An Exploration of Individual Differences in Synthesizing Entity-Relationship Data Models

Dr. Vijay V. Raghavan
Northern Kentucky University, Highland Heights, KY, USA.

raghavan@nku.edu

Abstract

Entity Relationship (ER) modeling (1976) is a popular approach to formulate a conceptual data model for designing properly structured databases. In spite of some criticisms of the model leading to numerous extensions added to the original ER model, it is generally believed to be a method of choice for designing common databases. Not surprisingly, ER Modeling is inextricably a part of all database-design classes. Teachers of this modeling construct often encounter students experiencing problems in synthesizing ER models from verbal or written descriptions. This study explores whether individual differences contribute to such difficulties. Gender, length of Information Technology (IT) experience, length of database experience, length of business experience, national origin and learning styles are hypothesized as the individual differences that might contribute to a student's ability to synthesize a conceptual ER model. Ability to synthesize ER models was evaluated using a textbook type ER modeling problem.

Keywords: ER Modeling, Conceptual Modeling, Learning Styles; Teaching Techniques, Individual Characteristics.

Background

Students learning to synthesize conceptual data models from written and verbal descriptions are sometimes puzzled by the fact that they can develop alternative ER models to represent the same situation. There is an element of subjectivity in interpretting the of user requirements that lead to conceptual data models. Students of this data modeling technique are also expected to make appropriate assumptions to synthesize models to represent a business context. Research involving ER modeling has traditionally focused on certain ER modeling principles and notations (e.g. Bodart et. al., 2001). The question of why some are able to synthesize ER models from verbal and written descriptions fairly easily as others find it inherently difficult, has not been addressed. This study explores individual differences in synthesizing ER models.

Connolly and Begg (2002) assert that to gain an understanding of the nature of the data and how it is used by the enterprise, we need to have a model for communication that is non-technical and free of ambiguities. A perspective to the contrary is offered by Kimball (1997) who claims that ER modeling is not a beneficial technique especially for data warehouses. He argues that ER modeling is a discipline used to represent associations among data. He argues that ER model is focused on removing redundancy in data. Hence, designers loose sight of an important goal of data modeling—to create databases that can be queried efficiently. This study, without taking a position on whether ER modeling as an appropriate technique for designing all databases, recognizes that it is still one of the popular approaches for teaching conceptual modeling. The purpose of this paper is to explore whether there are any individual differences between those who can correctly conceptualize data models from written and verbal descriptions and others who seem to have difficulties in synthesizing ER models.

Theoretical Foundations

Although different investigators have differed slightly in the labels and meaning of personality factors, in general, research finds five major dimensions of personality that provide distinctive behavioral clusters. These include Extroversion, Likeability, Conscientiousness, Emotional Stability, and Intellectual Curiosity/Creativity (Loehlin, 1992). "Considerable confusion appears in the literature concerning the terms *cognitive style* and *learning style*. Numerous authors use the terms interchangeably. Garity (1985) notes that learning style has been used as a description for the cognitive process of thinking, perceiving, and remembering. McFadden (1986) states that most definitions of learning style as well as cognitive style, illustrate variations in individual information processing and that no single definition for learning style or cognitive style has been identified (Heineman, 1995)." However, Felder and Silverman's (1988) model of teaching and learning styles is an important contribution in the area of engineering education. Their index of learning styles is widely used and has been translated into many languages. Felder made two recent changes to the original model and it currently includes the following dimensions: sensory-intuitive, visual-verbal, active-reflective, and sequential-global.

Learning Styles: Felder (1993) identifies four dichotomous dimensions of learning style. They are sensing/intuitive, visual/verbal, active/reflective, and sequential/global. Felder's model of learning style is summarized in Table 1:

Goal of Measurement	First Dimension	Second Dimension
Type of information preferentially perceived	Sensory: sights, sounds, physical sensations	Intuitive: memories, ideas and insights
Modality of perception of sensory information	Visual: pictures, diagrams, graphs, demonstration	Verbal: sounds, written and spoken words.
Preference in processing information	Actively: through engagement in physical activity or discussion	Reflectively: through introspection
Progress towards understanding	Sequentially: in a logical progression of small incremental steps	Globally: in large jumps, holistically

Table1: Model of Learning Styles: Adapted from Felder (1993)

Felder's Learning Styles (2002) includes a preface that explains the changes made to the model since 1988 and the rationale for such changes. As a result of its popularity and widespread use in a closely allied field, it was decided to use Felder's Learning Styles index in this study.

Participant's prior exposure to the IT work environment may also positively influence the learning outcome of training to formulate ER models. Prior Exposure to the IT environment within an organization, direct exposure to database technologies or general exposure to real world business problem contexts as identified by business work experience in general may favorably influence a participant's ability to develop ER models. Another demographic factor included this study is "gender." Research finds that the only consistent difference in mental abilities between men and women is in spatial abilities (Jacklin, 1989). However, girls continue to be concentrated among jobs historically held by women and more oriented toward working with people (Beutel & Marini, 1995). Since ER modeling is a logical, deductive and analytical process it is appropriate to include gender as an independent factor in our study. Thus, gender, experience of the participant in any area of IT, and learning styles are used as independent variables that influence the participants ability to synthesize a correct ER model from a written description of a problem context. (See Figure 1.)

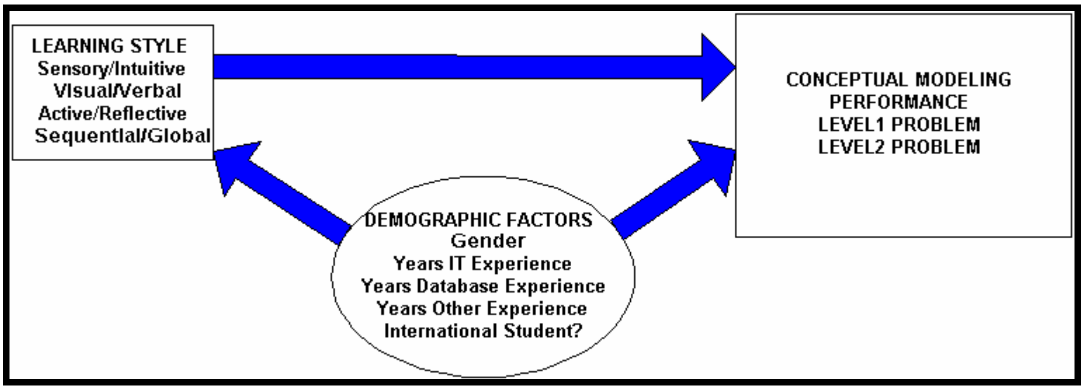

Figure 1: A Model of Individual Factors, Learning Style and Conceptual Modeling Performance

Experimental Design

Participants in the study were students enrolled in a Master's Level Database Class at the Northern Kentucky University, KY, USA. There were 37 students enrolled in the fall semester of 2002. The experiment was conducted as part of instruction during two regular class meetings. The problems themselves were not graded for the course but only as a tool to provide feedback on their learning of ER modeling concepts and their ability to draw ER diagrams based on a problem context. The students were preparing for future tests based on this experience. This provided enough motivation for all participants to take the experiment seriously. There were two problems selected from Connolly and Begg: one at an elementary level and the other at an advanced level. The two problems selected for the experiment are reported in Appendix. A.

It is beneficial for students to first understand the purpose and benefits of a high-level conceptual data model such as the ER model to design a database. Before introducing the concepts associated with the ER model, it is helpful for students to see a possible end product of ER modeling. Hence the participants were first shown an Enterprise Model and the instructor emphasized that the diagram seen is the end product. In the real world, developing an ER model is often a process where participants work in groups. Working in groups exposes the participants to problems associated with differing interpretation of user requirements. However, the students in this study were not allowed to work in groups as the goal of the current research is to study the individual differences. In addition, the problems used in this study are not as complex as real world models involving numerous entities and relationships. After the instruction phase, had one week to study the concepts from the textbook (Connolly and Begg, 2002) and work through two exercises presented at the end of the relevant chapter. Next week, students were given an elementary level problem and then a more advanced problem (Appendix A) in a regular class meeting. All participants solved the problem in the same room at the same time. A scoring rubric was used where each entity identification was given 2 points and each cardinality specification was given one point. This resulted in a total raw score of 22 points for the Level 1 problem and 28 points for the Level 2 problem.

After the experiment was over students were given the ILS questionnaire (Appendix B) and were asked to give a print out of the web version of their results. This was collated with demographic variables and scoring rubric to create the data set for statistical analyses.

Analysis

There were 14 female students and 22 males students who participated in the study. Sixteen of the thirty-six students were international students. Eighteen of the thirty-six students had no prior database experience. Others had one or more years of database experience. Twelve students had no Information

Technology experience. Ten participants had no prior work experience. All statistical analyses were conducted using SPSS version 11.5. At this stage of data collection, due to a relatively a small size, simple Chi-square tests to study the association between various individual factors and ER Modeling performance was preferred to other robust procedures that would provide greater insight into the relationships. Also, the scoring on the dependent variable was categorized using the SPSS procedure to provide two distinct categories to study the association of demographic factors. Table 2 presents the results of this analysis. The continuous scale ER Modeling scores was retained to compare means for different learning style dimensions.

H_1: Database experience and ability to Synthesize ER models are related

H_2: IS experience and ability to Synthesize ER models are related

H_3: General business experience and ability to Synthesize ER models are related.

H_4: Participant's national origin and ability to Synthesize ER models are related.

H_5: Gender and ability to Synthesize ER models are related.

Factor Studied	Level 1 Problem			Level 2 Problem		
	N	**Chi-Square**	**Significance**	**N**	**Chi-Square**	**Significance**
Database Experience	35	0.696	0.404	34	11.691	.001*
IS Experience	35	0.801	0.371	34	0.002	.968
Other Experience	35	0.218	0.640	34	0.537	.464
International Student	35	1.621	0.203	34	3.316	.069*
Gender	35	0.551	0.458	36	0.808	.369

Table 2: Results of Tests of Association of Demographic Factors

Hypotheses H_1 is supported at 0.05 significance level while H_4 is supported at .1 significance level for Level 2 problem which was at an increased level of complexity. Independent Sample T-tests were performed by dividing the experimental group based on the four dimensions of learning style. The results are reported in Table 3.

H_6: Mean scores in ER Synthesizing is different for active and reflective learners.

H_7: Mean scores in ER Synthesizing is different for sensing and intuitive learners

H_8: Mean scores in ER Synthesizing is different for visual and verbal learners.

H_9: Mean score in ER Synthesizing is different for sequential and global learners.

Learning Style Dimension	Level 1 Problem		Level 2 Problem	
	t-value	Significance	t-value	Significance
ACT-REF	0.663	0.516	1.86	.080*
SEN-INT	2.517	0.022**	0.787	.442
VIS-VER	Not Analyzed		Not Analyzed	
SEQ-GLO	-0..316	0.757	0.816	.428

Table 3: Results of Comparison of Means for Different Learning Style Dimensions

Hypotheses H_6 was supported for level 2 problems while H_7 is supported for Level 1 problems. There was insufficient data to test whether ER Synthesizing means are different for visual and verbal learners as only one participant scored on verbal end of the spectrum.

Discussion and Conclusion

Although a sample size of 30 is enough to assume normal distribution of ER synthesizing scores, any interpretation of the support of these hypotheses should bear in mind the relatively small sample size. In addition, the finding suggests that in general, as the problem complexity increases individual differences may become more of a factor. Hence, our discussion focuses mainly on interpreting the results for the Level 2 problem.

Prior exposure to the database environment was found to have significant association with the participant's ability to synthesize ER models. Intuitively prior exposure to a database context is likely to enhance participant's understanding of the experimental problem. Exposure to the database context may have sensitized the participants to take the problems solving task more seriously than others and hence they may have attempted the solution with greater rigor. The difference in scores due to national origin should not be interpreted as their inherent ability or lack thereof to synthesize ER models. Since the experimental problem was a textual description relating to a geographical context, difference in the ability to understand the context might have contributed to a difference in scores. Finally, active learners scored better than reflective learners in the experimental problem. The mean score for active learners was 22.27 while that of reflective learners was 20.75 in the Level 2 problem. Active learners learn by actively engaging in the process of learning through discussing or applying a concept. Reflective learners learn by just thinking about a context. This finding supports that ER modeling is likely a concept learnt by actively studying it rather than just thinking about it.

This finding on learning style also suggests that the trainers may need to tailor their training style to suit to different types of learners. As an example, a training method that uses group exercises may facilitate learning by certain types of users. Also reflective learners may benefit by exposure many example problems. Although this study did not vary the training methods, incorporating alternative training approaches will be a useful future extension of this study. Also additional data collection is necessary to use more robust statistical analyses. Efforts to increase the sample size are currently under way.

References

Anselin L., & Cho, W.K. (2002). Tam Conceptualizing Space: Reply . *Political Analysis* 10, 3 (Summer 2002): 301-303. Retrieved November 13, 2002, from the World Wide Web http://cho.pol.uiuc.edu/wendy/papers/acrej.pdf

Beutel, A., & Marini, M. (1995). Gender and values. *American Sociological Review*, 60, 436-448.

Bodart, F., Patel A., Sim M., & Weber R. (2001). Should optional Properties be used in Conceptual Modeling? A Theory and Three Empirical Tests. *Information Systems Research*, Vol.12. No.4. December 2001, pp.384-405.

Chen, Peter P. (1976). The Entity-Relationship Model – Toward a Unified View of Data. *ACM Transactions on Database Systems*, 1(1), 9-36.

Connolly, T. & Begg, C. (2002). *Instructor's Guide to Database Systems: A Practical Guide to Design, Implementation and Management*. Harlow: England.

Dunn, R. & Dunn, K. (1978). *Teaching students through their learning styles: A practical approach.* Prentice Hall.

Felder, R.M. & Silverman, L.K. (1988). Learning and Teaching Styles in Engineering Education. *Journal of Engineering Education*., 78(7), 674-681. Retrieved November 13, 2002, from the World Wide Web http://www.ncsu.edu/felder-public/Papers/LS-1988.pdf

Felder, R. (1993) Reaching the Second Tier: Learning and Teaching Styles in College Science Education. *J. College Science Teaching*, *23*(5), 286-290 (1993).

Heinman, P.L. (1995). Cognitive Versus Learning Style. Retrieved November 13, 2002, from the World Wide Web http://www.personality-project.org/perproj/others/heineman/cog.htm

Howell, J. & Erickson, M. (1978). *Matching teacher and learning styles.* Toronto, Ontario: Educational Research.

Jacklin, C. (1989). Female and male: Gender issues. *American Psychologist*, 44, 127-133

James, W.B., & Blank, W.E. (1993). Review and critique of available learning style instruments for adults. In Flanner, D. (Ed.), *Applying cognitive learning theory to adult learning* (pp.47-57), New Directions for Adult and Continuing Education, No. 59, Fall, 1993. San Francisco: Jossey-Bass.

Kimball, R.A. (1997). Dimensional Modeling Manifesto: Drawing a Line Between Dimensional Modeling and ER Modeling Techniques, Vol.10 No.9 – August 1997 Retrieved November 13, 2002, from the World Wide Web http://www.dbmsmag.com/9708d15.html

Kolb, D. (1985). *Learning style inventory.* Boston, MA: McBer.

Loehlin, J. (1992). Genes and environment in personality development. Newbury Park, CA: Sage Publications.

McFadden, E.A. (1986). *Clinical decision making and its relationship to learning style and personality type*. Unpublished doctoral dissertation, University of Maryland.

Soloman B.A. & Felder R.M. (1999). *Index of Learning Style Questionnaire*. Retrieved from http://www.ncsu.edu/felder-public/ILSdir/ilsweb.html on August 10, 2002.

Biography

Vijay Raghavan is Associate Professor of Information Systems at the Northern Kentucky University. He received his PhD from the Kent State University. His research interests are in the areas of software development process and issues, organizational impact of new information technologies and IT education. He has published in *the Journal of Systems Management*, *Journal of Information Technology Management*, and *Managing Software Development*. He has widely consulted with "*Fortune-100*" firms in the areas of database and intranet technologies. His current research projects include formation of and issues relating to virtual teams, offshore vendor selection process, and security issues during software development.

Appendix A

Level1 Problem

Both these problems are adapted from Connolly and Begg (2002).

Pets World Case Study

Represent each of the following requirements with an ER diagram.

(a) A company called Pets World runs a number of clinics in the Greater Cincinnati area. Each clinic has many staff assigned to it but only one member of staff manages at most one clinic (not all staff manage clinics). Each clinic has a unique clinic number (clinicNo) and each member of staff has a unique staff number (staffNo).

(b) When a pet owner contacts a clinic, the owner's pet is registered with the clinic. An owner can own one or more pets, but a pet can only register with one clinic. Each owner has a unique owner number (ownerNo) and each pet has a unique pet number (petNo).

(c) Combine a and b into one ER diagram to represent the entire firm.

Level 2 Problem

Reliable Rentals Case Study

The requirements collection and analysis phase of the database design process has provided the following data requirements for a company called Reliable Rentals, a private rental company which rents out vehicles (cars and vans). The Company has various outlets (garage/offices) throughout the Greater Cincinnati area. Each outlet has a number, address, phone number, fax number, and a manager who supervises the operation of the garage and offices at each site.

Each site is allocated a stock of vehicles for rent, however, individual vehicles may be moved between outlets, as required. Only the current location for each vehicle is stored. The registration number uniquely identifies each vehicle for rent and is used when hiring a vehicle to a client.

Clients may rent vehicles for various periods of time (minimum 1 day to maximum 1 year). Each individual rent agreement between a client and the Company is uniquely identified using a rent number. Information stored on the vehicles for rent include: the vehicle registration number, model, make, engine size, capacity, current mileage, date MOT due, daily rent rate, and the current location (outlet) of each vehicle.

The data stored on a rent agreement includes the rent number, the client's number, name, address and phone number, date the client started the rent period, date the client wishes to terminate the rent period, the vehicle registration number, model and make, the mileage before and after the rent period. After each rental a member of staff checks the vehicle and notes any fault(s). Fault report information on each vehicle is stored, which records the name of the member of staff responsible for the check, date checked, whether fault(s) where found (yes or no), the vehicle registration number, model, make and the current mileage.

The Company has two types of clients: personal and business. The data stored on personal clients includes the client number, name (first and last name), home address, phone number, date of birth and driving license number. The data stored on business clients includes the client number, name of business, type of business, address, and telephone and fax numbers. The client number uniquely identifies each client and the information stored relates to all clients who have rented in the past and those currently hiring a vehicle.

Information is stored on the staff based at various outlets including: staff number, name (first and last name), home address, home phone number, date of birth (DOB), sex, National Insurance Number (NIN), date joined the Company, job title and salary. Each staff member is associated with a single outlet but may be moved to an alternative outlet as required, although only the current location for each member of staff is stored.

Create a conceptual schema for Reliable Rentals using the concepts of the Enhanced Entity–Relationship (EER) model. To simplify the diagram, only show entities, relationships and the primary key attributes. Specify the cardinality ratio and participation constraint of each relationship type. State any assumptions you make when creating the EER model (if necessary).

Appendix B

Index Of Learning Styles*

Barbara A. Soloman
First-Year College
North Carolina State University
Raleigh, North Carolina 27695

Richard M. Felder
Department of Chemical Engineering
North Carolina State University
Raleigh, NC 27695-7905

Directions

Circle "a" or "b" to indicate your answer to every question. Please choose only one answer for each question.
If both "a" and "b" seem to apply to you, choose the one that applies more frequently.

1. I understand something better after I
   (a) try it out.
   (b) think it through.

2. I would rather be considered
   (a) realistic.
   (b) innovative.

3. When I think about what I did yesterday, I am most likely to get
   (a) a picture.
   (b) words.

4. I tend to
   (a) understand details of a subject but may be fuzzy about its overall structure.
   (b) understand the overall structure but may be fuzzy about details.

5. When I am learning something new, it helps me to
   (a) talk about it.
   (b) think about it.

6. If I were a teacher, I would rather teach a course
   (a) that deals with facts and real life situations.
   (b) that deals with ideas and theories.

7. I prefer to get new information in
   (a) pictures, diagrams, graphs, or maps.
   (b) written directions or verbal information.

8. Once I understand
   (a) all the parts, I understand the whole thing.
   (b) the whole thing, I see how the parts fit.

9. In a study group working on difficult material, I am more likely to
   (a) jump in and contribute ideas.
   (b) sit back and listen.

10. I find it easier
    (a) to learn facts.
    (b) to learn concepts.

11. In a book with lots of pictures and charts, I am likely to
    (a) look over the pictures and charts carefully.
    (b) focus on the written text.

12. When I solve math problems
    (a) I usually work my way to the solutions one step at a time.
    (b) I often just see the solutions but then have to struggle to figure out the steps to get to them.

13. In classes I have taken
(a) I have usually gotten to know many of the students.
(b) I have rarely gotten to know many of the students.

14. In reading nonfiction, I prefer
(a) something that teaches me new facts or tells me how to do something.
(b) something that gives me new ideas to think about.

15. I like teachers
(a) who put a lot of diagrams on the board.
(b) who spend a lot of time explaining.

16. When I'm analyzing a story or a novel
(a) I think of the incidents and try to put them together to figure out the themes.
(b) I just know what the themes are when I finish reading and then I have to go back and find the incidents that demonstrate them.

17. When I start a homework problem, I am more likely to
(a) start working on the solution immediately.
(b) try to fully understand the problem first.

18. I prefer the idea of
(a) certainty.
(b) theory.

19. I remember best
(a) what I see.
(b) what I hear.

20. It is more important to me that an instructor
(a) lay out the material in clear sequential steps.
(b) give me an overall picture and relate the material to other subjects.

21. I prefer to study
(a) in a study group.
(b) alone.

22. I am more likely to be considered
(a) careful about the details of my work.
(b) creative about how to do my work.

23. When I get directions to a new place, I prefer
(a) a map.
(b) written instructions.

24. I learn
(a) at a fairly regular pace. If I study hard, I'll "get it."
(b) in fits and starts. I'll be totally confused and then suddenly it all "clicks."

25. I would rather first
(a) try things out.
(b) think about how I'm going to do it.

26. When I am reading for enjoyment, I like writers to
(a) clearly say what they mean.
(b) say things in creative, interesting ways.

27. When I see a diagram or sketch in class, I am most likely to remember
(a) the picture.
(b) what the instructor said about it.

28. When considering a body of information, I am more likely to
(a) focus on details and miss the big picture.
(b) try to understand the big picture before getting into the details.

29. I more easily remember
(a) something I have done.
(b) something I have thought a lot about.

30. When I have to perform a task, I prefer to
(a) master one way of doing it.
(b) come up with new ways of doing it.

31. When someone is showing me data, I prefer
(a) charts or graphs.
(b) text summarizing the results.

32. When writing a paper, I am more likely to
(a) work on (think about or write) the beginning of the paper and progress forward.
(b) work on (think about or write) different parts of the paper and then order them.

33. When I have to work on a group project, I first want to
(a) have "group brainstorming" where everyone contributes ideas.
(b) brainstorm individually and then come together as a group to compare ideas.

34. I consider it higher praise to call someone
(a) sensible.
(b) imaginative.

35. When I meet people at a party, I am more likely to remember
(a) what they looked like.
(b) what they said about themselves.

36. When I am learning a new subject, I prefer to
(a) stay focused on that subject, learning as much about it as I can.
(b) try to make connections between that subject and related subjects.

37. I am more likely to be considered
(a) outgoing.
(b) reserved.

38. I prefer courses that emphasize
(a) concrete material (facts, dat(a).
(b) abstract material (concepts, theories).

39. For entertainment, I would rather
(a) watch television.
(b) read a book.

40. Some teachers start their lectures with an outline of what they will cover. Such outlines are
(a) somewhat helpful to me.
(b) very helpful to me.

41. The idea of doing homework in groups, with one grade for the entire group,
(a) appeals to me.
(b) does not appeal to me.

42. When I am doing long calculations,
(a) I tend to repeat all my steps and check my work carefully.
(b) I find checking my work tiresome and have to force myself to do it.

43. I tend to picture places I have been
(a) easily and fairly accurately.
(b) with difficulty and without much detail.

44. When solving problems in a group, I would be more likely to
(a) think of the steps in the solution process.
(b) think of possible consequences or applications of the solution in a wide range of areas.