# SECONDARY EDUCATION STUDENTS' DIFFICULTIES ON DATABASE DESIGN AND REMEDIAL TEACHING STRATEGIES

G. FESSAKIS AND A. DIMITRACOPOULOU

*Learning Technology and Educational Engineering Laboratory, Department of Education,*
*University of Aegean, Av. Demokratias, 85100, Rhodes, GREECE.*
*E-mail: [gfesakis,adimitr]@rhodes.aegean.gr*

C. HALATSIS

*Department of Informatics & Telecommunications,*
*University of Athens, Panepistimiopolis Ilissia, 15771, Athens, GREECE.*
*E-mail: halatsis@di.uoa.gr*

Databases use and design has utilitarian and pedagogical educational interest. The utilitarian interest for database design education is determined by their economic impact. Databases pedagogical interest rises from their view as models and the possibility to use them for the design of general knowledge learning activities. The utilitarian and learning advantages rationalize the database design instruction in secondary education rising educational research questions. In this paper we report some key research findings about 11th grade students' difficulties in database design, focusing especially in the representation of relationships. For the clarification of students' difficulties research activities were designed and implemented that asked students to transform given ER schemata to the corresponding logical and vice-versa. The analysis of students' solutions resulted in a systematic categorization of their difficulties and the proposition of specific action strategies for database design instruction improvement.

**TOPICS + KEYWORDS**: Computer science education, Database design, Data modeling, Teaching/learning strategies, Knowledge and skills in virtual environments, Active & interactive learning, Secondary Education

## 1 Introduction

Databases use and design has utilitarian and pedagogical educational interest. The utilitarian interest for databases education is determined by their economic impact. The increasing demand for database development in combination with the availability of user friendly data base management systems (dbms's) results to database design by untrained people raising quality and cost questions [2,3].

In addition we consider that pedagogical interest of databases rises from their view as models [8] and the possibility to use them for the design of general knowledge learning activities. In this consideration database design is a modeling activity and dbms's are general purpose modeling environments using data models as representation systems. Database design can be used for the development of group based, learning activities, concerning authentic problem solving that conserve students motivation.

The above utilitarian and learning advantages of databases rationalize the database design instruction in secondary education but the way that this could be better approached is still an open educational research question. Research reports about human factors in database design are rather limited [2,3,9] and concern higher education students and professionals. The differences in the background between the two target groups do not permit simple transfer of research results from higher to secondary education. Our research interests concern databases didactics and educational exploitation in general, for secondary education students.

In this paper we present an action research about difficulties secondary education students in database design using Entity-Relationships (ER) [5] and Relational [6] models for conceptual and logical design respectively. In the research 27, 11th class students participated from a public vocational school named 2nd TEE of Rhodes in Greece. Students were assigned the obligatory subject titled "databases". The researcher was their normal teacher for both subjects. After the instruction of the subject students were assigned small-scale database design projects. Students' difficulties in typical database design especially in the representation of relationships were the starting point of the research. For the clarification of the starting point we designed and implemented research activities that asked students to transform given ER schemata to the corresponding logical and vice-versa. The research does not aim to reproduce some of the many known critics for ER [9] but to propose improvements for the instruction of data base design in secondary education using educational research. In the following sections, the research data analysis is presented along with proposed action strategies for instruction and learning improvement.

## 2 Logical level "Relationships" interpretation.

In this activity students were asked to produce ER schemata for given relational. This process is not usually a teaching subject. The transformation of logical to conceptual schemata is expected to produce rich information about students' understanding of the related concepts because they are going to activate their understanding of the subject. The logical schemata provided to students are organized and presented at table 1. Primary keys are formatted bold and underline. External keys have the same name with the corresponding primary keys. Short verbal descriptions specified schemata meaning to students. Students worked alone for 90 minutes maximum.

**Table 1.** Logical schemata given to students to produce corresponding conceptual ones.

| | |
|---|---|
| **C1. Single entity schema.** | |
| C1S1 | **T1:** SHOP(**Name**, Address, Telephone, BossName) |
| **C2. Three relations schema for to entities and a binary relationship.** | |
| C2S1 | **T1:** WAREHOUSE(**wCode**, Address), **T2:** PRODUCT(**pCode**, Description), **T3:** EXIST_IN(**wCode**, pCode, Quantity, Position) |
| C2S2 | **T1:** NEWSPAPER(**Name**, Owner, Telephone), **T2:** ANNOUNCEMENT(**aCode**, Client, Text, Category), **T3:** PUBLISH(**Name, aCode, Date**, Page) |
| C2S3 | **T1:** CAR(**cCode**, Model), **T2:** SPARE_PART(**pCode**, Description), **T3:** USES(**cCode**, **pCode**) |
| C2S4 | **T1:** STUDENT(**sCode**, Name), **T2:** SUBJECT(**Title**, Kind), **T3:**EXAM(**sCode**, **Title**, **Date**, Time) |
| **C3. Two relations schema for a recursive relationship.** | |
| C3S1 | **T1:** EMPLOY(**ID**, FirstName, SurName, Telephone, Position), **T2:** MARRIED(**Hasband_ID**, **Wife_ID**) |
| **C4. Four relations schema for three entities and a ternary relationship.** | |
| C4S1 | **T1:** REFEREE(**ID**, Name) TEAM(**Name**, Home), **T2:** STADIUM(**Stadium_Name**, Address), **T3:** GAME(**ID**, **HomeName**, **GuestName**, **Date**, Time) |

### *2.1 Solutions analysis*

Students' solutions could be classified in the following categories:

**Category C1 - «Correct».** C1 contains all the correct solutions. The correctness of cardinality is not evaluated in this research because of the basic difficulties found in the understanding of relationships.

**Category C2 - «Attaching relationship properties to entities».** A small percentage of students produced ER schemata with relationship properties attached to entities. Some students recognize relationships but they do not "like" them to have properties like entities.

**Category C3 - «Syntactical solutions».** In this category students propose an entity for each relation of the logical schema and connect them with 'artificial' relationships in order to get the conceptual schema readable as a natural language sentence. We call this kind of solution "syntactical". Students giving syntactical solutions are concerning ER schemata rather as concept maps where relationships are more informal and arbitrary. Students can be conscious for this kind of mistake when they try to produce the corresponding relational schema of a syntactical one. The group of syntactic solutions is the most populated.

**Category C4 - «Ignoring relationships».** Students in this category produce ER schemata without relationships, designing an entity for each relation. These students appear to be unaware of the relationship concept and its representation.

**Category C5 - «Unclassified».** This group contains solutions that could not join any of C1 thought C4 groups. Solutions of this kind use arbitrary entity and/or characteristic names etc.

### 2.2 *Summary of solutions' analysis*

Table 2 presents the categorical distribution of solutions for each problem and in total. Column labelled 'N.S' presents the number of students that did not give a solution. **Observing column C1** we find evidence that understanding of recursive binary (C3S1) as well as ternary relationships (C4S1) is more difficult for students. Furthermore, problem seems to effect on performance for binary relationships since the percentage of correct solutions for problems C2S1- C2S4 varies. **Observing column C3** we can see that most students treat relationships syntactically. In addition when the difficulty increases students give less correct and more syntactical solutions. Some students may be in a transient level of relationships and backtrack when difficulty increases. **Observing column C4.** As in the previous case when difficulty increases more students ignore relationships. It seems that ignoring relationships is a first level for relationships understanding where students backtracked when the difficult problem of recursive relationship occurred.

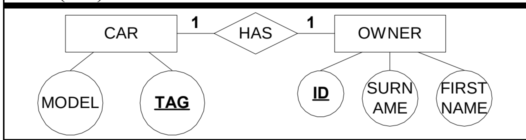
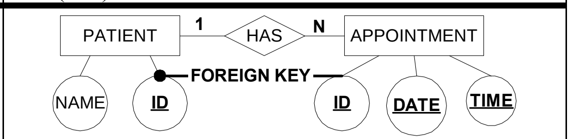
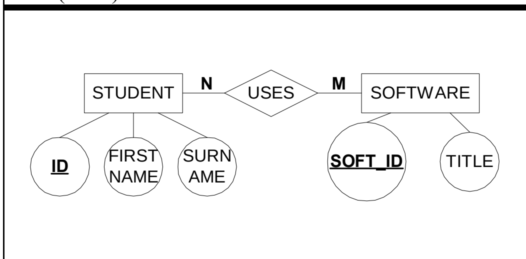
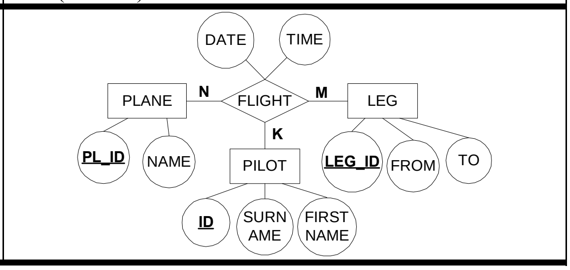
**Table 2.** Categorical Distribution of solutions for each problem and totally.

| | C1 | % | C2 | % | C3 | % | C4 | % | C5 | % | N.S | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **C1S1** | 24 | **88,89** | 0 | **0,00** | 0 | **0,00** | 0 | **0,00** | 3 | **11,11** | 0 | **0,00** |
| **C2S1** | 6 | **22,22** | 2 | **7,41** | 15 | **55,56** | 2 | **7,41** | 2 | **7,41** | 0 | **0,00** |
| **C2S2** | 13 | **48,15** | 0 | **0,00** | 11 | **40,74** | 2 | **7,41** | 0 | **0,00** | 1 | **3,70** |
| **C2S3** | 13 | **48,15** | 2 | **7,41** | 8 | **29,63** | 2 | **7,41** | 1 | **3,70** | 1 | **3,70** |
| **C2S4** | 9 | **33,33** | 0 | **0,00** | 13 | **48,15** | 2 | **7,41** | 0 | **0,00** | 3 | **11,11** |
| **C3S1** | 0 | **0,00** | 0 | **0,00** | 22 | **81,48** | 4 | **14,81** | 1 | **3,70** | 0 | **0,00** |
| **C4S1** | 1 | **3,70** | 1 | **3,70** | 20 | **74,07** | 1 | **3,70** | 0 | **0,00** | 4 | **14,81** |
| **TOTAL** | 66 | **34,92** | 5 | **2,65** | 89 | **47,09** | 13 | **6,88** | 7 | **3,70** | 9 | **4,76** |

## 3 Conceptual level "Relationships" interpretation.

In the second activity students were asked to produce Relational from ER schemata. Table 3 shows the problems given to students with the form of ER schemata categorized according to the kind of the relationships they include. Students have been instructed a certain methodology for ER to Relational schema translation. The analysis of students' solutions is presented in the following paragraphs.

**Table 3.** ER schemata given to students for transformation to relational.

| **P1.** (1-1) BINARY RELATIONSHIP SCHEMA | **P2.** (1-N) BINARY RELATIONSHIP SCHEMA |
|---|---|
|  |  |
| **P3.** (N-M) BINARY RELATIONSHIP SCHEMA | **P4.** (N-M-K) TERNARY RELATIONSHIP SCHEMA |
|  |  |









### 3.1 *Solutions analysis*

Students' solutions are of the following categories of decreasing ability to relationships' representation:

**Category C1 - «Correct».** Solutions of this category contain tables with proper primary and foreign keys for the representation of entities and relationships.

**Category C2 - «inadequate relationship representation».** Solutions of this kind propose relationship representation with minor or more significant errors. Some typical errors concern addition of arbitrary fields and/or elimination of others etc.

**Category C3 - «Ignoring relationship».** Solutions in this category contain a table for each entity without any foreign key and no representation for the relationship. Solutions of this kind represent a significant percent.

### *3.2 Summary of solutions analysis*

Table 4 presents the categorical distribution of solutions for each problem and in total. The column labeled 'N.S' presents the number of students that did not give a solution.

**Observing C1 (Correct solutions)** column it is obvious that students face increasing difficulties with ternary relationships. Students that solve P4 constitute a small group that understands the relationships. **Observing C3 (Ignoring relationships)** column it is interesting to analyze the 0% for P2. P2 problem ER schema has an explicit representation for the foreign key. Most correct solutions for P2 belong to students that systematically ignore relationships! These students produce a table for each entity and come up with correct solutions just by accident. Students that apply methodology rules in a rote manner produce a redundant table for P2 and do not review their solutions.

**Table 4.** Categorical Distribution of solutions for each problem and in total.

| | C1 | % | C2 | % | C3 | % | N.S | % |
|---|---|---|---|---|---|---|---|---|
| **P1** | 13 | **48.15** | 2 | **7.41** | 12 | **44.44** | 0 | **0** |
| **P2** | 14 | **51.85** | 13 | **48.15** | 0 | **0** | 0 | **0** |
| **P3** | 13 | **48.15** | 5 | **18.52** | 9 | **33.33** | 0 | **0** |
| **P4** | 4 | **14.81** | 11 | **40.74** | 9 | **33.33** | 3 | **11.11** |
| **TOTAL** | 44 | **40.74** | 31 | **28.70** | 30 | **27.78** | 3 | **2.78** |

## 4 Discussion

Databases design education in secondary education is interesting because of utilitarian and didactic reasons. The effective introduction of database design in secondary education needs thorough research. Combining the findings from both activities described previously it is possible to rationalize a set of action strategies for the improvement of database design learning by secondary education students:

**1. Most students treat relationships syntactically and use ER as a kind of conceptual map.**

The database design process as usually presented to students merges ontological analysis with conceptual design using ER model. The confrontation of domain understanding (recognition of the concepts-entities, their characteristics, etc) and the detailed and formal specification of the information needs of the problem simultaneously is considered a heavy duty. Thus, it is reasonable to propose the separation using concept maps for ontological analysis and a conceptual model for database design.

**2. Students that ignore relationships representation produce correct relational schemata from ER ones that explicit mention foreign keys.** In relational model relationship are implemented using foreign keys that are fields that work as references between tables. The representation of foreign keys in ER model is practically optional. This confuses the relational schema production. If the ER schema represents explicitly the foreign keys and there are only binary relationships without attributes, most students could produce a correct logical schema. The production relational schemata for given ER ones is important in order students to obtain feedback and review their designs. A didactically proper conceptual model should impose the foreign key representation.

**3. Students face difficulties in understanding relationships semantics and representation especially in the cases of recursive and ternary relationships.** Students need a more tangible relationship representation, for this purpose it is reasonable to propose the relationship concept introduction using a lower level representation as the tuple sets. In addition the understanding of relationship misconceptions could be based on feedback from the logical level according to normalization criteria. This observation recommends the automation of logical schema production for the conceptual (and vice versa) in order to get feedback as soon as possible for the meaning of their designs.

From the previous analysis we can conclude that a didactically proper Conceptual Model should: **1.** Permit the automatic conceptual to relational translation and vice versa in order to facilitate feedback. **2.** Use only binary relationships without attributes **3.** Impose the explicit representation of the foreign keys in the conceptual level and systematize their introduction to the conceptual schema reducing the problem of foreign key definition to a proper relationship selection decision.

Conceptual models that fulfill the above requirements include IDEF1X. IDEF1X is widely accepted for relational database design and is an official standard in USA [7]. A detailed presentation of IDEF1X is out of the purpose of the paper. For an illustration of IDEF1X didactical interest a small example is presented in Fig. 1 and 2. When the two entities of Fig. 1 are connected with a relationship most software tools add automatically the foreign key to the dependent entity which is marked with rounded squares. Consider a student that produces syntactical solutions. This student will come up with a correct solution or he/she through the automatic production of the corresponding database (logical level feedback) probably will find out soon that the proposed conceptual schema does not represent the problematic situation.

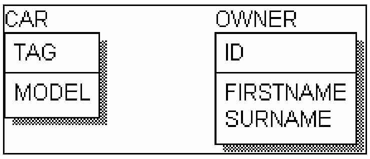



**Figure 1.** IDEF1X use. Entities before relationship definition

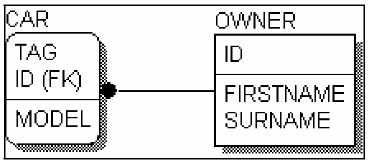



**Figure 2.** IDEF1X use. Entities after relationship definition

The presented work has been implemented in the framework of a phd research about database didactics in secondary education. The research is going to be continued with the validation of the action strategies proposed and the use of Multiple Correspondence Analysis [4] in order to find students groups with coherent solution strategies and to obtain a more evidence for the above mentioned key findings.

## References


1. Altrichter, H., Posch, P., and Somekh, B., Teachers investigate their work. An introduction to the methods of action research, *Routledge*, (1993)
2. Antony, S., and Batra, D., CODASYS: A Consulting Tool for Novice Database Designers, *ACM SIGMIS Database*, 33(3) (2002), pp. 54-68
3. Batra D., Hoffer J., and Bostrom R., Comparing Representations with Relational and EER Models, *Communications of the ACM*, 33(2) (1990)
4. Benzécri, J. P., Correspondence analysis handbook. *New York: Marcel Dekker*, (1992)
5. Chen, P., The Entity-Relationship model - toward a unified view of data, *ACM transactions on database systems*, 1(1) 1976, pp. 9-36
6. Codd, E., A Relational model of data for large shared data banks, *Communications of the ACM*, 13(6) 1970, pp. 377-387
7. Federal information Processing Standards Publication 184 . INTEGRATION DEFINITION FOR INFORMATION MODELING (IDEF1X), http://www.essentialstrategies.com/ publications/ modeling/ idef1x.htm, 1993
8. Fessakis, G., and Dimitracopoulou, A., Exploitation of data modeling for database design in secondary education learning activities: A case study concerning real stories analysis.**,** *Interactive Computer Aided Learning (ICL) 2003*, Carinthia Tech Institute, 24-26 Sep 2003 Villach, Austria
9. Hay C. D, A comparison of data modelling techniques, the database newsletter (1999 revision at http://www.essentialstrategies.com), 23(3) 1995