Exploitation of data modeling for database design in secondary education learning activities: A case study concerning real stories analysis

George Fessakis, Angelique Dimitracopoulou

EXPLOITATION OF DATA MODELLING FOR DATABASE DESIGN IN SECONDARY EDUCATION LEARNING ACTIVITIES: A CASE STUDY CONCERNING REAL STORIES ANALYSIS

George Fessakis[1], *Angelique Dimitracopoulou*[2].
[1,2]Learning Technology and Educational Engineering Laboratory,
Department of Education, University of Aegean, GR

Key words: *Data modelling, database design, story analyses*

Abstract:

Data models are representation systems for structural modelling of physical systems. Data models application results in databases that are amenable to computation enabling the modeller to interact with them for information retrieval, hypotheses testing, comparisons etc. In this paper we initially approach the didactic and learning value of data modelling for database design in terms of the features related to students' involvement in the related learning activities. In order to clarify the proposed analysis a case study is presented concerning the design of a database for the analysis of real stories about vehicular accidents. The case study results support the fundamental consideration that database design based learning activities offer opportunities for rich learning situations that activate high order thinking.

1 Introduction

Database educational interest has its own sources in supporting arguments from two main directions. The first direction concerns utilitarian arguments based on the economic significance of database technology, while the second concerns arguments about the value of databases as learning tools according to modern pedagogical approaches and learning theories. In this paper we initially formulate arguments from both directions about the educational interest of databases' design in secondary education and then we present a case study of real stories analysis by students using database design for the study of vehicular accidents. The case study supports the identification of the qualitative features of students' involvement in database design learning activities as they proposed in the first part of the paper. In addition the case study permits the examination of students' chances to improve their database use comprehension during the exploitation of databases in a general problem solving activity.

2 The educational interest of database design

2.1 Utilitarian interest for database education

Data processing systems are still of the most widely required application of information technology in organizations of any size. Databases are the cornerstone for most of the modern data processing systems, as well as, for the decision support systems that constitute the "nervous system" of the organizations. Databases enable dynamic information access and retrieval requiring very small physical volume and cost in contrast with their manual predecessors. The economic significance of databases and the spreading of their applications make the ability of their designing and development comparable to that of programming [2]. International organizations predict that database administration related jobs will face among the greatest demand increasing in the interval 1996-2006 mainly because of the internet applications that are database based, the Enterprise Resource Management systems that organisations are installing and the availability of desktop dbms's [1]. Non trivial database design by untrained and/or non expert people is not always successful [3]. The above facts generate utilitarian interest and social requirement for large scale education in database use and design.

2.2 Didactic interest for database education

Databases are models representing the structural attributes of corresponding physical systems, [5, 7, 9]. Dbms's are general purpose modelling environments using data models as representation systems [2, 4]. Data modelling process during database design has a clear ***constructivist character*** with final product the implementation of the database. In contrast to other structural modelling methods, (e.g. concept maps) data models' application produces databases that are amenable to computation enabling the modeller to interact with them. Interaction includes information retrieval, hypotheses testing, comparisons etc. The interactivity of databases facilitates the feedback "circuitry" construction to the designers. In addition database design requires often ***social collaboration*** with other designer and/or with the problem provider.

Working on dbms users exploit computational resources in problem solving using conceptual frameworks closer to the solver and/or the problem than to the architecture of the machine. One of the most obvious differences of databases to other modelling environments is that problems' exploiting databases can concern a large amount of entities which attributes' relationships are not simple algebraic ones or they are under investigation.

Using data base design, it is possible to propose interesting learning activities for a large variety of subjects where students construct interactive artifacts (databases) collaboratively fulfilling both the social constructivists learning theories issues and the modern student centred pedagogies. Databases are in this manner executable structural models that have general didactic and learning value.

In the following sections we support the above considerations/assumptions, presenting the general categpries of learning activities using databases design as well as the features of students' involvement in them.

2.2.1 General categories of learning activities using database design

According to Jonassen [6] databases are considered as a kind of mindtool that not only amplifies human cognitive power but transforms it requiring new ways of thinking [8]. Furthermore, Jonassen considers concept maps as another mindtool of the same category, namely "Semantic organization tools". We consider database design as a combination of informal and typical semantic analysis [11] that requires the construction of a concept map in order to produce a database schema. Combining the kinds of learning activities proposed by Jonassen for databases and concept maps, we can distinguish at least the following five categories of database design learning activities:

I. Ontological analysis of study field

In activities of this category students are supplied with a set of information sources (e.g. books, papers, www sites, authentic documents etc) in order to design a database that stores significant data in the framework of a study or the solution of a problem. During this kind of activities students could be committed to:

- Recognize and define concepts and/or object types from the problem domain as well as their relationships of interest. Furthermore, students must decide which attributes of the entities concerning the problem should be stored in the database in order to be useful for the activity.

- Collect data systematically to store in the database.
- Process data in order to discover or investigate relationships among the entities' attributes, to compare, to reason, to confirm, to contradict etc.

The active involvement in ontological and conceptual analysis could be helpful in the improvement of the understanding of a study field or problem domain as well as to the development of metacognitive ability. Databases can be used for the discovery of knowledge and relationships using constructive activities in cases, where usual teaching means and strategies result in memorization exercises due to the large amount of data.

II. Real stories analysis

A large part of human knowledge is organized as a network of cases that humans recall, when they confront similar situations. Human experts often explain or decide and plan their actions mentioning similar cases. The systematic analysis of case studies a basic technique in the education of sciences like medicine or law. Case based reasoning is also a known technique for the development of Artificial Intelligence (AI) systems that present common sense. AI case based reasoning systems have a knowledge base of cases that enables them to trait similar problems [9].

In 'real stories analysis' learning activities, students study a large amount of a problem family cases in order to design a database that they could exploit in a case of an analogous problem. The stories could be found in bibliographic references or can be a product of interview and/or observation, etc.

III. Using databases as study guides

Study guides are databases that a basic design has been implemented by the teacher and students store data as they study a subject. As an example, a student could store chemical characteristics of materials that he/she studies in a predesigned database. The resulted database could be used for information retrieval and general data analysis. In addition, the combination of the study guides of a large number of students could lead to common database containing a very large amount of data, which could enable interesting questions' formulation.

IV. Using database design for understanding assessment

The database design ability indicates at least structural knowledge for a problem. Furthermore, design errors often externalize misconceptions. These observations are the base of the last kind of learning activities in which students design databases in order to assess their understanding of a subject. In these activities it is purposeful to facilitate feedback mechanisms.

2.2.2 Characteristics of the involvement in learning activities using databases

Students involved in learning activities of the general categories described above are expected to have increased opportunities to experience many interesting characteristics that constitute meaningful learning [6]. More specifically students:

Activate their own cognitive models in order to plan their actions to the database through the interface of the dbms as well as to interpret the results of such manipulations.

Design and construct an interactive artifact in the framework of authentic problems that preserve their motivation. In addition, students often self manage their learning process, formulating the learning goals as well as the process to obtain them.

Reflect during the conceptual database design for the determination of the entities and their relationships. Conceptual modeling is a highly metacognitive and reflective process in the sense that externalizes the mental processing of students experiences. Students reflecting on a problem improve their structural knowledge about it and increase the possibility to advance their problem solving ability. Moreover the database design reflection can be supported during data analysis phase where students are able to check hypotheses, interpret information, compare data etc. For example student could reflect using a geographical database to relate general economic with social indices such as infants' death rate etc.

Collaborate in groups in the problem analysis as well as in the data collection and analysis phases. The collaboration advances their socialization and helps them to formulate common meaning using social negotiation.

Overcome the limitations of human memory and to decrease significantly the data processing time. As a result students can save time for critical analysis of data and understanding of interrelations between entities of the problem domain.

Reorganize students thinking combining the basic processing structures of dbms's for the solution of complex problems. Especially in relational dbms's students can solve complex problems exploiting relational algebra through the declarative language SQL.

The above characteristics and their learning facilitation value can be estimated through the following analysis of a case study where students were involved in a real story's analysis activity for the study of vehicular accidents.

3 A case study of real stories of vehicular accidents analysis using databases

In the case study presented in the following sections, 11th grade students participated from two classes of a public vocational school in Greece, named 2nd TEE of Rhodes. The 41 students were assigned an obligatory subject named "databases" in which they where familiarized with the desktop dbms that was available in the school lab while they had a typical introduction to the relational and Entity-Relationship data models for logical and conceptual database design respectively. For the familiarization of students with the relational databases design, and use, a series of lectures and hands on lab activities has been designed and implemented. The instruction of database design to students is an educational research questions generation point that it interests the authors but [10, 11] the analysis of which is out the scope of current paper.

The activity scenario required students to get interviews from people that have been involved to vehicular accident in any way in order to analyze the collected data to produce knowledge and formulate conclusions of any kind about the phenomenon. The scenario is organized in four discrete phases. In the 1st phase, students constructed a common concept map collaboratively, merging the maps produced by students groups, in order to understand the problem and determine the abstraction level that they will adopt. In the 2nd phase the produced concept map, as a "typical" problem definition was used as input for the database design process. Students designed the database in a collaborative manner for each class. The design of the first class was merged with that of the second class in order to use a common data base. In the 3rd phase students used the database design for the formulation of the record that should be constructed for each interview. Interview records were paper forms that reminded the interviewer what data should be recorded in a systematic manner. In the same phase also, the data collected by the interviews were stored in a common database. In the 4th phase students engaged in information retrieval and data analysis activities for knowledge production, hypotheses test etc. In the following sections the students activities are briefly described for each phase.

3.1 The 1st phase: Conceptual map conctruction

During the first phase of the activity, students constructed a conceptual map for the problem of vehicular accidents. Students decided about the concepts their attributes and characteristics concerning the problem clearing out ambiguous points. Students worked for two sessions of 45 minutes each. During the first session students worked in small groups while in the second as a whole class where they combined their concept maps in that of fig. 1 using the whiteboard. Each student took a hard copy of this concept map.

The concept map was constructed using brainstorming, social negotiation and limited teacher support. Students had to decide the abstraction level for the description of the concepts as well as to define their meaning. Disagreements are unavoidable in an operation like this, but students solved these problems using social negotiation in which they formulated arguments using terms from the problem domain and asking for teacher support. During conceptual map construction students' reflect on and become conscious about the conceptual structure of the problem domain. Students used the conceptual map in order to design the database as well as an interview record as it is described in the following sections.

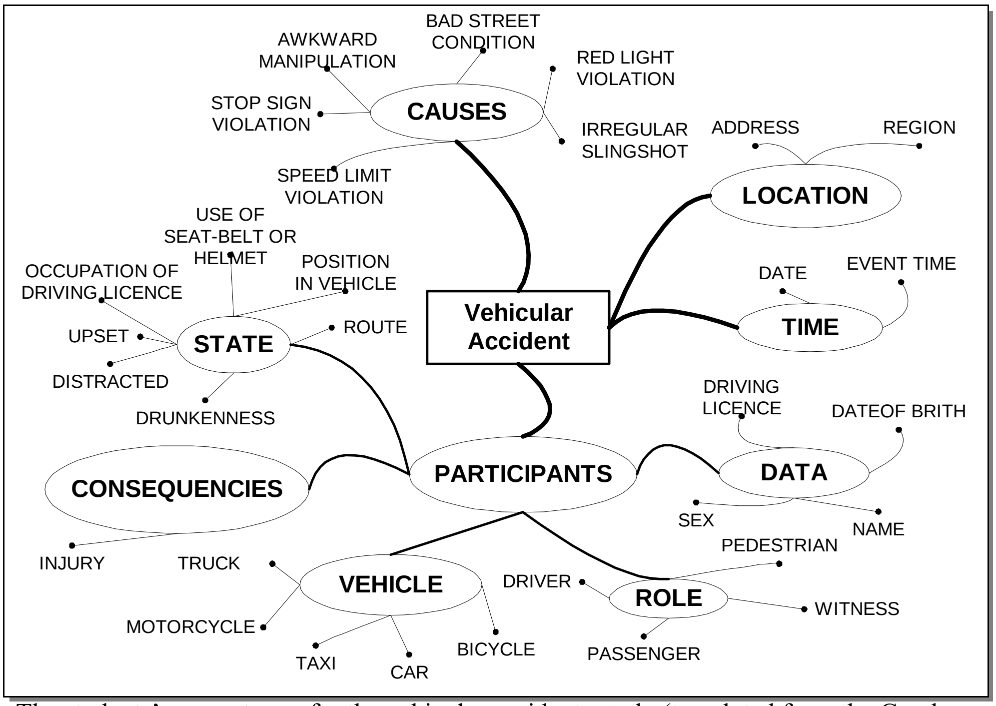

Figure 1. The students' concept map for the vehicular accidents study (translated form the Greek manuscript).

3.2 The 2nd phase: Database design

The database design complexity was significantly decreased for students since they had already decided about the information that should be stored and faced most of the understanding problems during the conceptual map construction.

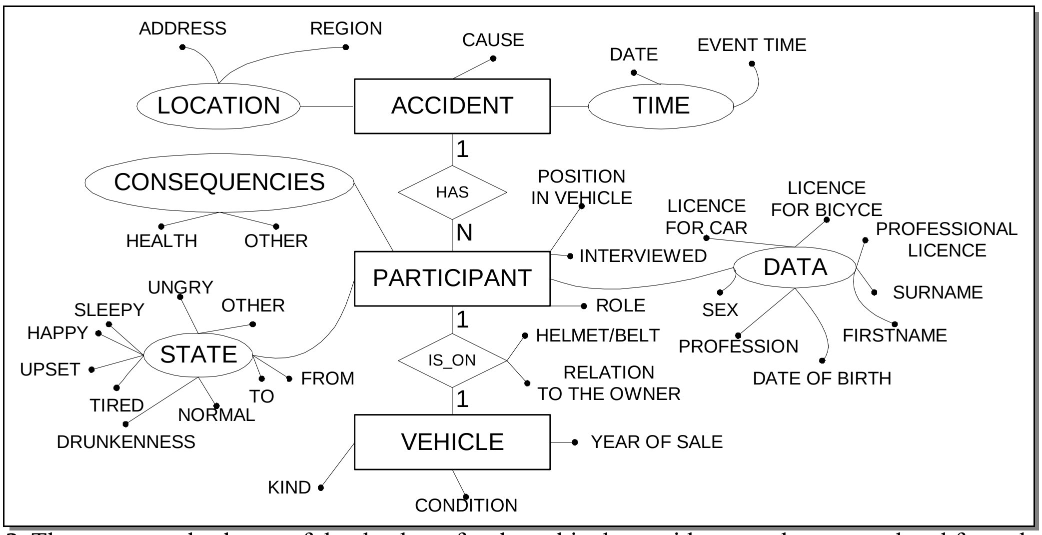

Figure 2. The conceptual schema of the database for the vehicular accidents study as translated from the Greek manuscript.

Fig. 2 presents the ER schema for the database produced by students in common. Database design was implemented during two sessions, 45 minutes each. As in the case of the conceptual map students worked in small groups in the first session and then as a class they

produced combined their designs in a common one. Database should be common not only among the groups of each class but also among the two classes in order to store data in one database of a more interesting size. In fig. 2 we can see that the ER conceptual schema it is about three entities full of attributes and related with two binary relationships.
Students face problems with database design but there is a percentage of students that can accomplish successfully the design process for many interesting cases. Distributing these students in groups in companion with the class function phase can eliminate the motivation elimination cases.

Students reviewed the database design in this phase was by test interviewing each other and considering questions that should be answered by the proposed database.

3.3 The 3rd phase: Data collection and entry

Students designed interview records using conceptual map and database design. In order to get familiarized with the interview records students interviewed each other. After the familiarization and the final corrections of the interview records, students had a period of a week in order to collect the real data. The people that students asked for interview where students from other classes, or other schools, teachers, parents, family friends etc. The data collection week was a pleasant disturb for the normal school life since it become the main issue at school and was accompanied by several incidents that equipped students with rich social experiences. Finally, students collected 45 interview records concerning 69 accident participants. A small group of students was assigned to implement the data entry using a special data entry form designed by them (fig. 3).

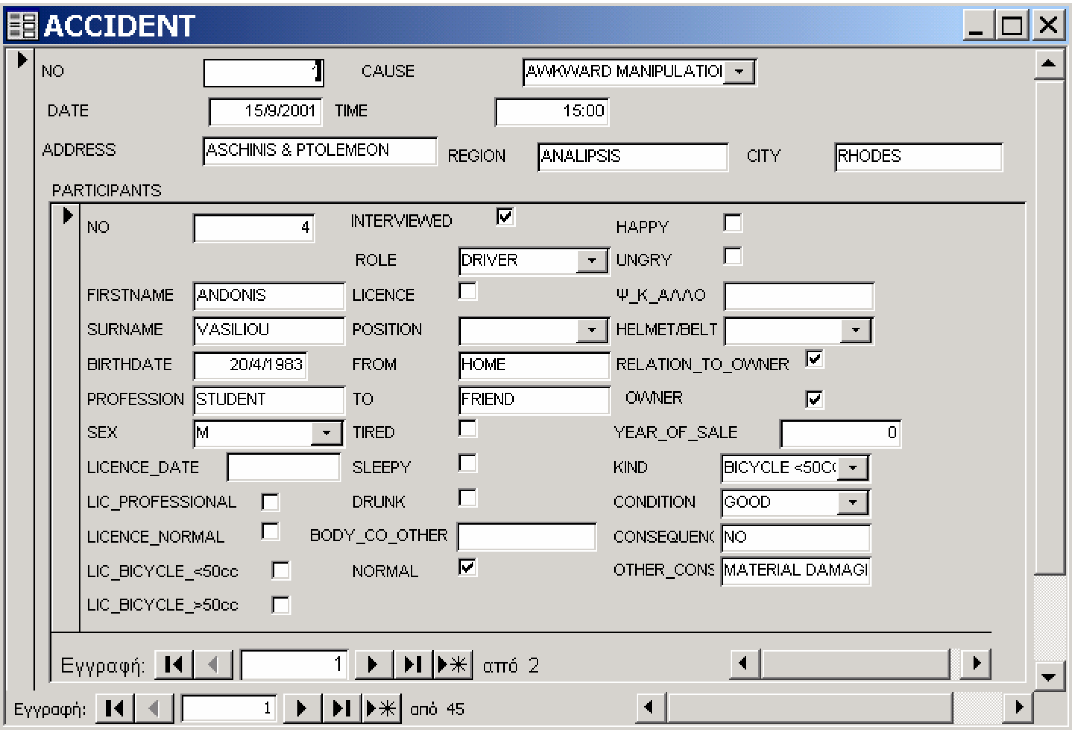

Figure 3. The data entry form. It contains a subform for the participants of each accident.

3.4 The 4th phase: Data analysis end exploitation

Data analysis was the last phase of the learning activity. Students' groups supplied with a, full of data, copy of the database in order to create answers to their own and/or teacher made questions. The teacher's action in this activity includes:

- Proposition of questions to groups that could not decide what to study and advanced questions in groups that completed some basic ones.

- Checking of the queries' correctness and interpretations reasoning.
- Supporting the dbms' use and the syntax of queries.
- Encouraging students' groups to produce interpretations of their results and to compare their findings with those of other groups for the same question.

Students worked in small groups for the first session and then presented their findings to the class. Students develop a competitive climate among the groups in terms of conclusion production and information retrieval that helped to advance their dbms use ability as well as the understanding of the subject. Situations with special learning interest appeared when different groups disagreed on the results for the same questions. Students tried to find which group was right triggering social negotiation and powerful feedback mechanisms.

3.5 Information discovered using database

Queries produced by students are a good indication about the quality of the learning situation that they have experienced, as well as of the related concepts understanding. Three of the most interesting and representative questions and underline concepts are presented in the following sections.

Query 1. Frequency distribution of participants to the vehicle kind.

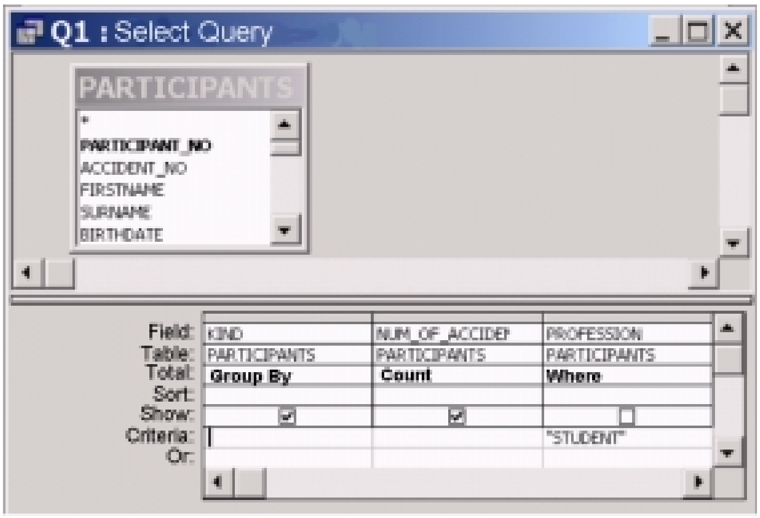

Figure 4. Query for the frequency distribution of participants to vehicle kind, translated from the Greek version of the database.

KIND	NUM_OF_ACCIDENTS
	2
BICYCLE <50CC	18
BICYCLE >50CC	10
CAR	2
TRUCK	2

Εγγραφή: 1 από 5

Figure 5. Frequency distribution of students per vehicle kind.

KIND	NUM_OF_ACCIDENTS
	4
BICYCLE <50CC	4
BICYCLE >50CC	3
CAR	20
SMALL TRUCK	1

Εγγραφή: 1 από 5

Figure 6. Frequency distribution of non students per vehicle kind.

Students TA02 and TA20 formulated the question *"is the participant vehicle motorcycle or car;"* meaning (as they explained) that they would like to see whether more of the accidents concern motorcycles than cars. Students hypothesized that it is more likely to have an accident using a motorcycle. In order to approach the question students produced a query that computes the frequency distribution of the participants to the vehicle kind as it is presented in fig. 4. The teacher asked students to produce the frequencies' distribution of participants in vehicle kinds, for students and non students separately. Students produced the frequency distributions of figures 5 and 6 respectively. Comparing the results, students were confused initially and then said that students' accidents concerned motorcycles more frequently than non students and that the total number of students' accidents in the database is larger than those of non students. After a short discussion students seemed to understand that in order to see if motorcycles participate more frequently in accidents they should compare the ratio of accidents concerning a motorcycle per number of motorcycles in use with corresponding ratios of other vehicle kinds, thus working on ***"simple statistical concepts"***.

Query 2. Safety means use frequencies.
Many students groups produced the query of figure 7 for the question *"How many of the participants did use the belt or the helmet"*. Students commented on the infrequent use of the safety means but they observed that the total number of participants is 68 instead of 69. This observation uncovered an orphan participant that was not assigned to any accident and triggered a discussion about the ***"foreign key"*** and the ***"referential integrity constraint"*** concepts.

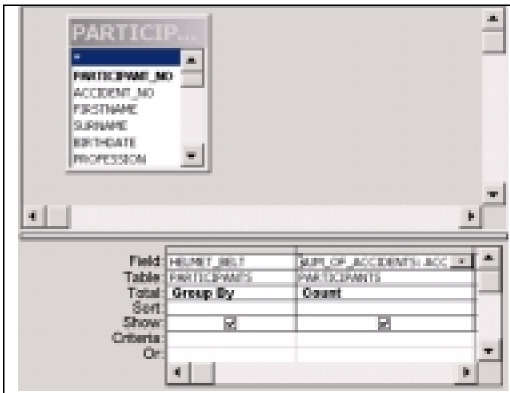

Figure 7. Query safety means use.

HELMET_BELT	NUM_OF_ACCIDENTS
	62
BELT	5
HELMET	1

Εγγραφή: 1 από 3

Figure 8. Frequencies of safety mean use.

Query 3. Accident causes.
Students TA05 and TA14 computed the frequencies distribution of accident causes as it is presented in figure 9 and 10. Causes' descriptions are as the interviewed people mentioned them. Students TA16 and TA19 answered the same question with a different query presented in figures 11 and 12. The whole class was asked to conclude about the correctness of the proposed solution. Students rejected the first solution (TA05 and TA14) because the sum of frequencies is not equal to the number of accidents (45). Investigating the source of the error students found that the difference between the queries is the redundant ***join*** of the first query imposing the combinatorial semantics of the ***"join"*** concept.

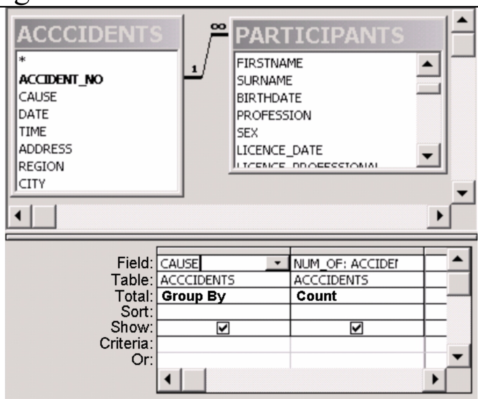

Figure 9. Accident causes frequencies query by TA05 and TA14.

CAUSE	NUM_OF
AWKWARD MANIPULATION	17
SPEED	12
STOP VIOLATION	10
SPEED LIMIT VIOLATIONΣ	10
DISTRACTION	6
IRREGULAR SLINGSHOT	4
SLEEPING DRIVER	3
VOLTE-FACE	2
STREET CONDITION	2
SLIPPERY STREET	1
RED LIGHT VIOLATION	1

Εγγραφή: 1 από 11

Figure 10. Accident causes frequencies by TA05 and TA14.

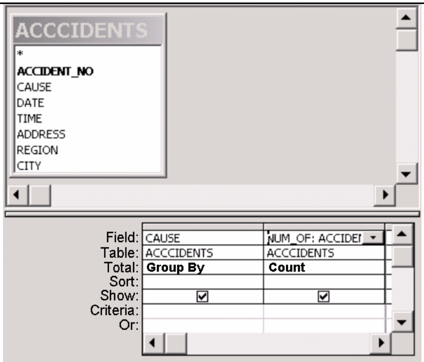

Figure 11. Accident causes frequencies query by TA16 and TA19.

CAUSE	NUM_OF
AWKWARD MANIPULATION	13
STOP VIOLATION	7
SPEED LIMIT VIOLATIONΣ	7
SPEED	6
IRREGULAR SLINGSHOT	4
DISTRACTION	3
VOLTE-FACE	1
STREET CONDITION	1
SLIPPERY STREET	1
SLEEPING DRIVER	1
RED LIGHT VIOLATION	1

Εγγραφή: 1 από 11

Figure 12. Accident causes frequencies by TA16 and TA19.

3.6 Discussion on the Learning activity general characteristics

In this section some essential factual observations are collected in order to support the assumptions about the didactical interest of database design based activities as they presented in paragraph 2.2.2 above. The previously described learning activity permits at least the following, didactical interesting, observations:

- It concerns the collection of a large amount of authentic data.
- Problem domain entities and their relationships are not given, but they are defined during problem analysis' phase using apart from references and teacher support, the brainstorming, and social negotiation processes.
- Entities attributes' relations are not known and/or simple algebraic, but they are a discovery object.
- Students worked in a group manner even in after school hours in social related activities.
- Students managed to retrieve interesting information querying the collected data in a relatively short time interval producing mainly frequencies' distributions. The time requirements for the solution of the same problems without database technology tools are not comparable. Furthermore, students could compare, try and improve several queries before the adoption of a solution.
- The comparison of the different solutions among student groups for the same questions provided feedback to students and triggered discussion about the subject of the study as well as the databases technology.
- Students are able to formulate interpretations for a study domain and to try to support those using queries for the database. During this process students can transfer knowledge from the database to the problem domain and vice versa. This transfer function is called conceptual reframework and is a main learning mechanism of modeling environments.
- Group work is possible to decrease the possibility of student failure engagement in database design based learning activities if there are a variety of ability levels among the group members.

The above observations of the learning activity implementation constitute evidence that students engaged in Real Story Analysis learning activities using database design present the characteristics of *activation, design and construction, reflection, collaboration, limitations overcoming, and thinking reorganization* as described in section 2.2.2. In other words, students have increased possibilities to experience high quality meaningful learning situations and activate high order thinking in a manner consisted to the modern social constructivist learning theories and pedagogical approaches.

4 Conclusions

The educational interest about instruction **for** the use and design of databases is based on utilitarian as well as pedagogical reasons. Utilitarian interest sources from the economic significance of databases technology while pedagogical interest from the possibility of implementation of general learning activities **using** databases' design and application. Students' systematic familiarization with database design and use is expected to increase the likelihood of successful involvement in such learning activities. Database design based activities can be at least of the following general categories: Ontological analysis of study field, Real stories analysis, Study guides, and Understanding assessment.

In this paper, it was presented a database design based learning activity, of the category of Real Stories Analysis. As the case study presented supports, database design based learning activities characteristics are consistent to modern learning theories and pedagogical approaches. Databases are considered as executable structural models and their design as a modelling activity. During data modelling, students have the opportunity to experience situations in which they activate high order thinking and reflection obtaining meaningful learning. Students designing databases improve the structural knowledge about the subject under study and increase their problem solving ability. In addition, due to the collaborative character of the design activities, they exercise their social and communication skills in the framework of authentic and meaningful problems.

References:

[1] Antony, S.; Batra, D.: CODASYS: A consulting Tool for Novice Database Designers, ACM SIGMIS Database, V:33, N:3, pp.:54-68, 2002

[2] Batini C.; Ceri S.; Navathe B. S.:, *Conceptual Data Base design,* CA: Benjamin-Cummings Pco, Inc., 1992

[3] Batra D.; Hoffer A. J.; Bostrom P. R:. *Comparing Representations with Relational and EER Models*, Communications of the ACM, 33(2) , 126-139, *1990*

[4] Brodie L. M.; Mylopoulos J.; Schmidt W. J. (Eds):, *On Conceptual Modelling,* 19-49, NY: Springer-Verlag Inc., *1984*

[5] Hancock, C.; Kaput, J. (1990). Computerized tools and the process of data modeling in G. Booker, P. Cobb, & T. deMendicutti (Eds.), Proceedings of the 14th Conference of the international Group for the Psychology of Mathematics Education, Mexico, V:3, pp.:165-172

[6] Jonassen H. D.: *Computers as mindtools for schools,*2nd Ed 1-57, NJ: Merrill, *2000*

[7] Minsky M. L.: Matter, Mind, and Models in: M.L. Minsky (ed.): Semantic Information Processing., MIT Press, 1968

[8] Pea D.: Beyond amplification: Using the computer to reorganize mental functioning, Educational Psychologist, Vol. 20, No. 4, 1985, pp.: 167-182

[9] Rich E.; Knight K.: *Artificial Intelligence,* 2nd Ed, 109-110, NY: MGraw-Hill, Inc., *1991*

[10] Fessakis G.; Dimitracopoulou A.; Komis V.:, Students' cognitive models about databases and difficulties in manual database design. *Third Pan-Hellenic conference with international participation: "ICT's in Education",* University of Aegean, Department of Education, ETPE, Rhodes 26-29 Sep. Kastaniotis Editions. pp.:267-276, 2002

[11] Fessakis G.; Dimitracopoulou A.; Komis V.:, Improving database design teaching in secondary education. Using action research for documentation of didactic requirements and strategies., under submission to the Computers in Humans Behavior, Elsevier, 2003

Author(s):

George Fessakis, BSc, MSc Informatics, Phd student
University of Aegean, Department of Education,
Learning Technology and Educational Engineering Laboratory
Griva 23, 85100, Rhodes, GREECE, gfesakis@rhodes.aegean.gr

Angelique Dimitracopoulou, Associate Professor
University of Aegean, Department of Education
Learning Technology and Educational Engineering Laboratory
Av. Demokratias, 85100, Rhodes, GREECE, adimitr@rhodes.aegean.gr