



XIX Congreso Argentino de Ciencias de la Computación



Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension (MODELS 2011)

G. Bavota¹, C. Gravino¹, R. Oliveto², A. De Lucia¹, G. Tortora¹, **M. Genero**³, J.A. Cruz-Lemus³

¹University of Salerno (Italy)
²University of Molise (Italy)
³University of Castilla-La Mancha (Spain)




Contents

- Introduction
- Related Work
- Empirical Evaluation
- Discussion and Threats to Validity
- Conclusion and Future Work

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011

2

Introduction



DATA MODEL

set of concepts that can be used to describe both the structure of and the operations on a database

COMPREHENSIVE NOTATION

desirable to understand and interpret data models, avoid misunderstanding, facilitate comprehension during maintenance


CURRENT SITUATION

E/R	VS	UML	
still remains the <i>de facto</i> standard for DB conceptual modeling		becoming the <i>de facto</i> standard for the analysis and design of software systems	

3

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011

Introduction



Several empirical comparisons conclude:

The support given by UML class diagrams in comprehension tasks is at least equal (or even higher) than the support given by ER diagrams.

Still missing:


Qualitative and quantitative analyses concerning the identification of the graphical elements of one notation that are more comprehensible than the corresponding element in the other notation.

These analyses could be used to:

- i. justify the need of preferring ER or UML CDs for data modeling
- ii. identify weaknesses in each notation

4

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011




Introduction

The **main goal** of this paper is to present a controlled experiment and two replications:

- Analyze the support given by ER and UML CDs for data models comprehension.
- Perform a fine-grained analysis:
 - Single building blocks:** Entity, Primary Key/ID, Composite Attribute, Multi-value Attribute, Recursive relationship, Relationship cardinality, Ternary relationship, Generalization IS-A, Weak entity, M:N relationships

5

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Introduction

Material available at : <http://sesa.dmi.unisa.it/UMLvsER.html>

UML vs ER - Windows Internet Explorer

http://www.dmi.unisa.it/UMLvsER.html

ARCHIVO ESQUEMA ESCRIBIRIA GEMERIZACION AYUDA

Windows Live Inicio Favoritos Herramientas Ayuda

Google

UMLvsER

IDENTIFYING THE WEAKNESSES OF UML CLASS DIAGRAMS DURING DATA MODEL COMPREHENSION THROUGH CONTROLLED EXPERIMENTS

ABSTRACT

Several studies have been carried out to analyze and compare the support given by ER diagrams and UML class diagrams in comprehension tasks on data models. In this paper we present a set of experiments aimed at comparing the support provided by ER and UML class diagrams during comprehension activities by focusing on the single building blocks of the two notations. This kind of analysis can be used to identify weaknesses in a notation and/or justify the need of transferring ER to UML for data modeling. The results reveal that UML class diagrams are generally more comprehensible than ER diagrams, even if the former has some weaknesses related to those building blocks. These findings suggest that a UML class diagram extension should be considered to overcome these weaknesses and further improve the comprehensibility of the notation.

Downloads

- The raw data of our experimentation: RawData.zip (12 Kb)
- The questionnaires used in our case study: Questionnaires.zip (942 Kb)

AUTHORS

Gabriele Bavota
University of Salerno
gbavota@unisa.it

Carmine Granata
University of Salerno
granata@unisa.it

Rocco Oliveto
University of Molise
rocco.oliveto@unimol.it


Andrea De Lucia
University of Salerno
adlucia@unisa.it

Marcela Genero
University of Castilla
marcela.genero@uclm.es

José A. Cruz-Lemus
University of Castilla
josea.cruz@uclm.es

6

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Related Work

Shoval and Shiran (1997)

EER vs OO models from the point of view of design quality. Similar results, but EER models are better in ternary and 1:1 relationships. Designers prefer EER.

Shoval and Frumermann (1994)

EER vs OO models related to user comprehension. EER is more comprehensible for ternary relationship. No significant differences in the other constructs.

Bock and Ryan (1993)


EER vs OO models from a designer perspective. Significant differences only a few cases (1:1, M:N relationships) and no differences in time.

Palvia et al. (1992)

EER vs OO models from an end-user perspective. They measure the comprehension on overall term. The results suggests that OO schemas are more comprehensible.

7

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Empirical Evaluation

Definition

Goal


To analyse whether UML CDs are more comprehensible than ER diagrams during the comprehension of data models.

Fine-grained analysis comparing:

Entity, Primary Key/ID, Composite Attribute, Multi-value Attribute, Recursive relationship, Relationship cardinality, Ternary relationship, Generalization IS-A, Weak entity, M:N relationships

8

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011




Empirical Evaluation
Planning: Context

156 subjects from Salerno

fresher subjects. (n=67 in 2nd replication)
bachelor subjects. (n=37 in experiment)
master students, (n=52 in 1st replication)

9

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Empirical Evaluation
Planning: Context


Data models:

- Company, managing projects conducted by a company.
- EasyClinic, managing a medical doctor's office.

System	# entities	# attributes	# relationships
Company	7	17	5
EasyClinic-BookingManagement	6	18	5

10

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Empirical Evaluation

Planning: Variable Selection

Independent variable Method, i.e., the notation used (ER or UML)

Dependent variable Comprehension level, measured by the recall and precision measures

$$F - measure_s = 2 * \frac{precision_s * recall_s}{precision_s + recall_s} \%$$


$$recall_s = \frac{\sum_i |answer_{s,i} \cap correct_i|}{\sum_i |correct_i|} \% \quad precision_s = \frac{\sum_i |answer_{s,i} \cap correct_i|}{\sum_i |answer_{s,i}|} \%$$

Co-factors

- System
- Lab session

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011

11



Empirical Evaluation

Planning: Hypotheses

Hypothesis (main factor):


H1₀: There is no difference in the **comprehension level** of subjects when they have to comprehend **ER** or **CD** diagrams.
(we expect to reject the null hypothesis)

Hypothesis (co-factors):

H2₀: There **system/lab** session does not affect the **comprehension level** of subjects when they have to comprehend ER or CD diagrams.
(It is desirable not to reject the null hypotheses)

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011

12




Empirical Evaluation
Planning: Design

Group	Treatment	
	ER	UML
A	EasyClinic, Lab1	Company, Lab2
B	Company, Lab2	EasyClinic, Lab1
C	Company, Lab1	EasyClinic, Lab2
D	EasyClinic, Lab2	Company, Lab1

13

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Empirical Evaluation
Planning: Tasks

- **Questionnaire 1:**
 - 10 multiple choice questions.
 - One or more correct answers.

Q4 Let us focus on the classes Project and Company.
Which of the following statements is true:

A company has a unique office


A project has a unique office

A company may have multiple offices

- **Questionnaire 2:**
 - A subjective questionnaire to specify preferences between notations.

14

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Empirical Evaluation

Planning: Tasks

Final Questionnaire

- Assess the adequacy of time and the clearness and difficulty of the comprehension tasks.


Questions

- S1 : I had enough time to perform the tasks
- S2 : The task objectives were perfectly clear to me
- S3 : The tasks I performed was perfectly clear to me
- S4 : Judging the difficulty of the comprehension task

S1, S2, S3 → 1 (strongly disagree) to 5 (strongly agree)
S4 → 1 (very low) to 5 (very high).

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011

15



Empirical Evaluation

Planning: Data Analysis Procedure

Behavior of the dependent variable

Descriptive statistics

Differences between treatments (H1)

Paired Wilcoxon one-tailed test, each subject performed a task on two different models with the two possible treatments.

Co-factors effect (H2)

ANOVA

Preferences between treatments


Histograms

Final questionnaire analysis

Box-plots

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011

16




Empirical Evaluation

Operation: Execution

- **Training** → on ER and CDs and on experiment execution.
- **Execution** → Tasks performed individually supervised by an instructor.

17

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Empirical Evaluation

Operation: Data Analysis

Comprehension level (descriptive statistics)

Subjects	ER			CD		
	Mean	Median	St. Dev.	Mean	Median	St. Dev.
Fresher	0.801	1.000	0.307	0.816	1.000	0.280
Bachelor	0.849	1.000	0.242	0.845	1.000	0.278
Master	0.849	1.000	0.277	0.838	1.000	0.272

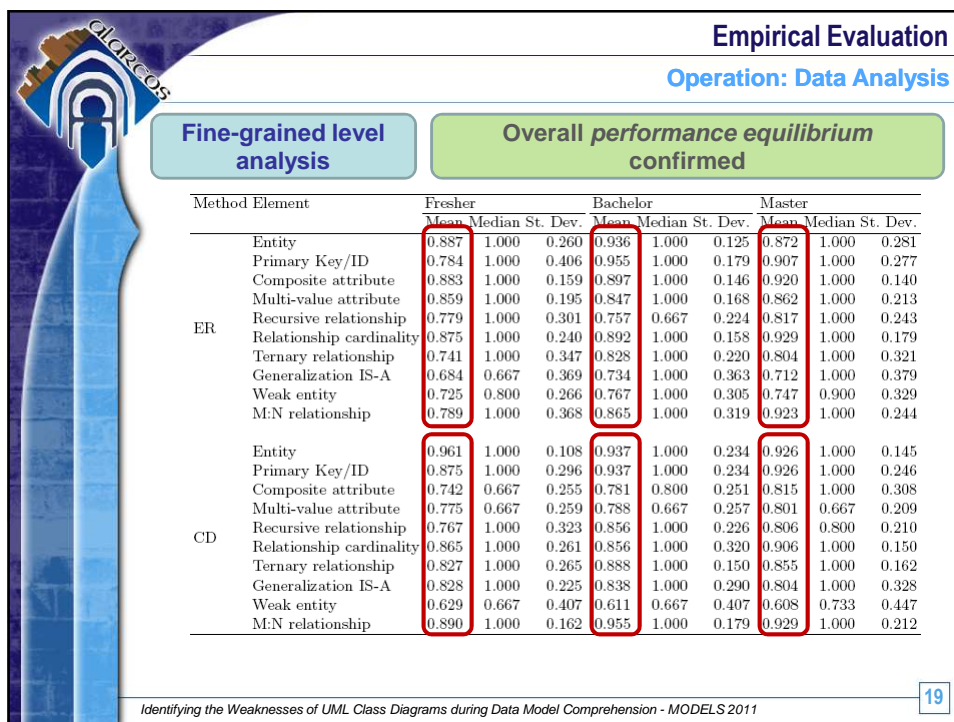
Both notations **provide comparable support** when performing comprehension activities on data models

Wilcoxon test

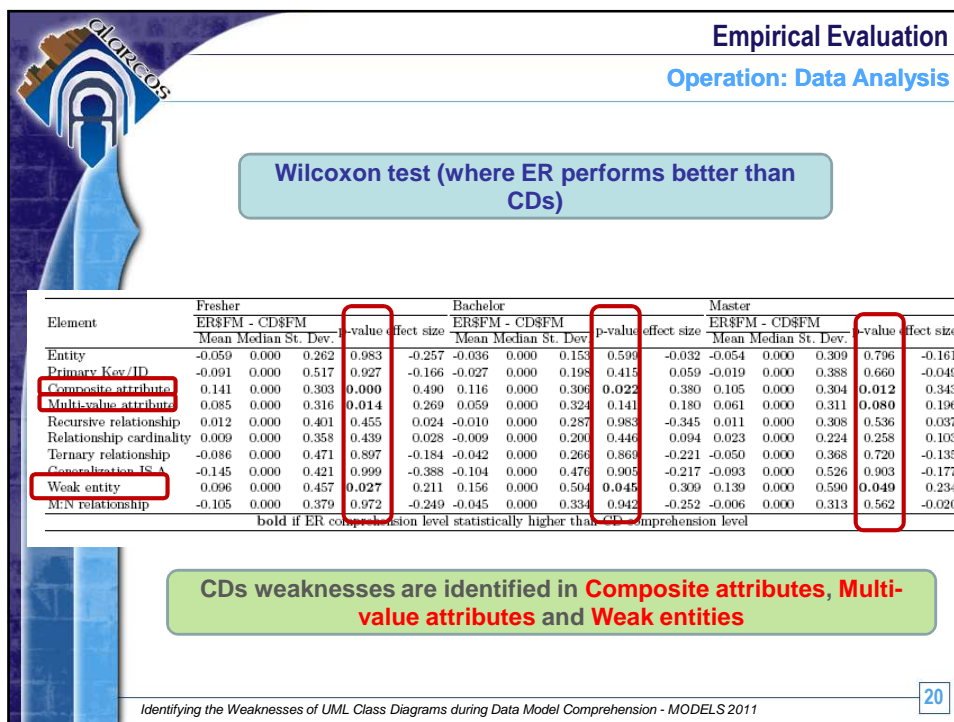
NO significant differences

Subjects	CD\$FM - ER\$FM			p-value	effect size
	Mean	Median	St. Dev.		
Fresher	0.014	0.000	0.404	0.343	0.037
Bachelor	0.003	0.000	0.330	0.420	-0.011
Master	-0.012	0.000	0.383	0.817	-0.030


18



Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011



Empirical Evaluation

Operation: Data Analysis

Comprehension level removing the identified weaknesses (Wilcoxon test)


Subjects	CD\$FM - ER\$FM			p-value	effect size
	Mean	Median	St. Dev.		
Fresher	0.066	0.000	0.410	0.000	0.161
Bachelor	0.052	0.000	0.290	0.010	0.120
Master	0.027	0.000	0.358	0.096	0.074

bold if CD comprehension level statistically higher than ER comprehension level

Using **CDs** implies obtaining a **better comprehension level**

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011

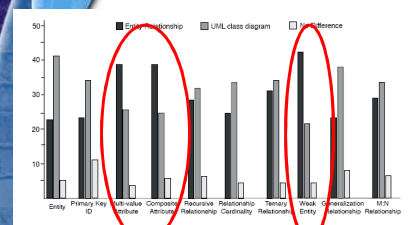
21



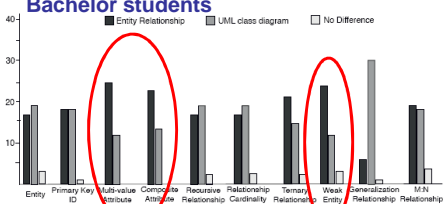
Empirical Evaluation

Qualitative comparison

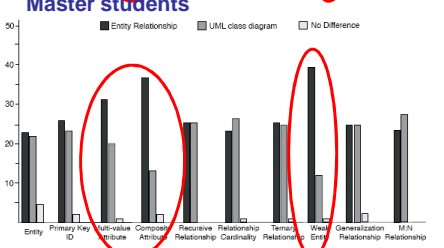
Fresher students



Bachelor students



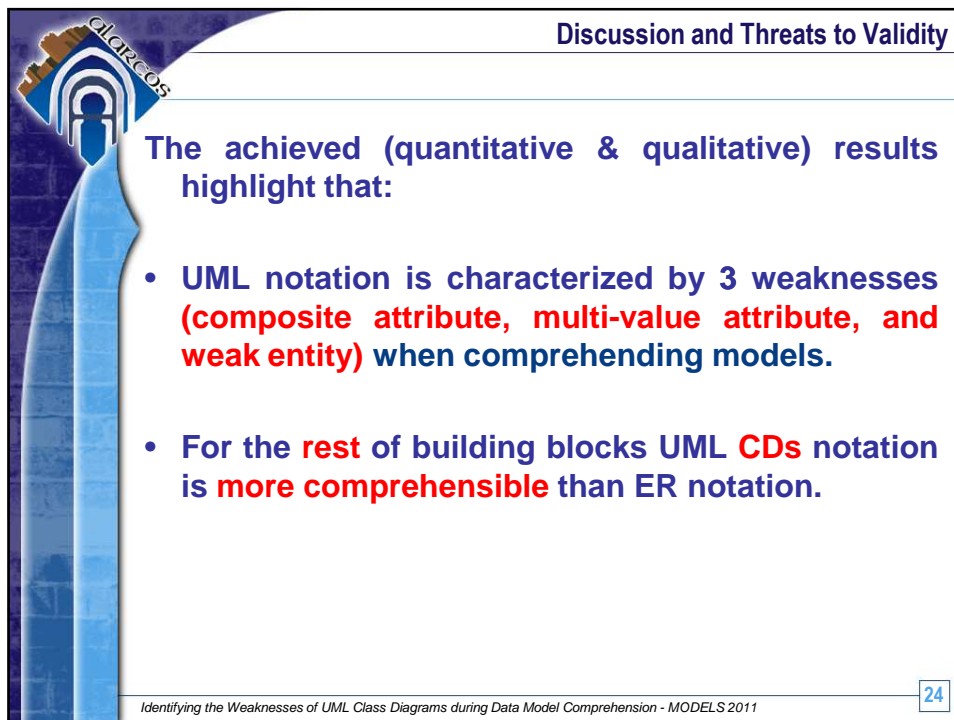
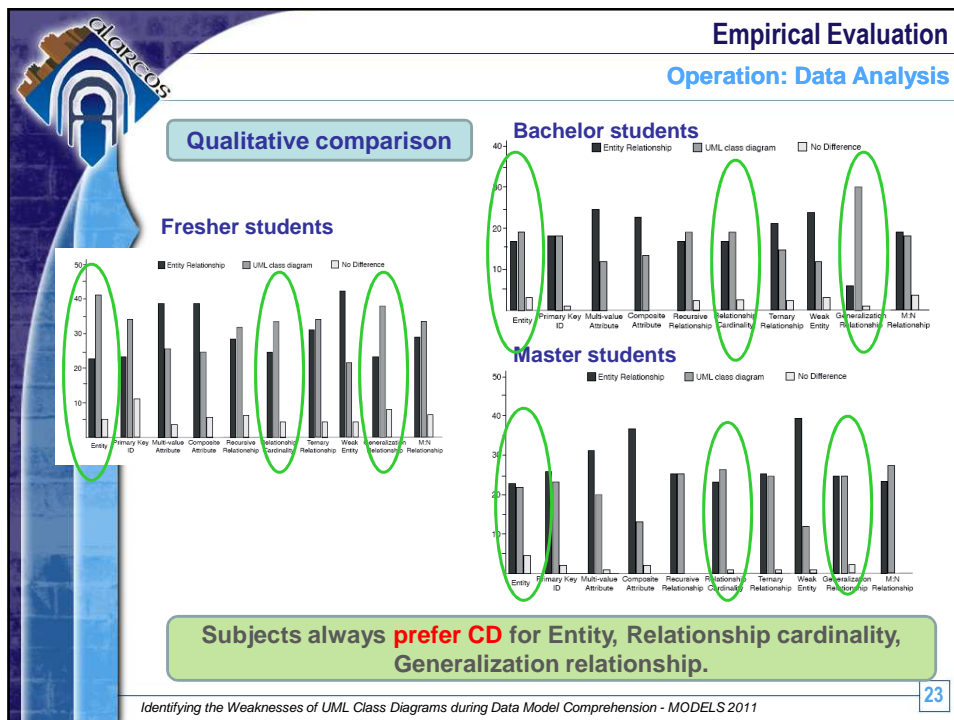
Master students



Subjects always **prefer ER** in the identified CDs weaknesses

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011

22



Discussion and Threats to Validity

- **Only comprehension** was studied, more properties (e.g. analyzability) to be evaluated in the future.
- Subjects had **enough time** to carry out the tasks (S1), the **objectives and tasks** to perform were **clear** (S2, S3), and they experienced no particular **difficulties** (S4).

25

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011

Discussion and Threats to Validity


- **Co-factors: no significant influence.**

Factor	Fresher	Bachelor	Master	All
Lab	No (0.787)	No (0.163)	No (0.175)	No (0.216)
System	No (0.793)	No (0.636)	No (0.113)	No (0.229)
Method vs Lab	No (0.817)	No (0.833)	No (0.305)	No (0.439)
Method vs System	No (0.793)	No (0.817)	No (0.618)	No (0.679)

- **Future replication** with practitioners.
- Also planned a **survey to confirm** if the identified elements are actual weaknesses and, if so, why.
- Future work devoted to compare the notations on **realistically sized artifacts**.

26

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011




Conclusions and Future Work

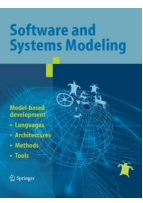
- We carried out → 1 controlled experiment and 2 replications aimed at analyzing the support given by ER and UML CDs for data models comprehension.
- Main findings:
 - UML CDs are generally more comprehensible (confirming previous studies' conclusions).
 - UML CDs weaknesses → weak entity, composite and multi-value attributes.
- Future work:
 - Replication in different contexts, with different subjects and objects.
 - Exploit stereotypes to extend UML CDs and solve these weaknesses.

27

Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension - MODELS 2011




Extended version




Bavota, G., Gravino, C., Oliveto, R., De Lucia, A., Tortora, G., Genero, M., Cruz-Lemus, J. A. **A fine-grained analysis of the support provided by UML class diagrams and ER diagrams during data model maintenance.**
 Software and System Modeling. DOI 10.1007/s10270-012-0312-6.

28




CACIC2013
XIX Congreso Argentino de Ciencias de la Computación



Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension (MODELS 2011)

G. Bavota¹, C. Gravino¹, R. Oliveto², A. De Lucia¹,
G. Tortora¹, **M. Genero**³, J.A. Cruz-Lemus³



¹University of Salerno (Italy)
²University of Molise (Italy)
³University of Castilla-La Mancha (Spain)