ENTITY-ANALYSIS AND VIEW-INTEGRATION
DATABASE DESIGN METHODOLOGIES:
A COMPARATIVE STUDY

Robert C. Nickerson
San Francisco State University
College of Business
1600 Holloway Avenue
San Francisco, CA 94132
rnick@sfsu.edu

All rights reserved. Not for publication without permission.
Abstract

An experiment was conducted to compare the effectiveness of the entity-analysis and view-integration database design methodologies. In the experiment, students in a database design course were divided into two groups. Each group received training in one of the design methodologies. Then the students were given identical database design problems and instructed to solve the problem using the methodology in which they were trained. Students using the entity-analysis database design methodology produced poorer designs and took longer on the average than students using the view-integrated methodology but the differences were not statistically significant. Students using the entity-analysis approach were more likely to make certain types of errors than those using the view-integrated approach. Finally, differences were noted in which steps in the design process each group considered hardest and easiest. Because the sample size was small, definitive conclusions cannot be reached.
1. Introduction

There are two general approaches to database design. In the first, the database application area is considered as a whole. Entities, attributes, and relationships for a single conceptual view (schema) are defined. User views (subschema) are then derived from the conceptual view. We call this the entity-analysis approach.

In the second approach, each user requirement in the application area is considered separately and a user's view for each requirement is developed. The separate user views are then integrated to produce a single conceptual view covering all user requirements. We call this the view-integration approach.

To date, no comparison has been made of these two approaches. An important question is whether, in a specific database design situation, they are equally effective at producing acceptable designs. If they do not produce satisfactory designs, then the question becomes whether there are common inconsistencies in the designs produced by them. The objective of this research is to compare the effectiveness of the entity-analysis and the view-integration database design methodologies.

To investigate these questions an experiment was conducted in which identical database design problems were administered to two groups of students in a database design course. The students in one group were trained in the entity-analysis approach and the students in the other group were trained in the view-integration approach. Each student was instructed to apply the approach in which he or she was trained to the design problem. The students' designs were then evaluated to determine their correctness and to look for common errors. Finally, the relationship between the students' designs and the approach they used was investigated.

2. The Entity-Analysis and View-Integration Approaches

Numerous papers have been written on the entity-analysis and view-integration database design methodologies. For example, Teorey et. al. (1986) describes the entity-analysis approach and Batini et. al. (1986) provides a thorough summary of the view-integration approach. Although other tools can be used, we choose to represent the database design produced by either methodology in an (extended) entity relationship or ER diagram. (See Teorey et. al. (1986).)

Appendix A gives the details of the steps in the two approaches. Basically the entity-analysis or EA approach is a top-down design methodology. The designer starts by identifying entities of interest to all users and listing the attributes of each entity. He or she then identifies hierarchies (generalizations and subsets) and draws an ER diagram for each hierarchy. Next the designer identifies relationships, drawing an ER diagram of each, and then eliminates redundant relationships. Finally, the designer draws an ER diagram for the complete design from the ER diagrams for the hierarchies, the ER diagrams for the relationships, and the list of attributes.

On the other hand, the view-integration or VI approach is a bottom-up design methodology. The designer starts by using the EA approach to design a separate database (user view) for each user's or user group's requirements, drawing an ER diagram for each design. Then
the designer integrates or merges the user views, two at a time, to produce the ER diagram for
the complete design. In integrating the user views, the designer may have to rename entities and
attributes, convert an attribute to an entity, identify new generalizations and subsets, and
eliminate redundant relationships.

3. Experimental Design

The students selected for the experiment were those in a graduate seminar in database
design at San Francisco State University. There were thirteen students in the seminar. The
students were ranked based on an admission index and grades in previous graduate major
courses. Then students with similar rankings were randomly placed in one of two groups. This
procedure was designed to reduce the effect of the small sample size and to create two groups of
students with similar performance potential. The groups were randomly assigned the EA or VI
approach.

Each group then met separately with the researcher for a lecture on their respective
design approaches. Each lecture outlined the steps in one of the approaches given in Appendix A
and demonstrated the approach with an example.

At the end of the lecture each student was given an envelope containing instructions, a
database design problem, a series of worksheets, a final solution sheet, and a comment sheet. The
database design problem was the same for both groups but the other sheets depended on which
approach was to be used. (See Appendices B and C.) All material in the envelope was identified
by a student number. A research assistant assigned the student numbers so that the researcher
was not aware of which student had which number.

Each student then completed the database design problem and turned in the result in a
blank envelope. The research assistant removed the final solution sheet for each student. This
sheet did not identify the design approach used and only identified the student by number. The
researcher then evaluated each final solution sheet using a key that provided for precise scoring
of each solution. The maximum number of points for the solution was 135. Finally, each final
solution sheet was examined for the types of errors the student made. A summary list of these
errors was prepared to look for common errors.

After all analysis of the final solution sheets was completed, the other sheets in each
student's packet were examined for any additional insights that could be gained.

4. Analysis of Results

Table 1 gives the total score of each student in each group along with the mean and
standard deviation. This table shows that the students who used the VI approach did better on the
average than those who used the EA approach. A t test was done to check for the significance of
the result. The test indicated that the means of the two samples were not significantly different.
Three errors were identified that were made by more than one student:
1. Not recognizing a subset hierarchy that arose out of two separate user requirements.
2. Not including an attribute of a relationship.
3. Not converting an attribute in one user's requirement to an entity in another user's requirement.

Table 2 lists the number of students who made each of these errors in each group. This table shows that the first type of error was about as common with both approaches but the other two types of errors were more common with the EA approach than with the VI approach.

<table>
<thead>
<tr>
<th>Error</th>
<th>EA Approach (N=6)</th>
<th>VI Approach (N=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not recognize a subset</td>
<td>4 (67%)</td>
<td>4 (75%)</td>
</tr>
<tr>
<td>Did not include an attribute of a relationship</td>
<td>4 (67%)</td>
<td>2 (29%)</td>
</tr>
<tr>
<td>Did not convert an attribute to an entity</td>
<td>4 (67%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Table 2. Numbers of students making common errors

On the comment sheet each student was asked to report the amount of time (to the nearest 1/4 hour) that he or she took to complete the problem solution. Table 3 summarizes the results. This table shows that the students who used the EA approach reported that they took longer on the average to complete the problem solution than those who used the VI approach. A t test, however, indicated that the means were not significantly different.
<table>
<thead>
<tr>
<th>Approach</th>
<th>EA</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.25</td>
<td>5.50</td>
</tr>
<tr>
<td></td>
<td>3.25</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>2.75</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>1.75</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Mean</td>
<td>2.42</td>
<td>1.93</td>
</tr>
<tr>
<td>SD</td>
<td>1.22</td>
<td>1.62</td>
</tr>
</tbody>
</table>

Table 3. Students’ self-reported times required to complete solution (hours)

The comment sheet also included questions about which steps in the design process were the hardest and which were the easiest to apply. Table 4 lists the responses given by at least three students in a group. The table shows that the EA group considered identifying relationships the hardest step to apply whereas the VI group considered this step the easiest to apply.

<table>
<thead>
<tr>
<th>Approach</th>
<th>EA</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardest steps to apply</td>
<td>• Identifying relationships</td>
<td>• Superimposing ER diagrams of user’s views</td>
</tr>
<tr>
<td>Easiest steps to apply</td>
<td>• Identifying entities</td>
<td>• Identifying relationships</td>
</tr>
<tr>
<td></td>
<td>• Identifying attributes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Identifying hierarchies</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Hardest and easiest steps to apply

5. Conclusion

Because of the small sample size, we cannot reach definitive answers to the original questions of the research. Students using the entity-analysis database design methodology produced poorer designs and took longer on the average than students using the view-integrated methodology but the differences were not statistically significant. Students using the entity-analysis approach were more likely to make certain types of errors than those using the view-integrated approach. Finally, differences were noted in which steps in the design process each group considered hardest and easiest. Further research with a larger sample is needed to reach a definitive conclusion.

6. Acknowledgment

I am grateful for the enthusiastic participation in this research project of the students in BICS 862 at San Francisco State University.
7. References


Teorey, Toby J., Dongqing Yang, and James P. Fry, "A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model," ACM Computing Surveys, 18, 2 (June 1986), 197-222.