Volume 11, No. 1
February 2013
ISSN: 1545-679X

INFORMATION SYSTEMS EDUCATION JOURNAL

In this issue:

The **Information Systems Education Journal** (ISEDJ) is a double-blind peer-reviewed academic journal published by **EDSIG**, the Education Special Interest Group of AITP, the Association of Information Technology Professionals (Chicago, Illinois). Publishing frequency is six times per year. The first year of publication is 2003.

ISEDJ is published online (http://isedjorg) in connection with ISECON, the Information Systems Education Conference, which is also double-blind peer reviewed. Our sister publication, the Proceedings of ISECON (http://isecon.org) features all papers, panels, workshops, and presentations from the conference.

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the ISEDJ journal. Currently the target acceptance rate for the journal is about 45%.

Information Systems Education Journal is pleased to be listed in the 1st Edition of Cabell's Directory of Publishing Opportunities in Educational Technology and Library Science, in both the electronic and printed editions. Questions should be addressed to the editor at editor@isedj.org or the publisher at publisher@isedj.org.

INFORMATION SYSTEMS
EDUCATION JOURNAL

Comparing Top-down with Bottom-up Approaches: Teaching Data Modeling

Hsiang-Jui Kung
hjkung@georgiasouthern.edu
Information Systems, Georgia Southern University
Statesboro, GA 30460, USA

LeeAnn Kung
hzt0007@auburn.edu
Aviation & Supply Chain Management, Auburn University
Auburn, AL 36849, USA

Adrian Gardiner
agardine@georgiasouthern.edu
Information Systems, Georgia Southern University
Statesboro, GA 30460, USA

Abstract

Conceptual database design is a difficult task for novice database designers, such as students, and is also therefore particularly challenging for database educators to teach. In the teaching of database design, two general approaches are frequently emphasized: top-down and bottom-up. In this paper, we present an empirical comparison of students' performance between these two approaches in a conceptual data modeling exercise. Our results indicate that, while prior database education had a significant effect on the quality of design performance, the chosen approach did not. The findings suggest that database educators should integrate both top-down and bottom-up approaches in database design showing the differences and similarities between the two approaches to improve students' learning of data modeling.

Keywords: data model, entity relationship diagram, relational model, normalization

1. INTRODUCTION

Teaching database design remains an important topic as *data and information management* remains a core course in the IS 2010 undergraduate IS curriculum (Topi, Valacich, Wright, Kaiser, Nunamaker, Sipior, & De Vreeda, 2009). While there are many database textbooks devoted to presenting various approaches, methods, and techniques for database design, teaching practices vary considerably, and there is an ongoing debate with regards to the effectiveness of certain approaches both within the classroom and in practice (Fotache, 2006). This paper presents empirical results of an investigation into the effectiveness of two common, but contrasting, approaches to database design (namely, top-down and bottom-up approaches) within a classroom setting.

The database design process aims to create database structures that will efficiently store and manage data (Rob & Coronel, 2004). Database design has four phases: requirements analysis, conceptual design, logical design, and physical design. Notwithstanding, it is common within

Information Systems (IS) university courses in data management to present the primary aim of database design as the development of an acceptable logical data model, i.e., relational schema design. The final stage of database design (physical design) is frequently deemphasized, as IS graduates are normally expected to be less knowledgeable in issues such as the design of indexes and denormalization. Within the field of database design, a recurring distinction is made between top-down and bottom-up approaches. This tradition of duality suggests two different paths towards the development of an acceptable logical data model.

Top-Down and Bottom-up Approaches to Database Design

Top-down approaches stress an initial focus on knowledge of higher-level constructs, such as identification of populations and collections of things and entity types, membership rules, and relationships between such populations. Adoption of a top-down approach will generally start with a set of high-level requirements, such as a narrative. These requirements start a process of identifying the types of things needed to represent data with as well as the attributes of those things, which may become attributes in tables.

In the top-down database design tradition, the database analyst initially attempts to develop a conceptual data model by identifying highly abstracted data objects (things/entity types) that may exist within the domain—i.e., the analyst attempts to construct a domain ontology. Techniques applied by the analyst typically include making observations, conducting interviews, and other data collection strategies. Usually, inspiration for the data model also comes from a close analysis of the domain business rules. In addition, structural properties, such as relationships between entity types and relationship cardinality are identified. In many cases, an initial conceptual data model is drafted that does not include all data attributes. Once a satisfactory conceptual data model has been developed, the database analyst may turn his/her attention to the technological platform on which the final data repository will be deployed (i.e., development of the logical data schema). Development of the logical schema requires the database analyst to consider any mapping issues between the structures on the ER (Entity-Relationship) model and chosen persistent mechanism.

Historically, the most common persistent mechanism used by organizations has been either a relational or object-relational database. Commonly, top-down approaches have utilized diagrammatic approaches, such as conceptual data models (e.g., ER diagrams). Notwithstanding, ER diagrams have also been featured in bottom-up approaches. For example, Shoval, Danoch & Balabam (2004) present a bottom-up approach to developing conceptual data models that produce ER diagrams at increasingly higher levels of abstraction; while Teory, Wei, Bolton & Koenig (1989) present a bottom-up approach based on the principle of entity clustering.

In contrast, bottom-up approaches view database design as proceeding from an initial analysis of lower-level conceptual units, such as attributes and functional dependencies and then moving towards an acceptable logical data model through logical groupings of associated attributes. In other words, bottom-up approaches tend to view the task of population identification as a process of generalizing object identity from examples of structural dependencies (e.g., bundling/categorizing attributes that appear to co-occur). Input into a bottom-up approach, for example, could be views of data, such as screen shots or reports (printouts), or patterns of co-occurring attribute values identified within large datasets. A well-known approach to database design that can be used as a bottom-up approach is normalization (Connolly & Begg, 2000). By addressing potential deficiencies in a relational schema design associated with different levels of normal form, relations are defined to minimize redundancy and dependency. It is also common that normalization is infused with top-down approaches, such as using ER diagrams, as a logical check on the adequacy of the final relational schema.

The distinction between top-down and bottom-up approaches to database design is also highlighted in early theoretical work on conceptual data modeling and database design. Bernstein (1976) pioneered an approach to database design based upon the *synthesis* of relations (*synthesis* in this context relates to its philosophical meaning: "logical deduction"). It is of interest to note that Bernstein's paper, which was published in the same year as Chen's

(1976) seminal work on ER modeling, presented a distinct alternate approach to database design to that proposed by Chen. Although both papers focused on producing provably sound logical database schemas and addressing semantic constraints, Chen's approach can be considered an exemplar of top-down design, while Bernstein's approach presents a bottom-up database design methodology. Bernstein's synthesis approach is clearly predicated upon Codd's (1970) seminal work on normal forms and therefore provided a direct contrast to Chen's (1976) work – Chen's work was actually originally presented as an alternate approach to Codd's (1970) approach to database design, but one "with clearer semantics" and an approach not using "the transformation operation" (Chen, 1976, p.28).

Another form of the top-down versus bottom-up process comes from Hoffer, Ramesh & Topi (2010), who advocate two distinct approaches for identifying supertype/subtype structures within ER diagrams: specialization (top-down) and generalization (bottom-up). With generalization, the design process proceeds in a bottom-up manner, in which multiple entity sets are synthesized into a higher-level entity set on the basis of common features. The process of designating subgroupings within an entity set is called specialization. Choice of technique would depend on "several factors such as the nature of the problem domain, previous modeling efforts, and personal preference." (Hoffer et al., 2010).

Some data management textbooks have been criticized for incomplete and confusing treatment of important concepts within database design, such as definitions of a relation and first normal form (e.g., Philip, 2007). In addition, Fotache (2006) found a degree of confusion with respect to the role and importance of normalization within database design: some popular textbooks on database design did not feature normalization at all, or very little. Moreover, with regards to integrating normalization with top-down approaches, such as using ER diagrams, there are also different approaches and opinions (Fotache, 2006). Another concern is that data management textbooks seldom offer concrete advice as to under which circumstances a specific approach should be applied.

Overall, we contend that with many different opinions of the application of top-down and bottom-up approaches, it is not surprising that students may actually become more confused as to the true merits of each approach and their theoretical distinctions. Moreover, as many data management textbooks fail to clearly acknowledge the strengths and limitations of the top-down and bottom-up approaches, students commonly draw false conclusions that both approaches will always produce the same relational schema design, or that both approaches need to be applied before a final, acceptable relational schema can be produced.

2. RESEARCH QUESTIONS

From a teaching perspective, while most database and systems analysis and design textbooks cover both the ER modeling and the relational data model, it remains unclear as to how to best integrate both of these design methods. In addition, little empirical data exists to substantiate the true strengths and weaknesses of each approach. Such concerns are summarized through the following research questions:

- Does a certain teaching approach, emphasizing either top-down or bottom-up, result in better student database design performance?

- Do students experience difficulty in integrating the two design approaches formulating their final database design?

In this study, we address these research questions by comparing the performance of students across different database design methods, in which either a top-down or bottom-up approach was emphasized (e.g., an ER modeling approach vis-à-vis an approach based upon the relational data model).

The following section of this paper describes the research design and data collection procedure. We then present the data analyses and results of the study. The concluding section summarizes contributions and limitations of the study.

3. RESEARCH FRAMEWORK AND HYPOTHESES

The research framework is shown in Figure 1. Designer performance is the dependent variable, and is measured by error rate. The model predicts that designer performance will be affected by the teaching of data modeling approach and designer experience (course).

Our main interest is to identify any performance differences between the different approaches to the teaching of data modeling (top-down versus bottom-up) and course (Systems Analysis and Design (SA&D), Data Management (DM), and Business Systems Analysis (BSA)). As no prior empirical work has compared the two data modeling teaching approaches directly, it is therefore difficult to predict which approach will result in superior performance; however, given that most textbooks and database-related courses have traditionally emphasized a top-down approach to database design over a bottom-up one, it is plausible to support the notion that novice database designers using the ER modeling approach will perform better than those using the relational model (normalization) approach.

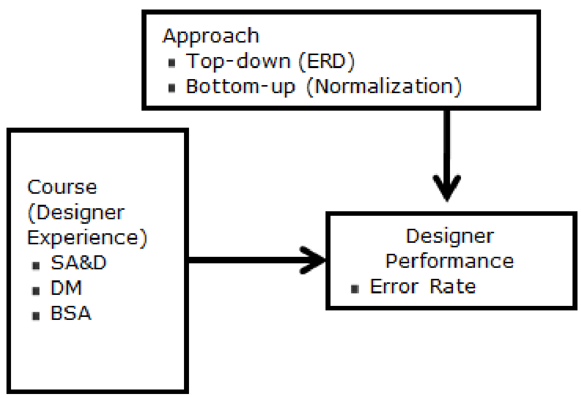


Figure 1: Research Framework

The degree of IS application domain knowledge (e.g., understanding of functional requirements) can potentially affect a designer's ability to design a quality database (Khatri, Ramesh, Vessey, Clay & Park, 2004). The level of database design knowledge is therefore an important indicator of design performance. Subjects in DM and BSA courses have some data modeling experience, while the majority of subjects in SA&D have no such experience. In order to take a DM course, subjects at the studied university had to have completed the SA&D course with a 'C' grade or better. In contrast, BSA is a course for MBA students seeking a concentration in Information Systems (IS). Considering the different levels of database design experience within our subject population, we added "course" as an independent variable to the research framework (see Figure 1).

The hypotheses (presented in null form) addressed in this study are as follows:

H1: No difference in students' performance between the different approaches will exist.

H2: No difference in students' performance across different courses will exist.

H3: No difference in students' performance across different ER modeling constructs will exist.

H4: No difference in students' performance across different relational data model constructs will exist.

4. RESEARCH METHODOLOGY

This study contains two parts. The first part was a laboratory experiment in which subjects were instructed to produce a database schema. The second part required subjects to complete a qualitative survey question, which was used to elicit further information about our subjects' attitudes toward the database design task.

Sample

One hundred and three students enrolled in an undergraduate SA&D and a DM courses, and students enrolled in a postgraduate MBA BSA course completed the in-class exercise. Each undergraduate course had two sections and each section had about the same number of students. Table 1 summarizes the distribution of subjects' demographics.

Table 1: Subjects' Demographics

Course	SA&D	44
	DM	45
	BSA	14
Status	Junior	46
	Senior	43
	Graduate	14
Gender	F	25
	M	78

The SA&D and DM courses are required core courses for the students' study program. The SA&D course is a prerequisite for data management. Most students in SA&D had no prior database design experience. About half of the subjects in the BSA course majored in Information Systems or Computer Science and had taken either one or both SA&D and DM courses in their undergraduate studies.

Procedure

Subjects in each course were exposed to five 75-minute sessions on data modeling processes. In the first session, the instructor explained the purpose of an ERD, including definitions of entity types, relationship, and cardinality. In the second session, the drawing of ERDs from business rules was demonstrated by the instructor and then practiced by subjects. The importance of database normalization was discussed during the third session and normalization techniques to the third normal form (3NF) were demonstrated and practiced in the fourth session. In the fifth session, the instructor explained the value-determined relationship to bridge ER and relational models and applied it to the same examples used in the previous session.

Exercise

In the sixth session, subjects were asked to complete an in-class data modeling exercise. An example of this exercise is presented in Appendix A. After completion of the exercise, subjects answered an open-ended survey question aimed at eliciting their perceptions of the difficulty of the two design approaches.

In the exercise, the top-down approach consisted of students 1) reading a textual description of the domain that identified the applicable business rules; 2) identifying entities; 3) identifying cardinality and relationships; and, 4) drawing a simplified ER diagram without attributes. The top-down exercise is the Step 1 of Appendix A. In the bottom-up condition, students were required to 1) identify domain attributes and consolidate functional dependencies into canonical form based on a given list of domain functional dependencies (FDs); 2) create a normalized relational schema; and 3) draw a final ERD diagram (including attributes) based on relation schemas. The bottom-up is the Steps 2 and 3 of Appendix A. The Appendix B contains the solutions of the top-down and bottom-up exercises. The authors randomly assigned one section of each course to top-down design problem and the other section of each course to bottom-up design problem, and all subjects completed the same problem domain.

Performance (Error Rate)

We operationalized performance as the ratio of incorrect problem domain objects to the total objects of one concept. Thus, for each concept *i*,

$$\text{Performance}_i = \frac{Number\_Of\_Error}{Total\_Object} \quad (1)$$

The in-class exercise was scored according to the number of errors/mistakes in terms of entities, relationships, cardinalities, attributes, normalized relations, and primary keys, with a higher score indicating poorer performance. For example, for the top-down approach, subjects' ERDs should have featured four entities, three relationships, and six maximum cardinalities. The performance is therefore calculated by taking the number_of_error divided by the denominator, thirteen (derived from the sum of four entities, three relationships and six cardinalities). For the bottom-up approach, subjects should have featured four relations: *Patient, Physician, Visit* and *Appointment*. The *Patient* relation has one primary key and four non-key attributes; *Physician* and *Visit* have one primary and two non-key attributes and *Appointment* has two primary keys and one non-key attribute. Therefore, the performance score is determined by the ratio of number of errors to 19 (the denominator 19 was derived form the sum of 14 attributes and 5 primary keys in 4 relations in the third normal form).

5. DATA ANALYSES AND RESULTS

The research design is a 2×3 factorial between subjects and within subjects' methods: their approach (top-down and bottom-up) and the course (SA&D, DM and BSA). Such a design will also reveal whether interactions occur between approach and course (i.e., whether an approach favors a specific level of expertise). IBM SPSS 19 was used to perform the statistical data analysis.

Hypotheses Testing

Hypothesis *H1* predicted that no difference in students' performance between the different approaches will exist. Sixty-seven subjects completed the allocated exercise correctly using top-down approach, while sixty subjects completed the allocated exercise correctly using the bottom-up approach. With zero being the best, Table 2 illustrates that subjects generally produced a higher error rate in the bottom-up design approach (about 19%).

A two-way between-groups ANOVA was performed (see Table 3). The main effect of the approach was not significant (F_2 = .059, p = 0.808). To test the designer performance difference between approaches, we used a paired-*t* test (pair-wise) for each subject. The paired-*t* test procedure compared the means of two variables for a single group, computed the differences between the values of the two variables for each subject, and tested whether the average differed from 0. The mean performance difference between the top-down and bottom-up for each subject was not statistically significant at .05 level (t_{102} =1.225 , p = .223) even though the gap was wider than the between-groups results. Hypothesis *H1* was therefore supported by the between-groups ANOVA and paired-*t* tests that there is no difference in performance between approaches.

Table 2: Error Rate Means of Each Approach Across Courses

	Top-down	Bottom-up	Overall by Course
SA&D	0.311	0.327	0.319
DM	**0.064**	0.112	0.088
BSA	0.082	**0.053**	**0.068**
Overall by Approach	**0.167**	0.192	

Table 3: ANOVA of the Two Factor Factorial Design

Source	Approach	Course	Approach × Course	Error
Type III Sum of Squares	0.005	2.78	0.035	17.077
Df	2	2	4	200
Mean Square	0.005	1.395	0.018	0.085
F	0.059	16.279	0.206	
Sig.	0.808	**0.000**	0.814	

Hypothesis *H2* stated that no difference in students' performance across different courses will exist. Subjects in SA&D, with little experience in data modeling, tended to make more errors than subjects in DM and BSA courses. To test the performance differences between different courses, we ran pair-wise comparisons between courses. The pair-wise comparisons showed that subjects' performance fell into two clusters. Subjects' performance in DM and BSA had no significant difference. Subjects in SA&D fell into another cluster that was significantly different from DM and BSA (see Table 4). Although the approach-course interaction plot (Figure 2) showed some sign of interactions between the two factors, the ANOVA results showed otherwise (F_4 = .206, p = .814).

Subjects in BSA had the lowest error rate across all three courses, while subjects in SA&D had the highest error rate. Our results therefore support the notion that previous database design experience had a significant effect on subjects' task performance. *H2*'s prediction that no performance difference will exist between different courses is therefore rejected (F_2 = 16.279, p = .000) (see Table 3).

Table 4: Pair-wise Comparisons between Courses

Course (I-J)	Mean Difference (I-J)	Std. Error	Sig.
SA&D—DM	0.248	0.055	**0.000**
SA&D—BSA	0.229	0.08	**0.005**
DM—BSA	-0.019	0.079	0.814

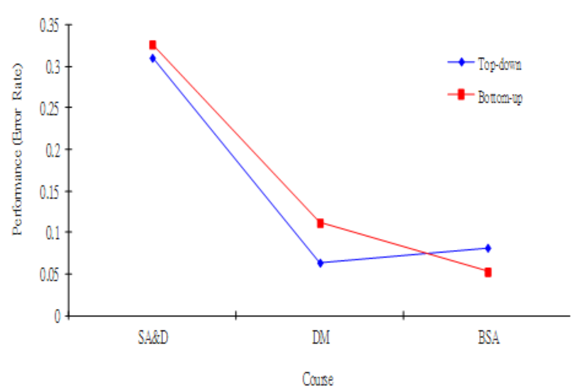


Figure 2: Approach—Course Interaction Plot

The means were plotted on a graph (Figure 2). Subjects in the BSA course produced lower error rates in the bottom-up approach. The overall average error rate of the top-down (16.7%) and bottom-up approaches (19.2%) indicated that the using bottom-up approach resulted in a slightly higher error rate than using the top-down approach.

Hypothesis *H3* stated that no difference in students' performance across different ER modeling constructs will exist. The three concepts tested were: entity, relationship, and cardinality. In the top-down approach, the performance of all three concepts had significantly different paired-*t* values (see Table 5). Entity was the easiest concept to grasp. The overall mean error rate of entity was 9 percent. The most difficult concept was cardinality. The overall mean error rate of cardinality was 34 percent. Relationship was in the middle with 15 percent error rate. Hypothesis *H3* was rejected because the error rates for all three ERD concepts were significantly different from one another.

Table 5: Paired *t*-test—Concepts Performance in Top-down Approach

	t	DF	P-Value
Entity vs. Relationship	-3.14	102	**0.002**
Entity vs. Cardinality	-4.981	102	**0.000**
Relationship vs. Cardinality	-4.04	102	**0.000**

Hypothesis *H4* posited that there will be no difference in students' performance across different relational data model constructs. Table 6 displays the paired t-test results. In the bottom-up approach, subjects had lower error rates in decomposing relations *Patient* and *Physician*. Subjects had higher error rates in decomposing Relations *Visit* and *Appointment*.

Table 6: Paired t-test—Concepts Performance in Bottom-up Approach

	t	DF	P-Value
Patient vs. *Physician*	-0.33	102	0.741
Patient vs. *Visit*	-3.79	102	**0.000**
Patient vs. *Appointment*	-2.86	102	**0.005**
Physician vs. *Visit*	-3.86	102	**0.000**
Physician vs. *Appointment*	-2.72	102	**0.008**
Visit vs. *Appointment*	-0.26	102	0.795

The relation *Appointment* had a composite key that made it an associative entity in the ER model which requires higher-level of understanding. Hypothesis *H4* was rejected since all four relation concepts in the bottom-up approach are significantly different from each other.

Overall Performance: Top-down vs. Bottom-up

A general overview of subjects' performance of the in-class exercise was shown in Table 2. The means of performance of the two factors were calculated. The lower means of error rates are shown in bold and underlined.

Table 7: Concept Performance of the In-Class Exercise

Concepts (Top-down)	Mean	Concept (Bottom-up)	Mean
ERD (Overall)	0.167	Normalization (Overall)	0.192
Entity	0.09	Relation *<u>Patient</u>*	0.151
Relationship	0.149	Primary Key	0.185
Cardinality	**<u>0.342</u>**	Non-key Attribute	0.146
		Relation *<u>Physician</u>*	0.154
		Primary Key	0.194
		Non-key Attribute	0.142
		Relation *<u>Visit</u>*	0.25
		Primary Key	0.233
		Non-key Attribute	0.259
		Relation *<u>Appointment</u>*	**<u>0.254</u>**
		Primary Key	0.204
		Non-key Attribute	0.291

The subjects' demonstration of different concepts in the two approaches is shown in Table 7. The most error-prone concept in each approach is shown in bold and underlined. Subjects had more errors in assigning correct cardinalities using the top-down approach. Cardinality was the most difficult concept to master for most subjects. The subjects created more errors in the relation *Appointment* in the bottom-up approach. The relation *Appointment* was an associative entity, had a composite key, and was the most difficult concept to master in the bottom-up approach. Relations *Patient* and *Physician* had the lowest error rates since most subjects could relate those to their real-world experiences.

In the top-down approach, the performance of all three concepts, entity, relationship, and cardinality, had significantly different paired-*t* values (see Table 5). Entity was the easiest concept to grasp. The most difficult concept was cardinality. The difficulty level of relationship was medium. This result highlights the needs that database educators should ensure that the concepts of cardinality and relationship concepts are well explained and understood by students.

In the bottom-up approach, subjects had lower error rates in decomposing relations *Patient* and *Physician*. Subjects had higher error rates in decomposing relations *Visit* and *Appointment*. The relation *Appointment* had a composite key that made it an associative entity in the ER model. The combination of associative entity and composite key made it the most difficult concept to master in the bottom-up approach because of its complexity. This emphasizes the importance that database educators should ensure that concepts of associative entity and the composite key are understood by students.

Phase 2: Qualitative

Following the quantitative laboratory experiment (Phase 1), an open-ended question was used in Phase 2 to collect subjects' perspectives. Subjects were to give their opinions on which approach and concept were more difficult to learn/master. Forty-seven subjects answered the question: five subjects considered both approaches were easy, 10 said both approaches were difficult, 11 thought ERD was difficult, and 21 indicated normalization was difficult. Combining qualitative and quantitative approaches to this study, we intended to triangulate findings to find contradictions and new perspectives. In general, the qualitative results supported the quantitative analyses.

6. CONCLUSIONS AND LIMITATIONS

This study has several limitations. The use of students as subjects from a single university is always an issue in terms of the ability to generalize findings. The second limitation was the time constraint to complete the experiment in six 75-minute sessions, which limited the training time for the two database design approaches.
The experiment looked at two factors: approach and course (previous experience). The results indicated that experience has higher impacts on students' performance than approach. We did not find statistically significant difference between approaches. No significant interaction effect between the approach and course was found. Overall, the subjects in BSA had the lowest error rates. Looking into each approach, we found the subjects in BSA performed best using the bottom-up approach, and DM was the best in top-down. For individual courses, the subjects in DM and SA&D had the same overall approach ranking pattern (better in top-down), opposite of the BSA results. This indicated that with proper training/experience subjects could do better in bottom-up design approach. The most error-prone concepts in each approach were cardinality in top-down, and associative (transaction) relation/table in bottom-up.

The need for training designers in data modeling becomes more important due to the growth of database usage in the business world. Effective teaching of data modeling is one of the important issues/challenges for IS/IT educators. Novice designers are likely to make errors, and design flaws can lead to significant costs in the maintenance phase. This study proposed to examine the relationship between top-down and bottom-up design approaches and the error-prone concepts in each design approach. The results indicated that top-down design led to lower error rates for most cases but the bottom-up design sometimes outperformed when designers were equipped with adequate experience. Not all concepts in every design approach have the same level of difficulty. This study results suggest that IS educators should allocate enough time to teach the concepts of cardinality, associative entity/table, and composite key for database.

7. REFERENCES

Bernstein, P. A. (1976). Synthesizing third normal form relations from functional dependencies. *ACM Transactions Database Systems,* 1, 4, 277-298.

Chen, P. P. (1976). The entity-relationship model — Toward a unified view of data. *ACM Transactions on Database Systems*, 1(1), 9-36.

Codd, E. F. (1970). A relational model of data for large relational databases. *Communications of the ACM*, 13 (June), 377-387.

Connolly, T., & Begg, C. (2000). Database Solutions. A step-by-step guide to building databases. Harlow, England: Addison-Wesley.

Fotache, M. (2006). Why normalization failed to become the ultimate guide for database designers? Available at SSRN: http://ssrn.com/abstract=905060 or http://dx.doi.org/10.2139/ssrn.905060

Hoffer, J. A., Ramesh, V., & Topi, H. (2010). Modern database management. 10th Ed., Prentice-Hall, Upper Saddle River, NJ.

Khatri, V., Ramesh, V., Vessey, I., Clay, P. & Park S. (2004). *Comprehension of conceptual schemas: exploring the role of application and IS domain knowledge*, Working paper.

Philip, G. C. (2007). Teaching database modeling and design: Areas of confusion and helpful hints, *Journal of Information Technology Education*, 6, 481-497.

Rob, P., & Coronel, C. (2004). Database systems: design, implementation, and management. Course Technology, Boston, MA.

Shoval, P., Danoch, R., & Balabam, M. (2004). Hierarchical entity-relationship diagrams: The model, method of creation and experimental evaluation, *Requirements Engineering*, 9(4), 217-228.

Teory, T., Wei, G., Bolton, D., & Koenig, J. (1989). ER model clustering as an aid for user communication and documentation in database design, *Communications of the ACM,* 32(8), 975–987.

Topi, H., Valacich, J., Wright, R., Kaiser, K., Nunamaker, J., Sipior, J., & De Vreeda, G. (2009). IS 2010: Curriculum guidelines for undergraduate degree programs in Information Systems, *Communications of AIS*, 2009, 359-428.

Appendix A: Data Modeling Exercise

The information on this page relates to designing a database that stores information for a medical clinic. You will need to develop a data model using the top-down design approach (Step 1) and bottom-up approach (Steps 2 and 3).

1) Draw a simple Entity-Relationship diagram (ERD) (without attributes) that reflects the following business rules that were provided by your client:

   A patient, over time, may make many visits to the clinic, and each visit relates to a single patient. Each visit, which is allocated a unique visit number, may involve many appointments, with each appointment related to a single visit. A physician may deal with many appointments, and each appointment is dealt with by a single physician.

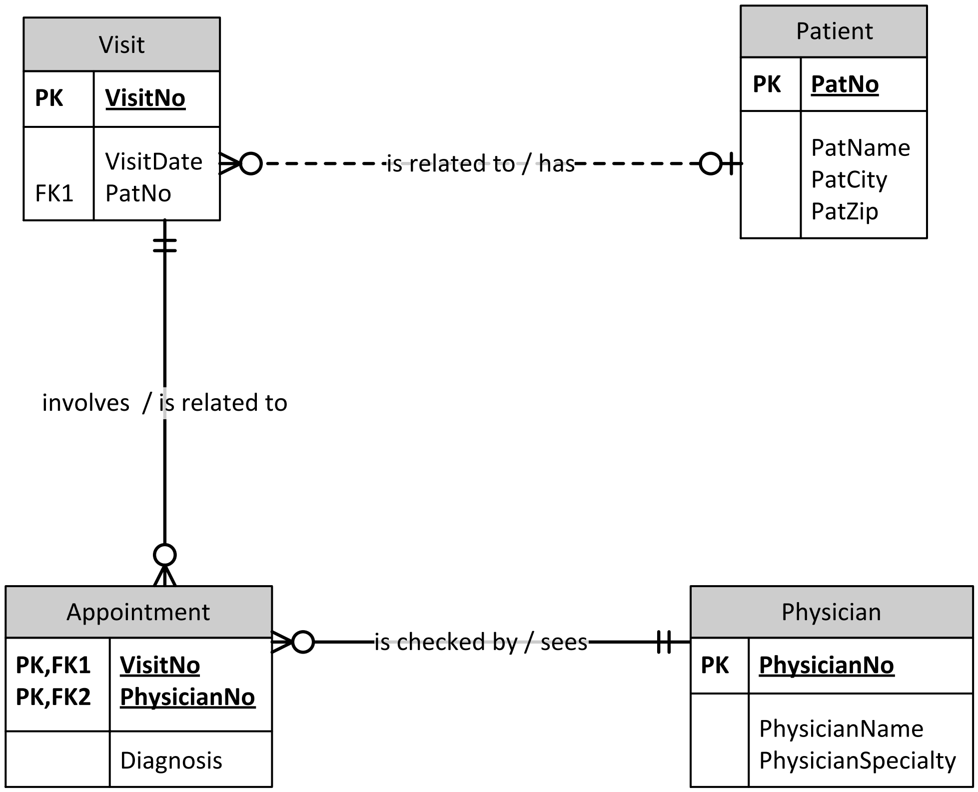
2) An experienced DBA inspected the sample data and identified the universal relation Clinic and functional dependencies (FDs). Your task is to normalize the universal relation Clinic to the third normal form (3NF). Show your answer in relation format.

   Universal Relation:
   Clinic (<u>VisiNo, PhysicianNo</u>, VisitDate, PatNo, PatName, PatCity, PatZip, PatPhone, PhysicanName, PhysicianSpecialty, Diagnosis)

   FDs:
   VisitNo, PhysicianNo → VisitDate, PatNo, PatName, PatCity, PatZip, PatPhone, PhysicianName, PhysicianSpecialty, Diagnosis
   PhysicianNo → PhysicianName, PhysicianSpecialty
   VisitNo → VisitDate, PatNo, PatName, PatCity, PatZip, PatPhone
   PatNo → PatName, PatCity, PatZip, PatPhone

3) Draw an ER diagram (with attributes) from Step 2.

Appendix B: Exercise Solutions



Patient (**<u>PatNo</u>**, PatName, PatCity, PatZip, PatPhone)
Physician (**<u>PhysicianNo</u>**, PhysicianName, PhysicianSpecialty)
Visit (**<u>VisitNo</u>**, VisitDate, PatNo)
Appointment (**<u>VisitNo</u>**, **<u>PhysicianNo</u>**, Diagnosis)