
    CONSTRAINT CLPEMP
        PRIMARY KEY (NSS),
    CONSTRAINT CLESUPEREMP
        FOREIGN KEY (NSSUPER) REFERENCES
EMPLOYEE(NSS)
        ON DELETE SET NULL ON UPDATE CASCADE,
    CONSTRAINT CLEDEPTOEMP
        FOREIGN KEY (ND) REFERENCES DEPARTAMENT
(NUMBERD)
        ON DELETE SET DEFAULT ON UPDATE CASCADE);

CREATE TABLE DEPARTAMENT(
    NAMED    VARCHAR(15) NOT NULL,
    NUMBERD  INT NOT NULL,
    NSSGTE   CHAR(9) NOT NULL,
    DATEINICGTE DATE,
    CONSTRAINT CLPDEPTO
        PRIMARY KEY (NUMBERD),
    CONSTRAINT CLSDEPTO
        UNIQUE(NOMBRED),
    CONSTRAINT CLEGTDEPTO
        FOREIGN KEY (NSSGTE) REFERENCES EMPLOYEE
(NSS)
        ON DELETE SET DEFAULT ON UPDATE CASCADE);

CREATE TABLE SITE_DEPTS(
    NUMBERD  INT NOT NULL,
    SITED    VARCHAR(15) NOT NULL,
    PRIMARY KEY (NUMBERD, SITED),
    FOREIGN KEY (NUMBERD) REFERENCES DEPARTAMENT
(NUMBERD)
        ON DELETE CASCADE ON UPDATE CASCADE);

CREATE TABLE PROJECT(
    NAMEPR   VARCHAR(15) NOT NULL,
    NUMBERPR INT NOT NULL,
    SITEPR   VARCHAR(15),
    NUMD     INT NOT NULL,
    PRIMARY KEY (NUMBERPR),
    UNIQUE (NAMEPR),
    FOREIGN KEY (NUMD) REFERENCES DEPARTAMENT
(NUMBERD));

CREATE TABLE WORK_IN(
    NSSE     CHAR(9) NOT NULL,
    NUMP     INT NOT NULL,
    HOURS    DECIMAL(3,1) NOT NULL,
    PRIMARY KEY (NSSE, NUMP),
    FOREIGN KEY (NSSE) REFERENCES EMPLOYEE (NSS),
    FOREIGN KEY (NUMP) REFERENCES PROJECT (NUMBERP));

CREATE TABLE DEPENDENT(
    NSSE     CHAR(9) NOT NULL,
    NAME_DEPEND VARCHAR(15) NOT NULL,
    SEX      CHAR,
    DATEAN   DATE,
    RELATION VARCHAR(8),
    PRIMARY KEY (NSSE, NAME_DEPEND),
    FOREIGN KEY (NSSE) REFERENCES EMPLOYEE (NSS);
    
```

We obtain the following values for the metrics (table 1):

	DRT	RD
EMPLOYEE	3	2
DEPARTAMENT	3	1
SITE_DEPTS	4	1
PROJECT	4	1
WORK_IN	5	2
DEPENDENT	4	1
SCHEMA	5	8

Table 1. Values of the metrics in the example

### 3. AN INTRODUCTION TO ZUSE MEASUREMENT FRAMEWORK

Several frameworks for measure characterisation have been proposed (Briand, Morasca and Basili, 1996) (Weyuker, 1988). In this paper we will follow the formal framework of [37] in order to describe the properties of the metrics defined above. This framework is based on an extension of the classical measurement theory, which gives a sound basis of software measures, their validation and criteria for measurement scales.

Dr. Zuse describes measurement as a detour, "a necessary one because humans mostly are not able to make clear and objective decisions or judgements" ([37]). Measuring is more than producing numbers, it is the combination of empirical entities with numerical entities. People are interested in establishing "empirical relations" between objects, such as "higher than" or "equally high or higher than". These empirical relations will be indicated by the symbols " $\bullet >$ " and " $\bullet \geq$ " respectively. We called Empirical Relational System a triple:  $A = (A, \bullet \geq, \circ)$ , where  $A$  is a non-empty set of objects,  $\bullet \geq$  is an empirical relation on  $A$  and  $\circ$  is a closed binary (concatenation) operation on  $A$ .

Zuse defines a set of properties for measures, which characterise different measurement structures. The most important ones are shown in table 2:

MODIFIED EXTENSIVE STRUCTURE	INDEPENDENCE CONDITIONS	MODIFIED RELATION OF BELIEF
<p>Axiom1: <math>(A, \bullet \geq)</math> (weak order)</p> <p>Axiom2: <math>A1 \circ A2 \bullet \geq A1</math> (positivity)</p> <p>Axiom3: <math>A1 \circ (A2 \circ A3) = (A1 \circ A2) \circ A3</math> (weak associativity)</p> <p>Axiom4: <math>A1 \circ A2 = A2 \circ A1</math> (weak commutativity)</p> <p>Axiom5: <math>A1 \bullet \geq A2 \Rightarrow A1 \circ A \bullet \geq A2 \circ A</math> (weak monotonicity)</p> <p>Axiom6: If <math>A3 \bullet &gt; A4</math> then for any <math>A1, A2</math>, then there exists a natural number <math>n</math>, such that <math>A1 \circ nA3 \bullet \geq A2 \circ nA4</math> (Archimedean axiom)</p>	<p>C1: <math>A1 = A2 \Rightarrow A1 \circ A = A2 \circ A</math> and <math>A1 = A2 \Rightarrow A \circ A1 = A \circ A2</math></p> <p>C2: <math>A1 = A2 \Leftrightarrow A1 \circ A = A2 \circ A</math> and <math>A1 = A2 \Leftrightarrow A \circ A1 = A \circ A2</math></p> <p>C3: <math>A1 \bullet \geq A2 \Rightarrow A1 \circ A \bullet \geq A2 \circ A</math>, and <math>A1 \bullet \geq A2 \Rightarrow A \circ A1 \bullet \geq A \circ A2</math></p> <p>C4: <math>A1 \bullet \geq A2 \Leftrightarrow A1 \circ A \bullet \geq A2 \circ A</math>, and <math>A1 \bullet \geq A2 \Leftrightarrow A \circ A1 \bullet \geq A \circ A2</math></p>	<p>MRB1: <math>\forall A, B \in \mathcal{S}: A \bullet \geq B</math> or <math>B \bullet \geq A</math> (completeness)</p> <p>MRB2: <math>\forall A, B, C \in \mathcal{S}: A \bullet \geq B</math> and <math>B \bullet \geq C \Rightarrow A \bullet \geq C</math> (transitivity)</p> <p>MRB3: <math>\forall A \supseteq B \Rightarrow A \bullet \geq B</math> (dominance axiom)</p> <p>MRB4: <math>\forall (A \supset B, A \cap C = \emptyset) \Rightarrow (A \bullet \geq B \Rightarrow A \cup C \bullet \geq B \cup C)</math> (partial monotonicity)</p> <p>MRB5: <math>\forall A \in \mathcal{S}: A \bullet \geq 0</math> (positivity)</p>
<p>As we know, binary relation <math>\bullet \geq</math> is called weak order if it is transitive and complete:  <math>A1 \bullet \geq A2</math>, and <math>A2 \bullet \geq A3 \Rightarrow A1 \bullet \geq A3</math>  <math>A1 \bullet \geq A2</math> or <math>A2 \bullet \geq A1</math></p>	<p>Where <math>A1 = A2</math> if and only if <math>A1 \bullet \geq A2</math> and <math>A2 \bullet \geq A1</math>, and <math>A1 \bullet &gt; A2</math> if and only if <math>A1 \bullet \geq A2</math> and not <math>(A2 \bullet \geq A1)</math>.</p>	

Table 2. Zuse's formal framework properties

There exists five scale types that are defined by admissible transformations. They are, in hierarchical order: nominal, ordinal, interval, ratio and absolute. Each scale type is defined by admissible transformations. Software measurement starts with the ordinal scale ([37]). Measures may be classified in a scale type, depending on whether they assume an extensive structure or not.

When a measure accomplishes this structure, it also accomplishes the independence conditions and can be used on the ratio scale levels.

If a measure does not satisfy the modified extensive structure, the combination rule (that describes the properties of the software measure clearly) will exist or not depending on the independence conditions. When a measure assumes the independence conditions but not the modified extensive structure, the scale type is the ordinal scale.

### 3.1 Formal verification of the proposed metrics

In relational database systems, and for our purposes, the Empirical Relational System could be defined as:  $R = (R, \bullet \geq, \circ)$ , where  $R$  is a non-empty set of relations (tables),  $\bullet \geq$  is the empirical relation "more or equal complex than" on  $R$  and  $\circ$  is a closed binary (concatenation) operation on  $R$ . In our case we will choose natural join as the concatenation operation. Depending on the characteristics of the combined tables, natural join can derive in Cartesian product. Furthermore, it is possible to make the natural join through foreign key-primary key or between any columns of two tables defined over the same domain. All these characteristics of the natural join will be useful in order to design the combination rule of the metrics:

1. In order to obtain the combination rule for DRT we may think of several possibilities: the natural join may derive in a Cartesian product or can be made between columns not related by referential integrity (in both cases the referential paths are not affected by the combination and the final value of the metric is the longest referential path), or the natural join may be made by foreign key (the length of the referential paths may vary, decreasing in one). So, we can generalise and define the combination rule as:  $DRT(R_i \circ R_j) = \max(DRT(R_i), DRT(R_j)) - v$  being  $v$  a variable.
2. In order to obtain the combination rule for RD, we can observe that if the concatenation (by natural join) between tables is made by foreign key, the number of foreign keys are affected (decreasing in one), and are not affected in other cases. So, we can characterise the combination rule for RD as:  $RD(R_i \circ R_j) = RD(R_i) + RD(R_j) - v$  where  $v$  is a variable that can be 0 or 1.

Table 3 presents the results obtained for the metrics with the combination functions described:

Properties	DRT	RD
Axiom 1	YES	YES
Axiom 2	NO	NO
Axiom 3	YES	YES
Axiom 4	YES	YES
Axiom 5	NO	NO
Axiom 6	NO	NO
Ind Cond. 1	NO	NO
Ind Cond. 2	NO	NO
Ind Cond. 3	NO	NO
Ind Cond. 4	NO	NO
MRB 1	YES	YES
MRB 2	YES	YES
MRB 3	YES	YES
MRB4	YES	YES
MRB5	YES	YES

Table 3. Characterization of rule's circumstance measures

So, we can characterise both metrics above the level of the ordinal scale assuming the modified relation of belief.

#### 4. EMPIRICAL VALIDATION

As Basili points out, software engineering is a laboratory science, but there is insufficient analysis and experimentation (Basili, 1998). Typically experimentation is looking for a relationship between two variables, such as process and product characteristics, and may be performed for many purposes: to study process effects, product characteristics, environmental constraints (cost or schedule). Besides, it is necessary to combine experiments to build a body of knowledge that is useful to the discipline.

In this section we summarise an experiment done in order to validate DRT and RD metrics. This empirical validation has been carried out following the experimental method applied to software engineering (Pfleeger, 1995) (Bourque and Côté 1991). Our purpose is to prove that the metrics DRT and RD can be used for measuring the complexity of the relational database schema which influences in its understandability.

##### 4.1 Hypotheses

The formal hypotheses are:

1. Null hypothesis: Different values of metrics do not affect the comprehension of the database schema.
2. Alternative hypothesis 1: The value of the DRT metric affects the comprehension of the database schema.
3. Alternative hypothesis 2: The value of the RD metric affects the comprehension of the database schema.
4. Alternative hypothesis 3: The combination of the DRT and RD metrics affects the comprehension of the database schema.

##### 4.2 Experimental design

Each level of one factor appears with each level of the other one; therefore, we have selected the crossing design. This crossing relationship is defined  $A \times B$ . For us, A is the RD metric and B is the DRT metric (see Table 4). There were two possible values for RD metric (eight or five) and for DRT metric (two or five). In order to increasing the power of the test,  $\alpha$  has been set to 0.1 instead of the more common level of 0,05 (Briand et al, 1997).

		Factor B (RD)	
		LOW	HIGH
Factor (DRT) A	LOW	2,5	2,8
	HIGH	5,5	5,8

Table 4. Crossed Design for the experiment



### 4.3 Experimental procedure

The complete documentation was given to each subject and contained all the material related to the four tests: relational schemata, tables with data and the question/answer paper.

Before we started, the following aspects of the experiment were explained, what kind of exercises were to be developed, the material given, how to answer the questions made, and how long each test lasted. When the time of each test finished (ten minutes), the subjects were informed and, immediately, they could change to another test.

### 4.4 Subjects

The participants in the experiment are Computer Science students at the University of Castilla-La Mancha (Spain), who were enrolled in the databases course lasting two semesters. Until the day of the experiment, the students did not know that they were going to do it. The experiment was developed by 60 students but only 59 were finally accepted.

We have tried to minimise variability among participants by choosing people of the same degree, the third one (the last in Computer Science BSc). Effects of irrelevant variables were minimised making the same tests for all the subjects with the same duration (ten minutes per test).

### 4.5 Experimental materials

The documentation accompanying each design was consisted of approximately seven pages long and included the schema database, the tables with their rows and the question/answer paper. For each design the database schema had six tables.

The subjects were asked to perform three tasks with the values of the database schema: insert, delete and update. Figure 2 shows the question/answer paper.

Before the subjects took the test, the experiment was conducted on a small set of people.

1. What tables and how many rows in each table are affected if we delete in the Table 5 the row with cod1=210?					
Table 1	Table 2	Table 3	Table 4	Table 5	Table 6
2. What tables and how many rows in each table are affected if we update the column X of the row with cod2=11 in the table 3?					
Table 1	Table 2	Table 3	Table 4	Table 5	Table 6
3. What tables and how many rows and columns are necessary to add if we want add a new row in the table 4? (Suppose that all the necessary data are news in the database)					
Table 1	Table 2	Table 3	Table 4	Table 5	Table 6

Figure 2. Question/answer paper

### 4.6 Experimental results

There are three major items to consider when choosing the analysis techniques: the nature of the data collected, why the experiment is performed and the type of experimental design used ([29]).

Due to the type of the experiment used, F statistic is the most appropriate technique for obtaining the results (Rohatgi, 1976). Table 5 shows the results for the F-statistic:

Source of Variation	Q	Degrees of Freedom	$S_i^2$	F-Ratio
DRT	18.457	1	18.5	1.67
RD	531.000	1	531	48.1
Interaction	31.339	1	31.3	2.84
Error	2560.304	1	11.0	
Total	3141.102	232		

**Table 5. Results of the F-statistic**

By comparing these values with  $F_{1,232}=2.73$ , we can ensure that:

1. Alternative Hypothesis 1. Since  $1.67 < 2.73$ , DRT does not affect the results of the experiment. Therefore, alternative hypothesis 1 is invalid because the value of the DRT metric does not affect the results obtained.
2. Alternative Hypothesis 2. Since  $48.1 > 2.73$ , RD affects the results of the experiment. As a consequence, alternative hypothesis 2 is valid because the value of the RD metric affects the results obtained.
3. Alternative Hypothesis 3. Since  $2.84 > 2.73$ , the interaction of the metrics affects the results of the experiment. As a result, alternative hypothesis 3 is valid because the combination of the values of the DRT and the RD metrics affect the results obtained.

We can conclude by saying that the number of foreign keys in a relational database schema is a more solid indicator of its understandability than the length of the referential paths and that the length of the referential tree is not relevant by itself, but can modulate the effect of the number of foreign keys.

Following (Schneidewind, 1997) RD can be classified as a dominant metric, and DRT is not needed to classify quality (is a redundant metric).

When we obtained the results of the experiment one, some questions arose: Is there another metric useful for measuring of understandability?, can we obtain a more powerful method for measuring understandability maybe with another of our metrics and the RD?. In order to answer these questions, we have designed another experiment, making the RD metric fix and working with another metric D also related to the referential integrity, that we have defined as:

Then, in the second experiment we worked with the D and the RD metrics in order to obtain another relation between these two metrics and the understandability of a relational database schema. All the process developed was similar to the other experiment and the conclusions obtained were that the sum of the square of the number of foreign keys per table in a relational database schema is not an indicator of its understandability and cannot modulate the effect of the number of foreign keys.

## 5. CONCLUSIONS AND FUTURE WORK

More research is needed into the aspects of software measurement (Neil, 1994) both from a theoretical and from a practical points of view (Glass, 1996). We think it is very interesting to dispose on metrics for relational databases. These metrics can be used to flag outlying schemata for special attention, a strong requirement for low testing and maintenance costs would argue for justify extra managerial attention to a quite significant fraction of the relational database schemata (Chidamber, Darcy and Kemerer, 1998).

We have put forward different measures related with referential integrity in order to measure the complexity of the relational database schemata. These metrics were developed and characterised in accordance with a set of sound measurement principles.

Relational databases measures assume, as object-oriented measures ([37]), more complex properties related to concatenation operation than classic measures. These measures do not assume an extensive structure but can be characterised above the ordinal scale by fulfilling all the properties of modified relations of belief.

We have done some experiments to validate the proposed metrics, but more others are being developed at this moment. However the controlled experiments have problems (such as the large number of variables that cause differences, or the fact that these experiments deal with low level issues, microcosms of reality and small sets of variables) and limits (e.g. they do not scale up, are done in a class in training situations, are made in vitro and face a variety of threats of validity). Therefore, it is convenient to run multiple studies, mixing controlled experiments and case studies (Basili, 1998). For these reasons, a deeper empirical evaluation is under way in collaboration with industrial and public organisations in "real" situations.

We are also adapting our framework to several RDBMS that followed the SQL2 model, simplifying the measures to be taken. In this way, we will be able to give a more precise guide to the metrics usage (Churcher and Shepperd 1995) and develop a tool for automatic metrics collection.

The metrics for relational databases are complemented with others for active databases (Piattini and Díaz, 1999) and object-relational databases (Piattini, Calero, Polo, M. and Ruiz, 1998).

Other interesting future research implies the use of these (and other) metrics for building prediction systems for database projects. In [22] a prediction system based on the number of entities, attributes and relationship of an E/R schema combined with other 4GL-oriented measures (number of menus, edit screens, reports, etc.) is shown.

## REFERENCES

- Basili, V. R. (1998). Using experiments to build a body of knowledge. Madrid. España.
- Batra, D., Hoffer, J.A. and Bostrom, R.P. (1990). A comparison of user performance between the relational and the extended entity relationship models in the discovery phase of database design. *CACM*, 33 (2): 126-139.
- Bourque P., Côte V. (1991). An experiment in Software Sizing with structured analysis metrics. *J. Systems Software*. 15:159-172.
- Briand L., Bunse C., Daly J., Differing C., (1997). An experimental comparison of the Maintainability of object-oriented and structured design documents. *Proc. Int. Conf. on Software Maintenance*, Harrold MJ, Vissagid G (Eds), Bari, 1-3 Oct, 130-138.
- Briand, L.C., Differding, C.M. and Rombach, D. (1996). Practical Guidelines for Measurement-Based Process Improvement. *Software Process-Improvement and Practice 2*, 253-280.
- Briand, L.C., Morasca, S. And Basili, V. (1996). Property-based software engineering measurement. *IEEE Transactions on Software Engineering*, 22(1): 68-85.
- Card, D.N. and Glass, R.L. (1990). *Measuring Software Design Quality*. Englewood Cliffs. USA.
- Chidamber, S.R., Darcy, D.P. and Kemerer, C. (1998). Managerial Use of Metrics for Object-Oriented Software: An Exploratory Analysis. *IEEE Trans. On Software Engineering*, Vol. 24, N° 8, 629-639.

- Churcher, N.J. and Shepperd, M.J. (1995). Comments on "A Metrics Suite for Object-Oriented Design". *IEEE Trans. on Software Engineering*, 21(3): 263-265.
- Codd, E.F. (1970). A Relational Model of Data for Larged Shared Data Banks. *CACM*, 13 (6), 377-387.
- Date, C.J. (1995). *An Introduction to Database Systems*. 6th. Addison-Wesley, Reading, Massachusetts.
- Elmasri R, Navathe S. (1997). *Database Systems*. Second edition. Addison-Wesley. Massachussets
- Fenton, N. (1994). Software Measurement: A Necessary Scientific Basis. *IEEE Transactions on Software Engineering*, 20(3): 199-206.
- Fenton, N. and Pfleeger, S. L. (1997). *Software Metrics: A Rigorous Approach* 2nd. edition. London, Chapman & Hall.
- Glass, R. (1996). The Relationship Between Theory and Practice in Software Engineering. *IEEE Software*, November, 39 (11), 11-13.
- Gray R.H.M., Carey B.N., McGlynn N.A., Pengelly A.D., (1991), Design metrics for database systems. *BT Technology J*, Vol 9, 4 Oct, 69-79
- Henderson-Sellers, B. (1996). *Object-oriented Metrics - Measures of complexity*. Prentice-Hall, Upper Saddle River, New Jersey.
- Jarvenpaa, S. and Machesky, J. (1986). End user learning behavior in data analysis and data modeling tools. *Proc. of the 7th Int. Conf. on Information Systems*, San Diego: 152-167.
- Juhn, S. and Naumann, J. (1985). The effectiveness of data representation characteristics on user validation. *Proc. of the 6th Int. Conf. on Information Systems*, Indianapolis: 212-226.
- Kim, Y.-G. and March, S.T. (1995). Comparing Data Modeling Formalisms. *CACM* 38(6): 103-115.
- Li, H.F. and Chen, W.K. (1987). An empirical study of software metrics. *IEEE Trans. on Software Engineering*, 13 (6): 679-708.
- MacDonell, S.G., Shepperd, M.J. and Sallis, P.J. (1997). Metrics for Database Systems: An Empirical Study. *Proc. Fourth International Software Metrics Symposium - Metrics'97*, Albuquerque. *IEEE Computer Society*, pp. 99-107.
- McCabe, T.J. (1976). A complexity measure. *IEEE Trans. Software Engineering* 2(5): 308-320.
- Melton, A. (ed.) (1996). *Software Measurement*. London, International Thomson Computer Press.
- Neil, M. (1994) Measurement as an Alternative to Bureaucracy for the Achievement of Software Quality. *Software Quality Journal* 3 (2), 65-78.
- Piattini, M., Calero, C., Polo, M. and Ruiz, F. (1998). Maintainability in Object-Relational Databases. . *Proc of The European Software Measurement Conference FESMA 98*, Antwerp, May 6-8, Coombes, Van Huysduynen and Peeters (eds.), 223-230.
- Piattini, M. and Díaz, O. (1999). Metrics for active databases maintainability. Accepted in *CAISE'99*. Heidelberg, June 16-18.
- Pfleeger S.L. (1995). Experimental design and analysis in software engineering. *Annals of Software Engineering*, 219-253, JC Baltzer AG, Science Publishers.
- Pfleeger, S. L. (1997). "Assessing Software Measurement". *IEEE Software*. March/April, pp. 25-26.
- Rohatgi VK (1976). *An introduction to probability theory and mathematical statistics*. Wiley Series in Probability and mathematical statistics.
- Rossi, M. and Brinkkemper, S. (1996). Complexity Metrics for Systems Development Methods and Techniques. *Information Systems* 21(2): 209-227.
- Schneidewind N.F. (1997). Software metrics for quality control. *Proceedings of the fourth*

- international software metrics symposium. IEEE Computer Society Technical Council on Software Engineering, 127-136
- Shoval, P. and Even-Chaime, M. (1987). Database schema design: An experimental comparison between normalization and information analysis. *Database* 18(3): 30-39.
- Sneed, H.M. and Foshag, O. (1998). Measuring Legacy Database Structures. Proc of The European Software Measurement Conference FESMA 98, Antwerp, May 6-8, Coombes, Van Huysduynen and Peeters (eds.), 199-211.
- Weyuker, E.J. (1988). Evaluating software complexity measures. *IEEE Transactions on Software Engineering* 14(9):1357-1365.
37. Zuse, H. (1998). *A Framework of Software Measurement*. Berlin, Walter de Gruyter.