OO and EER Conceptual Schemas: A Comparison of User Comprehension

Peretz Shoval and Israel Frumermann
Ben-Gurion University of the Negev

A database conceptual schema serves as a communication medium between professional analysts/designers and users who wish to comprehend and validate the conceptual schema. The conceptual schema is usually presented in a diagrammatic form that follows a specific semantic data model. The extended entity-relationship (EER) model is one of the most commonly used models, but being “threatened” by the object-oriented (OO) approach, which penetrates into the areas of systems analysis and design, as well as data modeling. The issue of which of the two data models is better for modeling reality and is easier to comprehend is still an open question. Our response to this question was to conduct a controlled experiment in which two groups of users, each trained to use one of the models, were tested for the comprehension of equivalent schema diagrams. Comprehension was measured by the number of correct answers to questions that addressed different constructs of the models.

The results of the experiment reveal that there is no significant difference in comprehension of facts dealing with attributes of entities or objects, binary-relationships and two relationships, but those dealing with ternary-relationships are significantly easier to comprehend with the EER model. Comprehension of other, unclassified facts, however, is easier with the OO model. We propose a special symbol for objects representing ternary and higher order relationships in order to overcome the weakness of OO diagrams.

The object-oriented (OO) approach penetrates into many areas of computing, including systems analysis and design, and databases, not to mention programming. Within the database arena, the OO approach is utilized both as a model for database management systems (DBMS), and a method for data modeling. As such, it can be viewed as a semantic data model. Semantic data modeling is an analysis and design activity, aimed at formalizing and representing the data structure of reality in terms of a conceptual schema. Afterwards the schema is converted into a database schema and implemented in a specific DBMS. A semantic model provides constructs that allow designers to grasp more of the semantics of reality than that which was formerly obtained with classical data models (e.g. relational model). Peckman and Marianski (1988), Batra et al. (1990), and others, point out that semantic models, as compared to classical ones, enable the user to more easily extract the same information, and permit him, through abstraction, to model and view the data in a manner which is consistent with how people view the world. The product of semantic data modeling—the conceptual schema—is usually represented in a diagrammatic form that portrays the data structure of the reality being modeled in terms of entities (or objects), attributes, and relationships. A conceptual schema diagram must be powerful in its semantic expressiveness and easily comprehensible, as it serves as a communication medium between professional designers and users (including managers) who interact during the stage of requirements analysis and modeling, and validate the design.

One of the most prominent data modeling methods is that of the entity-relationship (ER). The original ER model (Chen, 1976) has undergone many variations and extensions, which, for the purpose of this study, will be termed EER (extended ER). EER is used not only as a “stand alone” data modeling method, but is incorporated within various methods for information systems analysis and design, and has become an integral part of almost any “upper” CASE (computer-aided software engineering) tool. Generally, the analyst/designer interacts with the user in order to understand the data structure of the reality being modeled, and creates a model in the form...
of an EER diagram—the conceptual schema. This is usually done in parallel with functional analysis and design activities, although it can be done independently of those stages. Once the EER diagram is validated by the user representatives (as an appropriate model of reality), it is transformed into a database schema, usually a relational DBMS. This is usually an algorithmic process, based on well defined procedures, yielding normalized relations.

Many studies on the evaluation and comparison of data models and methods have been conducted in the last decade, and the EER is the most frequently studied. Some studies compared the EER with other conceptual models, while other studies compare the EER with the classical, “record based” models, usually the relational model. Various other studies compared the data models and methods from a designer’s perspective in an attempt to find out which yields better and more precise schemas, and others compared the query languages of the data models in order to determine which languages compose queries more accurately or faster. Some investigators compared models from a user’s perspective in an attempt to determine ease of comprehension. Although the results of these investigations are not always clear-cut or consistent, there is a tendency to agree that EER is superior to other, record-based and conceptual models.

Recently, the OO approach has begun to challenge many existing approaches. In the area of systems analysis and design we see a vast amount of OO methods threatening to replace the more “traditional”, functional-oriented approaches. It is argued that the data structure of a certain reality is more stable than its functional structure, and that it is easier to identify the object structure of the system rather than its “intangible” functions. Due to the one-to-one correspondence of objects to entities in the real world, OO can be used as a semantic data model, i.e., as a method for conceptual representation and validation. As such, we focus on the OO as an alternative to the EER.

Models can be compared according to various dimensions. We compare the OO and EER models for one dimension—user comprehension. This dimension is crucial for validating the data structure of reality, as represented in the form of a conceptual schema. Users judge if the conceptual schema is correct, consistent and complete, and only then validate it. Obviously, this understanding depends on the simplicity/complexity of the data model that is being used for the conceptual schema. In this study we question if there is a difference between EER and OO with respect to user comprehension, and which diagram is easier to comprehend. We conducted a controlled experiment with two groups of “sophisticated users” who are given equivalent EER and OO schema diagrams. Their response to a set of questions about facts represented in equivalent schemas measured their comprehension of the schemas. It should be noted that user-comprehension is only one aspect of evaluating data models. Other aspects, such as the quality of the model as a design tool for designers, are not addressed in this study.

The rest of this paper is structured as follows. Section 2 provides a brief survey of previous comparisons between relevant data models, emphasizing user comprehension. Section 3 briefly describes and exemplifies the specific EER and OO models that we compare. In Section 4 we present the research design and hypotheses, and in Section 5 - the details of the experiment. The results are analyzed in Section 6, along with conclusions and recommendations for further research.

**Related Studies**

Several prior studies have compared different data models in an experimental setting (Juhn and Naumann, 1985; Ridjanovic, 1986; Jarvenpaa, 1989; Shoval and Even-Chaime, 1987; Leitheiser et al., 1988; Jih et al., 1989; Batra et al., 1990; Chan et al., 1991; Palvia et al., 1992; Kim and March, 1994). Most of the studies compared EER with both earlier (classical) models, mainly the relational, and with other semantic models. In almost all studies the subjects of the experiments were students with varying degrees of training in information systems and databases. In some of the studies the task of the subjects was model development, i.e., designing schemas, with the objective of finding out which data model or method provides a better, more correct and precise schema, which model designers prefer to use, and which model requires less time for completing the task. In some studies the aim was to find out which model enables users to compose queries more accurately or in less time. In other studies the objective was finding out which schema (i.e., data model) is easier to comprehend by users.

An overview and evaluation of previous studies can be found in Batra et al. (1990), Batra and Srinivasan (1992), Srinivasan (1992), Palvia et al. (1992), and Kim and March (1994), and is not reviewed here. As we are concerned with user comprehension of EER and OO schemas, we concentrate only on some of the related studies. Juhn and Naumann (1985) studied end-user comprehension, comparing semantic models (including ER) with “data storage” models (including the relational). They found that the semantic models were more effective in tasks related to understanding relationships (relationship-existence finding and cardinality finding), but the relational model outperformed other models with respect to identifier comprehension tasks. Leitheiser et al. (1990) studied end-user model comprehension and found that a semantic model (LDS) was easier to learn and resulted in a better understanding and recall of a database schema than a tabular representation. Jih et al. (1989) examined the performance of end users who employed the relational and ER models, by testing the quality of query composition. Their results showed that even though users of the relational model made fewer syntactic errors, users of the ER model not only made fewer semantic errors but were faster in writing queries. Chan et al.
(1991) also compared the two models from the point of view of query composition (using SQL for the relational model and KQL for the ER model). They tested the accuracy of queries, time for completing the task, and user confidence. Their results were conclusive because performance with the ER query language KQL was superior to performance with the relational language SQL for all dimensions. Batra and Srinivasan (1992) concluded that semantic data models lead to better performance in many situations than does the relational model, but further experimentation is necessary. With respect to a conceptual modeling task, they concluded that modeling relationships is more problematic than modeling entities and attributes. Hence, the usability of data models should be evaluated by their ability to easily model relationships. As will be discussed, our study emphasizes the ability to understand relationships.

Two recent studies that are quite relevant to our experiment are reviewed in more detail. The first is by Kim and March (1994), who compared two semantic models — EER and NIAM (Nijssen Information Analysis Method) on user comprehension. Their experiment consisted of 28 graduate business students who were randomly assigned to one of two treatment groups, and trained in the appropriate data modeling formalism. They used the case of a manufacturing company and described its conceptual schema in two equivalent diagrams. The EER schema contained 12 entities, 12 relationships (11 binary and 1 ternary) and 33 attributes. User comprehension was measured by the number of correct answers to questions about basic modeling constructs (no examples were provided). They found no difference in comprehension performance between the two semantic models. They concluded that given a small amount of training, users are capable of reading and validating data models of non-trivial size and complexity.

Palvia et al. (1992) conducted an experimental study on end-user comprehension of three conceptual data models: ER, DSD (Data Structure Diagram) and OO. Given the absence of a standard OO model they used the approach described by Kroenke (1992). (They admit that this OO version is a simple one, and suggest that further research be directed to examining various variations and details of the OO model.) As the focus of their study was on evaluation of data structure, they did not include “methods” of objects. Their subjects were students of an introductory MIS course in a business school, who had only a limited exposure to computers, so they resemble novice end-users. Three versions of a database, with identical data in each, were developed. The database included data on three entities and two binary-relationships between them. These schemas are certainly trivial in terms of size and complexity (they show only the ER schema). The subjects, roughly 40 in each group, received a version of the database corresponding to one data model, along with a test consisting of 30 questions that evaluate comprehension of the database. The questions varied in complexity (however, the authors did not make clear how they measured complexity). The total score on correct answers was used as a measure of database comprehension. The test for difference revealed that users of the OO model showed a significantly better understanding of the data than users of the other two models. Therefore the authors agree with the claim that OO models produce a richer and more natural representation of the user’s world than the DSD and ER models. They admit, however, that their conclusion is weak, because comprehension was measured on overall terms only. They suggested that future research be directed to studying specific aspects of comprehension, e.g., comprehension of different types of entities and relationships, which is the intent of our study.

In a second study (reported in the same paper) Palvia et al. (1992) evaluated the same models for a set of 17 characteristics (e.g., communication ability to users, ease of use, ease of learning, overall quality). Each dimension was measured for user perception in a survey that included 30 subjects, on a 1-5 Likert scale. The subjects where MBA students with greater MIS and database exposure than that of the first study, and had used at least one of the models or had a working knowledge of it. The results of this survey showed no significant difference among the three models for almost all dimensions. The authors subscribe the different results of the two studies to the greater computer literacy of the subjects used in the survey. Therefore they infer that comprehension will improve as users become more experienced in computer and data processing.

Our study also bases user comprehension on a questionnaire, but with the following differences: we use larger and more complex schemas, our subject are more computer literate, and the questions are classified according to various constructs of data models, distinguishing between different relationship types, so as to enable measurement of comprehension not only on overall terms but also on the various constructs. Moreover, we test comprehension of the conceptual schema, rather than comprehension of the databases as we emphasize the role of users who validate the schema. Before detailing our experiment, we provide a brief description of the two models used.

### The EER and OO Models

The objective of this section is not to provide a tutorial on the two models, but to overview and exemplify them. Since each of the models has many variants, we confine ourselves to two specific types, which are sufficiently general to validate the results of the comparison.

#### The EER model

As EER is more “standard” than OO, we provide only a brief overview. Figure 1 shows a small example of an EER diagram which includes all our constructs of the model. The diagramatic notation is similar to the one in Teorey et al. (1986). We distinguish between entity (rectangle), attribute (circle) and relationship (diamond) types. (For brevity, we will omit “type” from the construct names.) An identifying attribute is
underlined, and a multi-valued attribute has a dual circle. Cardinalities of relationships are signified by both the m:n notation (e.g., 1:n, or m:n:n) and by coloring the “n” edges of the diamond. Supertype-Subtype hierarchies between entities are signified by arrows linking the sub-entities to the super-entity. We do not show “weak entities” and constraints, e.g., total-role, as they have no equivalents in the OO model.

The OO model

Since the OO model is still evolving there is no “standard” OO. Moreover, even within a specific OO model there are variations, and the same reality can be represented in many different forms. As we are concerned with user comprehension, our model and notations detail all necessary information on the OO diagram, ignoring issues of redundancy and efficiency of implementation. Special attention is given to the representation of relationships between object classes. It has been shown that a relationship can be represented in different ways (see, for example, in Spaccapietra et al., 1992). There is no standard or “normal form” that can guide an OO designer to decide which representation is preferable, and much depends on the assumptions about the types of queries that will be posed at run time. Another issue is “methods”. There is no doubt that methods contribute to the modeling of systems behavior. As a matter of fact, this is a major argument in support of the OO approach. The focus of this study, however, is on data structure and user-comprehension of the conceptual schema diagram, and does not take into account aspects such as behavior of the system. Therefore, methods are not included in the OO diagrams.

An example of an OO schema is shown in Figure 2. It represents the same reality as the EER diagram in Fig. 1. An object class (rectangle) has attributes. (For brevity, we will omit “class” and simply say “object”.) An attribute may have an atomic value (e.g., attribute “state” of object City) or be multi-valued, in which case the attribute name is preceded by
“set” (e.g., set “telephones” of object Person). The type of value(s) of an attribute is usually implied from its name, without being explicitly specified on the diagram, unless it is not obvious. (For example, “state” of City is not typed, as it is obvious that we mean state name, but Truck has an attribute “load-capacity” typed “tons”, written in parentheses). An attribute which is a unique identifier of an object is underlined.

An attribute may refer to another object. This is signified by writing the referenced object name in brackets next to the respective attribute name. For example, “mayor” of City is a reference-attribute to Person. This actually signifies a relationship between the objects. A relationship is bi-directional, signified in two complementary ways: a) reference-attributes are included in each of the involved objects (therefore, in the above example, in addition to the attribute “mayor” of City, there is the attribute “is-mayor-of [City]” in Person); b) a link connects the two objects, with indication of the relationship cardinality, using the n:m notation (so the above relationship is marked 1:1). Note that in the case of an 1:n relationship between two objects, the object in the “1” side has a “set” reference attribute to the other object (e.g., “set inhabitants [Person]” in City), and in case of an m:n relationship both objects have set reference attributes. An attribute may be a tuple consisting of multiple (value or reference) attributes, and signified by { } brackets. For example, City has a set attribute “producers” which is a tuple of two: a reference attribute to Producer and a valued attribute “number-of-plants”. (More examples of tuple attributes can be seen on the diagram.)

Relationships are represented as follows: a binary-relationship is represented by two objects with mutual reference attributes, and a link indicating the cardinality. (An alternative representation could have been to include reference attributes in both objects, or, in the case of an m:n relationship, to create a separate object for the relationship, with or without reference attributes. In our opinion, the representation we adopt is simple; it shows all the semantic information on the
diagram, and is one that is commonly used in OO data modeling.) When there is a ternary- (or higher order) relationship, it is represented as a separate object, linked to the other objects involved in the relationship, and includes the appropriate reference attributes in all objects. In our example, we have to store facts about agents selling vehicles to persons. Therefore, we use the objects Person, Agent and Vehicle, and for the relationship we create the object Sales. Each of the three original objects has a set reference attribute to Sales, while Sales has three reference attributes to those objects. In addition, it has the valued attributes “price ($)” and “date-of-sale”.

Finally, object hierarchies are signified by links from the subtypes to the supertype, similar to the EER notation. In both models, the subtype objects (entities) inherit the attributes and relationships of the supertype, and include only their own specific attributes and relationships. Note that there is no visible difference between the two models with respect to the representation of object/entity hierarchies.

**Research Design and Hypotheses**

The experimental design

We compared the two models in a controlled experimental setting, in which a certain case (reality) is represented by means of two equivalent conceptual schema diagrams. We examined user comprehension as follows: two groups of users, each studying a different model, were given a diagram of the model they studied, and a set of identical questions (statements) about facts in the schema. The research model is depicted in Figure 3. The independent variable is the models, i.e., an EER diagram and an OO diagram, each represent the same reality. (For the experiment we use more detailed, non trivial schemas, including many entities, or objects, and various relationships.) The two controlled variables are users and tasks.

The users were students of behavioral science and management, an interdisciplinary program at our university. They were taken from the same class, have similar backgrounds, experience and are enrolled in the same course of studies. They are not computer majors, but all took the same three computer-related courses: Introduction to Computers and Data-Processing, Use of Micro-Computers for Management (including practice with spreadsheet, 4GL and a database package) and Information Systems Implementation. Therefore, the students should be considered as “sophisticated users”; i.e., users who can play an active role in systems development by interacting with professional analysts, expressing information needs and approving the analysis and design products. The students were randomly divided into two groups, one studying the EER model and the other the OO model. The studies emphasized comprehension of the schema diagrams, rather than design of diagrams, since we are concerned with user comprehension.

The task of each user was to answer a questionnaire that consists of a set of “true” and “false” statements about facts in the conceptual schemas. The statements were classified according to different constructs in the schemas. They included statements about attributes of entities or objects, about binary-relationships, ternary-relationships, and two relationships, and about other facts with no direct relationships. The purpose was to find out not only if there is a difference in overall comprehension between the two schema diagrams (as reported in Palvia et al., 1992) but also if there is a difference in the specific categories of constructs of each model. The user was expected to review the schema diagram according to the model he studied, and to mark each statement “true” or “false”.

The level of comprehension was measured by counting the number of correct answers. This was done per statement, per category, and for all statements together. Based on the average scores of the users in each group we determined if there are significant differences in comprehension.

The hypotheses

The main question addressed is which of the two diagrammatic models is more comprehensible. Although the representation of entities is different than that of objects, and although there is a difference in the representation of relationships, we have no a-priori reason to assume that one diagram is “better” than the other. Therefore we hypothesized that there would be no difference in comprehension. As mentioned, we distinguished between overall comprehension that is based on all statements together, and comprehension of specific categories of constructs. For each of the following null (H0) hypotheses, there is an opposite (H1) hypothesis indicating that there is a difference:

a) overall: there is no difference in the overall comprehension of the two models;
b) attributes: there is no difference between the two models when attributes of entities or objects are involved.
c) binary-relationships: there is no difference between the two
The Experiment

The schema diagrams

We prepared two equivalent schema diagrams according to the principles of each model, as explained in Section 3. The diagrams included exactly the same facts, and involved all types of constructs for which we have formed hypotheses. They are shown in Figures 4 and 5. In the EER diagram there are 12 entities, 7 binary-relationships, and 2 ternary-relationships. The relationships have different cardinalities, and the entities or relationships have various attributes. There is also a supertype-subtype hierarchy. In the OO diagram there are 14 objects, each with various types of attributes: atomic, multi-valued (sets), reference-attributes, tuples, and appropriate links between the objects. Note that two objects represent ternary relationships. Although we cannot say that these...
Table 1: Examples of True (T) and False (F) Statements Classified by Types of Constructs

Attributes:
- (T) “an employee can be identified by SSN or emp-no”.
- (F) “two suppliers may have the same name”.

Binary-relationships:
- (T) “an order includes a list of parts and quantities ordered”.
- (F) “one can tell all dates a certain department contacted its suppliers last year”.

Two relationships:
- (T) “one can tell to which department belongs the employee who heads a certain project”.
- (F) “an order is sent to one supplier but includes parts that are ordered for different departments”.

Ternary-relationships:
- (T) “an employee may perform many tasks on a certain project”.
- (F) “an employee may work on many projects in a certain city”.

Other facts:
- (T) “an order that is issued to a supplier may include parts which that supplier cannot supply”.
- (F) “the cost of a project may not exceed its department’s budget”.

Figure 5: OO Diagram
diagrams represent large-scale situations, they are by no means trivial (especially as compared to the cases from different studies discussed earlier).

The questionnaire
The questionnaire consisted of 48 statements. Twenty one of the statements are “true” and 27 are “false”. (Any statement can be phrased in both ways, but since we test comprehension we chose to phrase statements in both forms. After the test, we examined whether there is a significant difference in correct answers to “true” or “false” questions, and found none.) With respect to the types of hypothesized constructs, the 48 statements are classified as follows: 9 are on “attributes” of entities or objects, 11 are on “binary-relationships”, 9 on “two relationships”, 9 on “ternary-relationships”, and 10 are classified as “other facts”. Table 1 presents 10 statements for example, two in each category: one “true” and the other “false”. In order to avoid bias because of order, we prepared 5 different sets of questionnaires, each with a different ordering of the statements.

Training of the participants
The subjects were students of behavioral science and management who took the same courses. The experiment took place at the end of the course entitled Information Systems Implementation, the third required computer course, where they studied (among other topics) the systems development life cycle, basics of systems analysis, database models, and the relational model. The class was randomly divided into two groups: one including 41 students and the other 37 (the small difference is due to an administrative error). The same instructor taught one group the EER model and the other group the OO model. We used the same instructor to avoid bias in teaching quality. The same amount of time — 1.5 hours — was allocated to teach each model, and the schemas of Figures 1 and 2 were used as examples. In order to motivate the students, they were told in advance that their performance in the experiment would be considered as part of the final course grade.

The experiment
The experiment took place two weeks after training. Each participant was given a diagram, according to the model s/he studied, and a questionnaire. (Recall that we prepared five sets with different orderings of statements.) The students were allowed to use their class notes and examples. We allocated 50 minutes to complete the task. This was based on an earlier pilot study with other participants. Actually, we allowed ten extra minutes to enable all participants to complete the task.

Table 2: Summary of the Results

<table>
<thead>
<tr>
<th>Construct</th>
<th>No of Questions</th>
<th>Model</th>
<th>Mean Score</th>
<th>Mean in %</th>
<th>STDS</th>
<th>Value of Z</th>
<th>Significance (at α=0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes</td>
<td>9</td>
<td>EER</td>
<td>7.83</td>
<td>87.00</td>
<td>1.34</td>
<td>-0.37</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OO</td>
<td>7.73</td>
<td>85.89</td>
<td>0.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binary-Relationships</td>
<td>11</td>
<td>EER</td>
<td>9.98</td>
<td>90.73</td>
<td>0.96</td>
<td>+0.32</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OO</td>
<td>10.05</td>
<td>91.36</td>
<td>0.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two Relationships</td>
<td>9</td>
<td>EER</td>
<td>6.73</td>
<td>74.78</td>
<td>1.23</td>
<td>+0.83</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OO</td>
<td>6.95</td>
<td>77.22</td>
<td>1.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ternary-Relationships</td>
<td>9</td>
<td>EER</td>
<td>8.20</td>
<td>91.10</td>
<td>0.80</td>
<td>-8.55</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OO</td>
<td>6.10</td>
<td>67.80</td>
<td>1.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Facts</td>
<td>10</td>
<td>EER</td>
<td>7.41</td>
<td>74.10</td>
<td>0.89</td>
<td>+3.47</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OO</td>
<td>8.35</td>
<td>83.50</td>
<td>1.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>48</td>
<td>EER</td>
<td>41.02</td>
<td>83.71</td>
<td>2.49</td>
<td>-1.4</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OO</td>
<td>40.03</td>
<td>81.69</td>
<td>3.64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: EER Model vs. OO Model

User Comprehension

![Graph showing User Comprehension](image-url)
(Note that we did not measure the time to complete the task, but we did limit the time frame; otherwise it is plausible that time would compensate for lack of understanding.)

Analysis of Results and Conclusions

For each user we calculated the number of correct answers (true and false) within each of the 5 categories of statements, and for all statements together. Then we summed and averaged the grades of all members within each group. The results are summarized in Table 2. For each category (including “All”) and each model we show the mean grade in absolute values and in percentages, and the standard deviation. The last two columns indicate the significance of difference between the mean grades. Figure 6 shows the main results in the form of a bar diagram.

To test the significance of difference of means we conducted two-tail Z tests at a level of confidence $\alpha = 0.05$. The column before last in Table 2 shows the Z values of the tests, and the last column indicates the significance of difference. Overall, we cannot reject the null hypothesis that there is no difference of comprehension between the two models, although there is a slight difference in favor of the EER model: average grade of 83.71% for EER compared to an average grade of 81.69% for OO. A closer look at the specific types of constructs that have been examined reveals the following: EER has a slight, insignificant advantage for “attributes”, and OO has a slight, insignificant advantage for both “binary-relationships” and “two relationships”. There is a significant difference in favor of EER for “ternary-relationships”: an average grade of 91.1% for EER vs. only 67.8% for OO. On the other hand, there is an advantage to OO for “other facts”: 83.5% to OO vs. 74.1% to EER.

The relatively slight differences between the models with respect to “attributes”, “binary-relationships” and “two relationships” can be explained by the similarity in diagrammatic representation. As would be expected, in both models the statements involving two binary-relationships are more difficult to comprehend than statements involving just attributes or single relationships, binary or ternary — average scores were only 74.78% for EER and 77.22% for OO.

Since we found a significant difference between the models in the case of ternary-relationships, we further examined these results by distinguishing between statements dealing with m:n:1 (many-to-many-to-one) relationships and statements dealing with m:n:n (many-to-many-to-many) relationships. In both cases we found a significant difference in favor of the EER model. In the case of m:n:1, the average score of EER was 93.33, compared to an average score of only 78.33 for OO. In the case of m:n:n the average scores were 90.67 and 62.17, respectively. The advantage of EER over OO in comprehension of ternary-relationships can be explained by the lucidity of the diamond symbol which connects the three entities. It seems that this symbol is clearer and more visible than the “normal” object rectangle which is used in OO for any type of object. In other words, in OO there is no special highlight of the compound, ternary-relationship. Another explanation may be that in the OO model each object includes sets of tuple attributes, which seem to be more complicated to comprehend compared to the EER model, where the diamond representing the ternary-relationship has only its own attributes attached to it, while every other entity (rectangle) is attached to its own attributes. It turned out that this complication is insignificant in binary-relationships, but is effective when it comes to ternary-relationships. We did not have more complicated relationships (e.g. 4-arys) to further examine these results.

With respect to “other facts”, OO turned out to be superior. This is actually not a specific category as it includes different questions which deal with entities or objects that exist in the diagram but have no direct relationships. Nor can the answer to the question be deducted from the given diagram. OO may be easier to comprehend for this category because most of the information in this model is encapsulated within the objects, as opposed to EER where it is “spread-out” among the entities and relationships.

In summary, the results of the experiment do not clarify which of the two models is easier to comprehend. Overall, the hypothesis that there is no difference cannot be rejected. Since we found a significant weakness of OO in comprehending ternary-relationships, we propose to improve the OO diagram by using a different special symbol (for example, a rectangle surrounded by a circle) for an object that represents a ternary or higher-order relationship. This will enable users to easily identify objects that encapsulate “complex” facts, and thus distinguish them from more “simple” objects.

From the point of view of users, EER diagrams can be used to describe the conceptual schema even when using the OO approach to systems development. (This conclusion does not suggest that the structural approach to systems development is superior to other approaches, for example, the functional approach.) This conclusion applies to user comprehension only; whether there is a difference between the two models from the point of view of designers is an important issue that will be addressed in a separate study.

Our conclusions are based on a specific experimental setting, i.e., certain tasks, models and subjects, as described above. With respect to tasks, the schemas we used were not trivial in terms of size and complexity (especially when compared to previous studies). The subjects were not novices and certainly can represent users who take part in schema validation, and the models were non-trivial versions of ER and OO models, and we examined comprehension of their different constructs. A point that deserves emphasis is that the OO model enables us to represent relationships in different ways, as discussed in Section 3.2. However, in this experiment we adopted specific representations. For example, an m:n relationship is represented by two objects having mutual reference.
attributes and a link indicating cardinality. These representations are simple and complete, and are commonly used by OO modelers. Yet, we did not examine (compare) different, alternative representation (for example, an alternative representation for an m:n relationship could be formed by creating a separate object for the relationship, with reference attributes and links to the involved objects). Although we had to confine ourselves to certain tasks, models and representations, there is certainly room for further experimentation and comparison of different representations of relationships in the OO model.

In this study we have not addressed behavioral aspects of the system. It should be noted, however, that unlike EER schemas, which only describe data structure, OO schemas also encompass “methods”. There is no question as to the advantage of modeling behavior of objects. Since this has not been considered in the current experiment, it may seem unreasonable to draw final conclusions as to which model is preferable. On the other hand, EER schemas may also be augmented with methods, as are OO schemas. There are such studies in the EER literature (e.g., Sakai, 1983; Eder et al., 1986; Markovitz, 1990; Tanaka et al., 1991; Lazarevic and Misic, 1991). Nevertheless, as the results are not conclusive, there is room for further experimentation and comparison of the two models. Such experiments should be from the point of view of both users and designers, and should be based on larger and more complex schemas. They should be directed to comparing different representations of relationships in the OO model, and should include methods attached to OO objects and EER entities.

References


Related Content

Ontology Based Object-Oriented Domain Modeling: Representing Behavior
www.irma-international.org/article/ontology-based-object-oriented-domain/3400/

Development of Interactive Web Sites to Enhance Police/Community Relations
www.irma-international.org/chapter/development-interactive-web-sites-enhance/18565/

Model Driven Engineering for Quality of Service Management: A Research Note on the Case of Real-Time Database Management Systems
www.irma-international.org/article/model-driven-engineering-for-quality-of-service-management/178634/

Thirty Years Later: Some Reflections on Ontological Analysis in Conceptual Modeling
www.irma-international.org/article/thirty-years-later/181666/

Interoperability of B2B Applications: Methods and Tools
www.irma-international.org/chapter/interoperability-b2b-applications/4380/