

# An Animated Introduction to Relational Databases for Many Majors

Suzanne W. Dietrich, Don Goelman, Connie M. Borrer and Sharon M. Crook

**Abstract**—Database technology affects many disciplines beyond computer science and business. This paper describes two animations developed with images and color that visually and dynamically introduce fundamental relational database concepts and querying to students of many majors. The goal is for educators in diverse academic disciplines to incorporate the animations in their existing courses, in order to meet their pedagogical needs. The introduction of the animations was assessed and evaluated within several contexts, including non-majors courses offered by computer scientists as well as a course on computational molecular biology. The evaluation indicated that student knowledge of fundamental database concepts improved significantly with the use of the courseware. The animations provide visual learners with an engaging method to learn the topics presented with the ability to replay the dynamic presentation. Overall, the evaluation indicates the accessibility of database topics across the disciplines as well as the specific concepts that need further elaboration.

**Index Terms**—Computer science education, courseware, querying, relational databases, visualization

## I. INTRODUCTION

Databases are ubiquitous; they affect many disciplines beyond computer science and business. Students of many majors can benefit from an introduction to fundamental database concepts that goes beyond information fluency [1]. For example, within the sciences, database methodology is an enabling technology for computational statistics [2]. However, many subfields are emerging outside of the sciences that include the use of databases, such as computational journalism [3]. Further, database educators have been challenged with offering a database course for non-majors [4].

There is not enough time in an introductory database course to include all the topics that database educators believe are fundamental for majors [5]. Consequently, it is even more

daunting to determine which fundamental database topics should be covered for non-majors. An introductory database course typically relies on computer science students having had prerequisite courses that cover programming, data structures, algorithms, and discrete mathematical structures. This prior knowledge cannot be assumed for an interdisciplinary audience. In addition, a course designed for non-majors may not provide a universal solution given that each university would face many factors in offering such a course. The approach that is reported in this paper is the development of courseware using Adobe® Flash® to create animations with images and color that visually and dynamically present the fundamental concepts of relational databases. Two animations were developed to cover the anticipated use of relational databases by many disciplines. The first animation advocates the use of databases for information storage and retrieval, highlighting databases' advantages over spreadsheets. The second animation illustrates principles related to retrieving data, including the industry standard query language SQL. The goal is for educators in diverse academic disciplines to incorporate the animations in their existing courses, in order to meet their pedagogical needs either as an in-class presentation by an instructor or as a self-study online assignment for the student. The animations and supporting resources are freely available on the Web [6].

This paper reports on the design of the animations and the evaluation of incorporating the animations as pedagogical tools in various contexts. Section II gives an overview of the use of visualization as an effective tool in computer science, particularly in the area of databases. Section III outlines the topics and goals of the introduction to relational databases animation. This courseware was introduced for the first time in several contexts in the 2010-2011 academic year, among them non-majors courses offered by computer scientists and courses on computational molecular biology. Since this was the first time that the animations were introduced, pre- and post-tests were designed to assess whether student knowledge of fundamental database concepts improved from the use of the courseware. This section also includes an evaluation of the animation in its second year of use. A second animation introducing students to querying relational databases, including the use of the SQL industry standard, is described and evaluated in Section IV. Section V provides a qualitative content analysis of the open-ended learner feedback gathered during the assessment process. Section VI concludes the paper

Manuscript received January 15, 2014. This material is based upon work supported by the National Science Foundation under Grant No. DUE-0941584 and DUE-0941401. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

S. W. Dietrich and C. M. Borrer are with the School of Mathematical and Natural Sciences, Arizona State University, Phoenix, AZ 85069 USA (email: dietrich@asu.edu and cborrer@asu.edu)

D. Goelman is with the Department of Computing Sciences, Villanova University, Villanova, PA 19087 USA (e-mail: don.goelman@villanova.edu).

S. M. Crook is with the School of Mathematical and Statistical Sciences and School of Life Sciences at Arizona State University, Tempe, AZ 85287 USA (e-mail: sharon.crook@asu.edu).

with a discussion of this initial study and plans for future work.

## II. VISUALIZATION AS AN EFFECTIVE PEDAGOGICAL TOOL

The current research was influenced by prior work on the topic by the authors and others. The incorporation of visual learning [7] and animation [8] has been encouraged for years, especially in the science, technology, engineering, and mathematics (STEM) disciplines, to enhance learning and to engage student interest. The book [9] supports the effectiveness of animations to assist student learning. In particular, it indicates that illustrated text has positive effects on learning and that the dynamic visualization of ideas can enhance cognitive meaning. Indeed, Robert Gagné, a renowned contributor to instructional design, insists that pedagogical media must be interactive [10]. In computer science, program animation was found to improve the learning of average students by “enabling them to build concrete mental models” [11]. There have also been efforts in visualizing various algorithms [12] and animations that introduce multiple computer science topics [13]; however, the topic of databases is not one of those presented.

A variety of database courseware is available [14] for majors. One example is the WinRDBI educational tool [15], [16] that evaluates both formal relational query languages (relational algebra and relational calculus) and SQL over a common database instance. There are also intelligent tutoring systems, such as DB-Suite [17] (that includes SQL-Tutor), NORMIT, and KERMIT; these cover SQL, normalization, and Entity Relationship diagrams, respectively. ADVICE [18] is an educational system supporting exercises with automated correctness checking that covers a majority of the topics in a database course for majors. The most closely related database courseware is the Animated DataBase Courseware (ADbC) project [19], [20]. However, the audience for ADbC consists of computer science or business majors enrolled in a database course. ADbC provides examples of in-depth database concepts using a Web-based graphical user interface implemented in Java (but not animation per se) to walk a student through a particular technical concept meant for majors only. The above database educational software targets majors. Typically these are more complex tools that educators use to enhance learning after having first presented the concepts in class. This approach is in strong contrast to the objective of the animation courseware presented in this paper, which is meant to be the methodology used to introduce students, especially those with no prior background, to the fundamental concepts of databases. It is important to note that in their database courses for majors the authors now also use the animations to introduce the concepts, and then incorporate some of the abovementioned tools to provide interactive learning of the more detailed concepts.

Additional research on animation and visualization indicate that its use can improve how quickly students learn or provide additional motivation to learn [21]. Also, research on visualization strongly encourages the use of an active learning activity to further engage learners [22], [23]. Consequently,

the incorporation of the animations in this research involved the use of cooperative learning activities to reinforce the learning objectives. As described in Section V, learner feedback supported both the benefits of the visual approach for motivation as well as the practice of the concepts through active learning.

## III. ANIMATION 1: INTRODUCTION TO DATABASES

### A. Overview

The choice of topics for the animation that introduces databases was strongly motivated by the research literature [4], [24] and by the student’s intuition of spreadsheets. Therefore, the underlying theme selected for the animation is based on explaining some potential shortcomings of a spreadsheet, and how relational databases can avoid those issues. In this context, the term “spreadsheet” refers to a non-normalized table, which may be vulnerable to redundancy and anomalies. The final design of the animation was the result of an iterative process, which was substantially influenced by experience in teaching the concepts and feedback from the visualization of the animation as it was developed.

Table I shows the order of topics included in the animation, along with a brief description. After showing the ubiquity of data, the animation relies on students’ intuitive understanding of data storage in a spreadsheet. It shows how a spreadsheet can easily answer some types of questions, but other types of ad hoc questions require a copy of the data to be made for further manipulation. When a spreadsheet stores repeated information, there are some problems, known as “anomalies,” that occur when data are inserted, updated, or deleted. The concept of a database is introduced as providing efficient shared access to persistent, related data, with a single copy of the data and no unnecessary repetition. The original spreadsheet is then broken down into smaller parts, representing separate concepts. (In order to provide a gentler introduction to a diverse audience, the relevant technical terms, *decomposition* and *normalization*, are not used in the animation.) Relations are introduced as these smaller spreadsheets that do not suffer from the anomalies of the larger one. The concepts of primary and foreign keys are presented as a way to identify data uniquely, and to relate the rows in different tables so as to answer queries; this is illustrated by an example.

### B. Assessment

The animation was introduced into various non-major settings in two universities: a technology-based course for non-majors (CST), a database course for non-majors (DB0), and a computational molecular biology course (BIO). Table II provides an overview of the topics covered in each course. Both the CST and DB0 had no prerequisites, whereas the prerequisite for the BIO junior-level course was one semester of calculus. In the CST course, the instructor presented the animation during class after covering Excel to explain the difference between spreadsheets and databases, as well as to introduce relational databases and querying before starting coverage of databases using Access as a tool. Another class

TABLE I  
TOPICS FOR INTRODUCTION TO DATABASES ANIMATION

Topic	Description
Data	An overview of how data are used in our everyday lives.
Spreadsheet	An introduction to how data are stored using a spreadsheet.
Questions	Examples of asking questions over data stored in a spreadsheet.
Anomalies	An introduction to some problems that may occur when using a spreadsheet to store data.
Database	What is a database?
Breakdown	Illustrates how data from a spreadsheet can be broken down into several parts to avoid anomalies.
Relations	What is a relation?
Keys	An introduction to primary and foreign keys, and how a database uses them to identify and relate information.
Queries	A brief introduction to asking questions over a database.

was devoted to discussing the topics introduced, using a cooperative, in-class exercise. In the DB0 course, the advantage of relational databases over alternate strategies was illustrated quite early. The instructor presented the animation in class and reinforced the coverage with the publisher's PowerPoint slides [24], as well as a cooperative in-class exercise. In the BIO course, the animation replaced a static PowerPoint presentation that introduced terminology and concepts related to database management systems and relational databases. A cooperative, in-class exercise reinforced these topics, then further class time was used to discuss the use of databases in archiving and sharing molecular biology data. In all courses, the animation was made available to students on a course management system, and they were strongly encouraged to review the animation at their own pace to support a graded assignment and exam questions.

Pre- and post-assessments were designed to assess the topics covered by the animation: spreadsheet anomalies, the absence of anomalies in a set of normalized tables, the identification of primary keys, the recognition of foreign keys, and the identification of data and relationships for answering queries. Table III summarizes the topics and concepts assessed. For spreadsheet anomalies in questions 1 through 5, students were asked to determine redundant data in the spreadsheet and whether an insert, update, or delete operation changes an existing row, changes multiple rows, results in a loss of information, or cannot be represented. Given a breakdown of the larger sheet into three smaller spreadsheets, questions 6 through 9 prompted students to identify which of the smaller spreadsheets would need to be revised to handle the same four operations from questions 2 through 5. Questions 10 through 12 asked students to identify primary keys for each spreadsheet by asking which columns or combination of columns would always identify a unique row in the spreadsheet, even if the data in the cells change. Two of the spreadsheets had a simple primary key consisting of a single attribute, and the third sheet had a composite primary key consisting of the combination of two attributes. Questions

TABLE II  
COURSE AND TOPIC OVERVIEW

Course	Topic Overview
BIO	Molecular biology databases, genetic sequence alignment, phylogenetic reconstruction, gene prediction
CST	Networks and how the internet works, XML (with DTD), Excel, Intro to DB and Querying, XHTML
DB0	Motivation for relational databases; terminology; referential integrity; SQL DDL and queries; MS-Access; QBE; the E-R model

13 through 15 covered the determination of foreign keys, if any, for each spreadsheet, by asking which columns or combination of columns refer to values that uniquely identify a row in another spreadsheet. The remainder of the questions, 16 through 20, asked students to indicate which spreadsheets were needed to answer various types of questions, including average, count, select, and join.

The pre- and post-tests consisted of the 20 questions outlined in Table III. However, the order of the questions within the topics and the order of the choices for each question were different with the exception of the first question, which asked students to identify the redundant data in the spreadsheet. This identification is important for answering questions 2 through 5 on anomalies. Both tests had enterprises that represented two entities and a many-to-many relationship between these entities, which is similar to the example presented in the animation itself. The mapping of question numbers to topic concepts shown in Table III provides a framework for the following evaluation of the gathered assessment data.

### C. Evaluation

The animation was introduced in two successive years into a diverse set of courses at two institutions. In both years, the scores on the pre-assessment and post-assessment instruments were on a scale of 0 to 20 and covered the same topics.

TABLE III  
ASSESSMENT QUESTIONS FOR INTRODUCTION TO DATABASES

Question	Concept	Topic
1	Identify redundant data	
2	No anomaly [Update: single row]	Spreadsheet
3	Insert	Anomalies
4	Update: multiple rows	(SA)
5	Delete	
6	No anomaly[Update: single row]	Database
7	Insert	Anomalies:
8	Update: multiple rows	None
9	Delete	(DA)
10	Simple	Identify
11	Simple	Primary Keys
12	Composite	(PK)
13	None	Identify
14	None	Foreign Keys
15	Composite	(FK)
16	Count	Identify
17	Select	Data for
18	Join	Query
19	Average	(QU)
20	Select	

However, the instrument was refined in the second year, from a multiple-answer paradigm to straightforward multiple-choice questions, considered effective at assessing student learning and comprehension [25]. The results are therefore presented by year.

The evaluation of the animations followed Human Subjects protocols, so the results shown in aggregate are for students agreeing to participate in the study - almost all of them. The fact that opportunities for gathering assessment data over the two-year period relied on course scheduling and sabbatical constraints at two universities had the effect of reducing subject numbers for the second year.

### Year 1

Table IV summarizes the pre- and post-assessment information gathered. The pre-assessment and post-assessment instruments were administered in the course offerings, and the post-test was a graded in-class exercise. The column labeled "n" provides the number of students participating in the study in each class. Since there were two sections of the databases for non-majors course (DB0) offered by the same instructor in each of two semesters, a decision was made to use the animation in just one offering, so as to gather some additional information in this initial assessment of the animation. The N suffix represents the section in which only traditional teaching methods were used, i.e., the control group, and the Y suffix represents the introduction of the animation, i.e., the treatment group. When registering for the course, students were unaware of the experiment and chose which DB0 section to attend based on the day and time that the course was offered. The mean pre-total and post-total scores are given, with the associated p-value for a paired analysis of the pre- and post-scores by student. The hypothesis of interest is whether the student post-scores are significantly higher than student pre-scores. The small p-values indicate that for each class the improvement of the mean post-score over the mean pre-score is statistically significant. Note that prior experience was not explicitly accounted for in the design of this study. However, conducting an analysis on the differences between pre- and post-scores by student (i.e., paired data), minimizes the effect of each student's individual ability on the results.

An examination of the control and treatment groups in DB0 is also warranted. In the Fall 2010 semester, Table IV shows that the treatment group outperformed the control group. However, in the Spring 2011 semester, the control group outperformed the treatment group. A two-sample t-test for independent samples was performed to compare the

TABLE IV

INTRODUCTION TO DATABASES YEAR 1 ASSESSMENT OVERVIEW

Course	n	Semester	Label	Pre (Mean)	Post (Mean)	p-value (paired t-test)
BIO	32	Spring	BIOS11	14.281	15.016	0.051
CST	11	Fall	CSTF10	12.318	13.455	0.014
DB0	14	Fall	DB0F10N	12.607	16.250	0.000
	17	Fall	DB0F10Y	13.735	16.912	0.000
	22	Spring	DB0S11N	12.841	17.182	0.000
	22	Spring	DB0S11Y	12.727	15.977	0.000
All	118	2010-2011		13.260	15.837	0.000

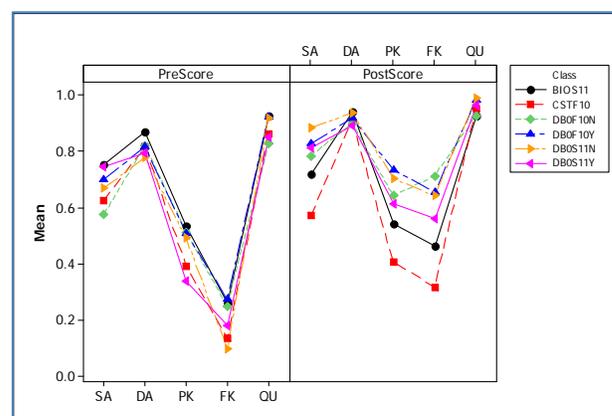


Fig. 1. Introduction to Databases Year 1 Mean of Pre- and Post-Scores by Topic and Class. Illustrates improvement in pre- and post-scores by topic as well as the accessibility of the topic to various majors.

performance of the control and treatment groups within each semester (e.g., Test 1: DB0F10N vs. DB0F10Y; Test 2: DB0S11N vs. DB0S11Y). Each test indicated no statistical significance in the difference of the means of these groups. Based on the results discussed in [21], this result is not uncommon. The introduction of animations itself does not necessarily improve learning but may provide additional motivation to learn. In addition, an examination of the composition of the students in these Spring semester courses shows that the control section included mathematics and engineering majors, as well as some students who later declared computer science as a major. The treatment section, on the other hand, included several students who withdrew from the course.

These initial evaluation results, illustrating the statistical significance of the rise from mean pre-score to mean post-score for each class, are encouraging, and further examination of the results prove interesting. The average scores for the pre- and post-assessment questions were broken down by animation topic, as shown in Fig. 1. Overall, the chart visually shows the improvement from the pre- to the post-tests. By and large, all of the charts strikingly illustrate the same ranking of student performance on the group of topics from highest to lowest: Querying, Database Anomalies (None), Spreadsheet Anomalies, Primary Key, and Foreign Key. This is a significant result of the assessment process, in that it provides valuable insight for database educators. Students of many majors can identify the tables required for answering a query (QU) and the tables that are modified for an insert, update, or delete operation on the database (DA). The identification of anomalies in a non-normalized spreadsheet was somewhat problematic. The concepts of primary and foreign keys appear to be the most difficult for students.

### Year 2

Table V summarizes the pre- and post-assessment results for the second year, which consisted of one offering of CST and two offerings of DB0, one section per semester. Table V provides the mean pre- and post-exam scores per class for the multiple-choice assessment. The p-values for a paired t-test are also given. Again, the hypothesis of interest is whether the student post-scores are significantly higher than student pre-

TABLE V  
INTRODUCTION TO DATABASES YEAR 2 ASSESSMENT OVERVIEW

Course	n	Semester	Label	Pre (Mean)	Post (Mean)	p-value (paired t-test)
CST	9	Fall	CSTF11	12.778	16.000	0.001
DB0	13	Fall	DB0F11	14.385	18.308	0.000
	10	Spring	DB0S12	15.583	18.333	0.000
All	32	2011-2012		14.371	17.600	0.000

scores. Here, too, the small p-values indicate that for each class the improvement of the mean post-score over the mean pre-score is statistically significant.

A similar breakdown of the mean pre- and post-assessment question scores by topic and class is shown in Fig. 2. For the most part, this chart has a similar shape to that of Year 1 (Fig. 1). Both figures show the same ranking in difficulty of the animation topics from highest to lowest: Querying (QU), Database Anomalies (DA), Spreadsheet Anomalies (SA), Primary Key (PK), and Foreign Key (FK).

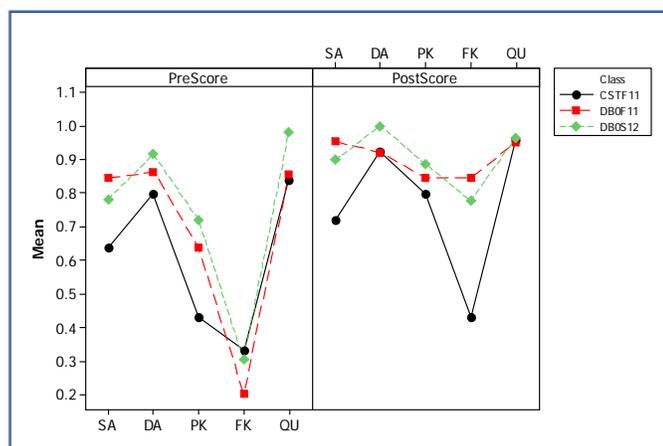


Fig. 2. Introduction to Databases Year 2 Mean of Pre- and Post-Scores by Topic and Class. Illustrates similarity to Year 1 results with same ranking in difficulty of the animation topics.

#### IV. ANIMATION 2: INTRODUCTION TO QUERYING

Databases provide for efficient storage and retrieval of data. The second animation emphasizes the retrieval through the use of queries. The ultimate goal is to help students understand the logic and concepts of retrievals and to apply them using the industry-standard query language SQL. Therefore, based on years of teaching experience and related literature [5], [24], topics, among them set operations and relational algebra, were chosen to explain the foundation of SQL. These topics are introduced in a visual and user-friendly approach so that students can understand how SQL works. This design was also the result of an iterative process that included feedback from the visualization of the animation as it developed.

##### A. Overview

Table VI shows the sequence of topics for the querying animation. Initially, the animation reviews the primary and foreign key relationships between the tables that are used to assist in the querying process. The coverage starts by considering the design of the query at a higher level, similar to the QBE panel provided by the MS-Access desktop database.

TABLE VI  
INTRODUCTION TO QUERYING ANIMATION TOPICS

Topic	Description
Query	Motivation to identify data and relationships.
Sets	Common set operators.
Filtering	Operations to horizontally and vertically filter data.
Joining	More ways of combining tables of data that require a form of filtering.
SQL	Introduction to querying using SQL.

The first step is to identify the tables needed to answer a query; the next step is to combine the tables correctly to get the answer. The possible operations on the data provide the sequence of topics leading up to the coverage of SQL. Specifically, the common set operations of union, intersection, and negation are visualized as a method for combining data in the tables. Additional operations on the data can include filtering the data to choose only some rows or columns. The terms horizontal and vertical filtering provide a gentle introduction to the select and project operators of relational algebra. The join operator is then introduced as a Cartesian product followed by a horizontal filtering of the result, which is usually based on the primary-foreign key relationships between the tables. With this foundation, the animation introduces the basic select-from-where statement of SQL with a step-by-step construction of an SQL textual query with the corresponding visual QBE design.

##### B. Assessment

Pre- and post-test multiple-choice assessment instruments were designed to aid in the evaluation of the effectiveness of the querying animation. The pre- and post-tests for querying used the same enterprises as the assessments for the introduction to databases. Instead of database tables, a pictorial representation of the database schema was given with the primary keys indicated by a yellow key. The foreign key relationships were explicitly stated. Table VII provides a summary of the assessment questions. The first three questions provide an English description of a query and ask the students to choose the set-based concept required to answer that question: union, intersection, or negation. Questions 4 through

TABLE VII  
ASSESSMENT QUESTIONS FOR INTRODUCTION TO QUERYING

Question	Concept	Topic
1	union	Set
2	intersection	Operations
3	negation	(SO)
4	horizontal	Filtering
5	vertical	Operations
6	both	(FO)
7	join with filtering	Join
8	filtering only	Operations
9	join only	(JO)
10	select	SFW
11	from	Concepts
12	where	(SC)
13	1 table	Identify
14	2 tables	SQL Queries
15	3 tables	(SQ)

TABLE VIII  
INTRODUCTION TO QUERYING ASSESSMENT OVERVIEW

Course	n	Semester	Label	Pre (Mean)	Post (Mean)	p-value (paired t-test)
CST	9	Fall	CSTF11	5.333	8.333	0.036
DB0	13	Fall	DB0F11	7.615	12.154	0.003
	10	Spring	DB0S12	6.400	13.400	0.000
All	32	2011-2012		6.594	11.469	0.000

6 emphasize filtering by determining which concept would be applied to achieve the desired fields in the result of the query: horizontal only, vertical only, both horizontal and vertical filtering, or neither. Similarly, questions 7 through 9 emphasize natural joins and ask which concepts are required to answer the query with the desired fields: natural join only, horizontal filtering only, vertical filtering only, or a combination of the above. Questions 10 through 12 ask students to relate the clauses of a generic SQL query to the concepts of horizontal filtering, vertical filtering, or Cartesian product. The last three queries elicit students to choose the correct SQL specification for the described query. The queries range from a simple query involving one table to a query with three tables.

### C. Evaluation

The querying animation was introduced in the second year in three courses: the technology-based course for non-majors (CST) and two offerings of the database course for non-majors (DB0) across two semesters. All offerings used the animation. The scores on the pre-assessment and post-assessment instruments were on a scale of 0 to 15, with each multiple-choice question being graded as 0 or 1 – incorrect or correct. Table VIII provides the mean pre- and post-test total scores by class, and includes the p-values for a paired t-test. The small p-values provide evidence that the improvement in the mean post-scores for the student over the mean pre-scores is statistically significant.

An analysis of the pre and post question scores by querying topic and class shown in Fig. 3 illustrates the improvement between the pre and post assessments. Unlike the introduction to databases animation in which the class graphs of the pre- and post-test question scores by topic had similar structure, the querying animation shows more variability between classes.

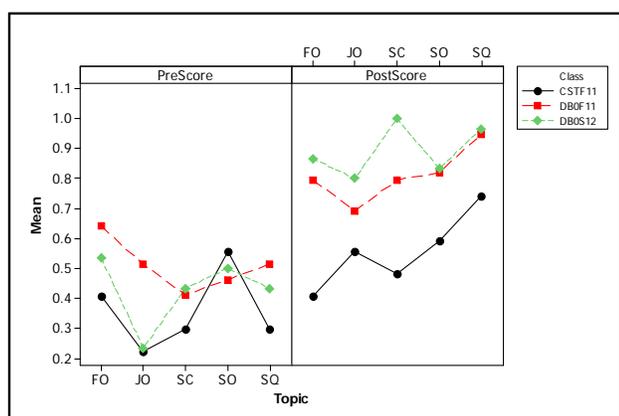


Fig. 3. Introduction to Querying Mean of Pre- and Post-test Scores by Topic and Class. Illustrates the improvement of pre- and post-test assessments by class and topic.

Querying builds on the fundamental introduction to databases in the first animation and involves more advanced topics. The most difficult topic appears to be the join operations, which is one of the harder concepts for students to understand in the classroom, too. The most improved topic is that of SQL queries (SQ), which illustrates that most students were able to identify the correct specification of queries in SQL after viewing the animation and participating in learning activities in support of the animation.

### V. LEARNER FEEDBACK

In addition to the assessment data gathered and evaluated for the pre- and post-tests, learner feedback was compiled as an important component of the evaluation [22]. In particular, the open-ended comments from the students were particularly encouraging. The following are a representative sample of the 77 comments volunteered by students:

- *Absolutely loved all the animations. I know I am a visual learner and having the animations was a huge help. Especially in this type of class where a lot of the information is complex to explain with just words and lecture. Really like the scroll feature and that the animation could be rewound or fast-forwarded by dragging along the bottom. Overall, excellent.*
- *The animation was a very helpful way to visualize the processes that normally occur “unseen.” The best example I can think of was its method of showing normalization, as it showed with each step certain fields being blanked out and implemented in a new table.*
- *Those little flash animations were definitely very helpful in learning some of the various concepts. It helped that the animations were dynamic, helping some concepts stick better than they would from a textbook or slideshow.*
- *The textbook and PowerPoint slides were not really helpful in my understanding. Going back to the animation was extremely helpful as it showed in a logical manner how queries work in both QBE and SQL.*
- *The animation was useful because it was available at all times and it could also be repeated multiple times if the repetition was necessary for comprehension.*

Some students reinforced that although the animations are helpful, practice is essential to learning:

- *I believe that ultimately the hands-on labs on queries and SQL queries was the most successful at helping us understand queries. However, the animation also was helpful in that it illustrated what needed to be done and what was meant by a certain query.*

The importance of practice coincides with the cooperative learning approach of this research as well as the recommendations of [22].

A qualitative content analysis [26], [27] was performed on the open-ended comments gathered in the assessment timeframe. A data-driven, inductive approach was used to determine the categories and subcategories in the coding frame shown in Table IX. The main categories are Helpful,

TABLE IX  
CODING FRAME FOR QUALITATIVE CONTENT ANALYSIS  
OF LEARNER FEEDBACK

Categories	Code	Subcategories
Helpful	H1	Helpful/useful/beneficial (no specific reason)
	H2	Visual learner
	H3	Visualization/dynamic presentation
	H4	Repeatable for learning and studying
	H5	Interactive replay capability
	H6	Better than slides
	H7	Better than textbook
	H8	Helpful in understanding: Redundancies
	H9	Helpful in understanding: Anomalies
	H10	Helpful in understanding: Breakdown
	H11	Helpful in understanding: Relations
	H12	Helpful in understanding: Keys
	H13	Helpful in understanding: Queries
	H14	Helpful in understanding: SQL
	H15	Liked the change of pace from lecture/slides
Improvements Needed	I1	Not enough depth of topic coverage
	I2	Improvement in visual presentation (color, clutter)
	I3	Difficult to understand: Redundancy
	I4	Difficult to understand: Anomalies
	I5	Difficult to understand: Foreign keys
Not as Helpful	N1	Practice better
	N2	Text better
	N3	Slides better
	N4	Instructor better
	N5	Not helpful (no specific reason)

Improvements Needed, and Not as Helpful. Each category has multiple subcategories corresponding to specific reasons for belonging to the category. Note that a code has been assigned to subcategories to assist in charting the results. One open-ended comment may map to several subcategories. For example, coding the first comment results in a count for visual learner (H2) and another for interactive replay capability (H5). The code H1 corresponds to a miscellaneous helpful comment where a specific reason was not given. Similarly, the code N5 corresponds to a miscellaneous “not as helpful” comment. Fig. 4 indicates the counts in each of the subcategories. The top four subcategories illustrate that students find the animations helpful, particularly due to visualization, and appreciate the ability to repeat the animation for learning and studying. In addition, 70 of the 77 comments (91%) included a mapping to the Helpful category, whereas three of 77 (4%) comments mapped only to the Improvements Needed category and four out of 77 (5%) mapped only to the Not as Helpful category.

## VI. CONCLUSION, DISCUSSION, AND FUTURE WORK

The design and evaluation of the animations presented in this paper represent a significant first step in introducing fundamental database concepts to students of many majors. The animations provide visual learners with an engaging method to learn the topics presented with the ability to replay the dynamic presentation. The evaluation indicated that

student knowledge of fundamental database concepts improved significantly with the use of the courseware. In particular, the evaluation of the Introduction to Databases animation indicates that students of many majors can identify the data needed to answer various types of queries. Also, the evaluation of the Introduction to Querying animation illustrates that students of diverse backgrounds are able to recognize the correct specification of a basic SQL query. These results provide evidence to support future work in making databases accessible to a wider audience. In addition, the learner feedback confirmed that the animations assisted the learning process by visualizing complex concepts. The animations have also been presented in various venues in regional and international conferences [28]-[31], and have been well received.

Another potential use of these animations may be for the Computer Science Principles project [32], which is a collaborative effort of computer science educators to design a new introductory-level college course on computing. Seven “big ideas” have been formulated in support of the learning objectives for this course; “Big Idea III” states that “data and information facilitate the creation of knowledge.” The expectation is that these animations can provide a resource for coverage of data within the context of CS Principles.

An additional objective of this on-going project is to provide the capability for educators to customize the animation with an example of their own choice. The animations described in this paper have a fixed scenario based on students, courses, and students taking courses. The customized example must also represent two entities and a many-to-many relationship between these entities. The customization uses XML to specify the text and data for the example scenario with a companion XML Schema Definition (XSD) that verifies the format of the XML, which is loaded into the animation using the XML capabilities provided by the object-oriented programming language (ActionScript 3) supported within the Flash development environment.

To facilitate the customization and long-term flexibility of the project, the animations were refined and implemented using Flash Builder, which is also known as FLEX. FLEX is a framework for building Flash animations that uses XML for defining the visual components and ActionScript to implement the underlying logic. This redesign in FLEX allowed the incorporation of formative feedback received from workshops to refine the animations. In particular, discrete steps were introduced within the topic that can be selected to restart the animation at that point. This concept of segmentation supports positive learning effects by providing meaning to the structure of the animation [33]. The original versions (Version 1.0) of the animations assessed in this paper are available on the Web site [6]. The revised versions (Version 2.0) are also available, including customizations for computational molecular biology (CMB) and geographic information systems (GIS). There are also resources for each customization that include an Access database instance as well as the SQL standard statements for creating and populating the database, since the animations are also useful for introducing these concepts to majors.

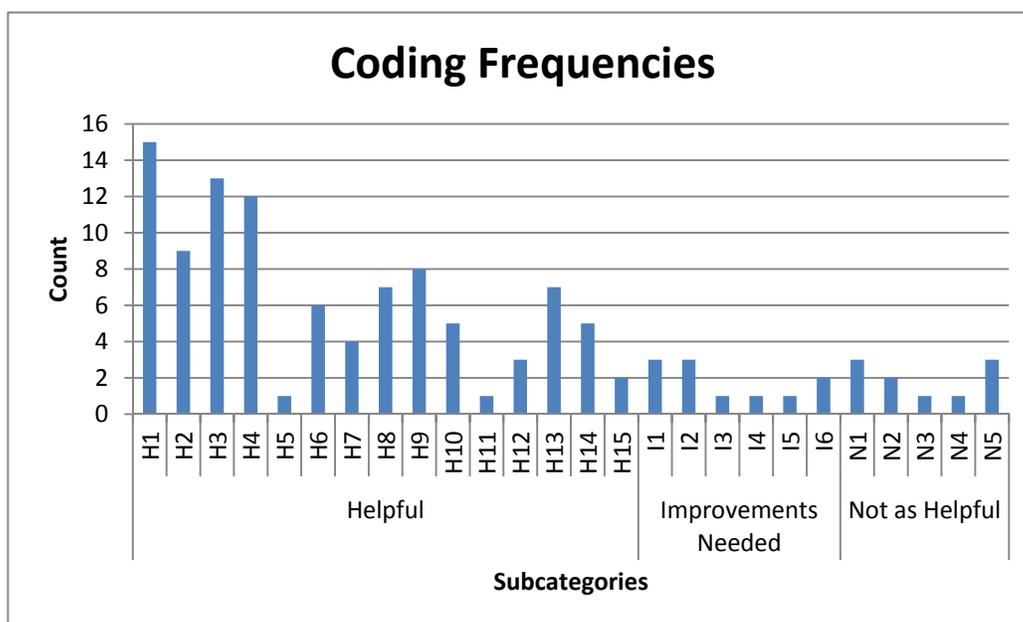


Fig. 4. Counts for Subcategories in the Coding Frame. The top responses are in the Helpful category, indicating that students find the animations helpful, particularly due to visualization, and appreciate the ability to repeat the animation for learning and studying.

There are several directions for future work. One objective is the design of a customizable, self-assessment framework to incorporate formative feedback into the animations. The planned framework will give students the ability to respond to questions as part of the animation to increase their level of comprehension. Responding involves answering questions about the learning outcomes of the visualization, and studies indicate that responding significantly improves learning over just viewing [22]. The proposed self-assessment for the animations should focus on improving the student's comprehension of the learning outcomes of the animations, while encouraging the student to continue to use the animation to reinforce learning. This ideal is known as a mastery goal in cognitive theory [34] versus a performance goal that focuses on a grade. Another objective is the development of tools to assist in the customization of the animation and the self-assessment framework. Though the customization process is feasible for computer scientists, visual and user-friendly tools are necessary and crucial to the adoption and broader impact of the animations for other disciplines.

#### ACKNOWLEDGMENT

Thanks are due to many who contributed to this effort, including Bangbo Hu as a domain partner for geographic information systems; Carol Weiss for assistance with assessment; Patricia Clark for artistic assessment, Brad Schneider for his contributions on the design and development of the FLASH animations, and Priya Dharshini Ponniah Rajagopal for assistance in the development of the FLEX animations. We would also like to thank Monisha Pulimood for customizing the animation for computational journalism and using the customized animation in her classes.

#### REFERENCES

- [1] L. Snyder, *Fluency with Information Technology: Skills, Concepts, & Capabilities*, 4th ed. Boston, MA: Addison-Wesley, 2010.
- [2] D. Nolan and D. T. Lang, "Computing in the statistics curricula," *American Statistician*, vol. 64, no. 2, pp. 97-107, 2010.
- [3] S. M. Pulimood, D. Shaw, and E. Lounsbury, "Gumshoe: a model for undergraduate computational journalism education," in *SIGCSE*, 2011, pp. 529-534.
- [4] D. Goelman, "Databases, non-majors and collaborative learning: a ternary relationship," in *ITiCSE*, 2008, pp. 27-31.
- [5] E. S. Adams, M. Granger, Don Goelman, and C. Ricardo, 2004. "Managing the introductory database course: what goes in and what comes out?" in *SIGCSE*, 2004, pp. 497-498.
- [6] S. W. Dietrich and D. Goelman (2014, May 19). *Databases for Many Majors* [Online]. Available: <http://databasesmanymajors.faculty.asu.edu/>
- [7] M. B. McGrath and J. B. Brown, "Visual learning for science and engineering," *IEEE Computer Graphics and Applications*, vol. 25, no. 5, pp. 56-63, Sep. 2005.
- [8] R. Lowe, "Interrogation of a dynamic visualization during learning," *Learning and Instruction*, vol. 14, no. 3, pp. 257-274, Jun. 2004.
- [9] C. Wetzel, P. Ratdtke, and H. Stern, *Instructional effectiveness of video media*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1994.
- [10] R. Gagné, L. Briggs, and W. Wager, *Principles of Instructional Design* 3<sup>rd</sup> ed. Orlando, FL: Holt, Rinehart and Wilson, 1988.
- [11] M. Ben-Ari, R. Bednarik, R. Ben-Bassat Levy, G. Ebel, A. Moreno, N. Myller, and E. Sutinen, "A decade of research and development on program animation: The Jeliot experience," *Journal Visual Languages and Computing*, vol. 22, pp. 375-384, 2011.
- [12] C. A. Shaffer, M. L. Cooper, A. J. D. Alon, M. Akbar, M. Stewart, S. Ponce, and S. H. Edwards, "Algorithm Visualization: The State of the Field," *ACM ToCE*, vol. 10, no. 3, pp. 1-22, Aug. 2010.
- [13] O. Balci, W. Gilley, R. J. Adams, E. Tunar, and N. D. Barnette, "Animations to assist learning some key computer science topics," *JERIC*, vol. 1, no. 2es, Aug. 2001.
- [14] ACM SIGMOD (2014, May 19). *Free/public domain software* [Online]. Available: <http://www.sigmod.org/resources/free-public-domain-software>
- [15] S. W. Dietrich (2014, May 19). *WinRDBI* [Online]. Available: <http://winrdbi.asu.edu/>

- [16] S. W. Dietrich, E. Eckert, and K. Piscator, "WinRDBI: a Windows-based relational database educational tool," in *SIGCSE*, 1997, pp. 126-130.
- [17] A. Mitrovic, P. Suraweera, B. Martin, and A. Weerasinghe, "DB-suite: experiences with three intelligent, web-based database tutors," *Journal Interactive Learning Research*, vol. 15, no. 4, pp. 409-432, 2004.
- [18] M. Cvetanovic, Z. Radivojevic, V. Blagojevic, and M. Bojovic, "ADVISE - Educational System for Teaching Database Courses," *IEEE Trans. on Education*, vol. 54, no. 3, pp. 398-409, Aug. 2011.
- [19] M. C. Murray (2014, May 19). *Animated DataBase Courseware: interactive approach for teaching the principles of database concepts* [Online]. Available: <http://adbc.kennesaw.edu/>
- [20] M. C. Murray and M. Guimaraes, "Recent upgrades to an animated database courseware (ADbc)," in *ACM SE*, 2008, pp. 515-516.
- [21] M. Byrne, R. Catrambone, and J. Stasko, "Evaluating animations as student aids in learning computer algorithms," *Computers & Education*, vol. 33, no. 4, pp. 253-278, Dec. 1999.
- [22] T. L. Naps, G. Rößling, V. Almstrum, W. Dann, R. Fleischer, C. Hundhausen, A. Korhonen, L. Malmi, L., M. McNally, S. Rodger, and J. Ángel Velázquez-Iturbide, "Exploring the role of visualization and engagement in computer science education," *SIGCSE Bull.* vol. 35, no. 2, pp. 131-152, Jun. 2002.
- [23] J. Urquiza-Fuentes and J. A. Velazques-Iturbide, "A survey of successful evaluations of program visualization and algorithm animation systems," *ACM TOCE*, vol. 9, no. 2, Jun. 2009.
- [24] D. Kroenke and D. Auer, *Database Concepts* 5th ed. Upper Saddle River, NJ: Prentice Hall, 2011.
- [25] K. Woodford and P. Bancroft, "Multiple choice questions not considered harmful," in *Proc. ACE 2005*, pp. 109-116.
- [26] S. Stemler, "An overview of content analysis," *Practical Assessment, Research & Evaluation*, vol. 7, no. 17, 2001.
- [27] M. Schreier, *Qualitative content analysis in practice*, London, UK: SAGE Publications, 2012.
- [28] S. W. Dietrich and D. Goelman, "Reaching out to many majors: a database approach," Poster, *SIGCSE* 2011.
- [29] D. Goelman and S. W. Dietrich, "Database animations for many majors: pre-conference workshop," *J. Comput. Sci. Coll.*, vol. 27, no. 3, pp. 7-7, Jan. 2012.
- [30] S. W. Dietrich and D. Goelman, "Database animations for many majors: conference tutorial," *J. Comput. Sci. Coll.* vol. 27, no. 4, pp. 174-174, Apr. 2012.
- [31] S. W. Dietrich and D. Goelman, "Databases for Many Majors: A Student-Centered Approach," NSF Showcase, *SIGCSE* 2012.
- [32] O. Astrachan and A. Briggs, "The CS Principles Project," in *ACM Inroads*, vol. 3, no. 2, pp. 38-42, Jun. 2012.
- [33] I. A. E. Spanjers, T. Van Gog, P. Wouters, and J. J. G. Van Merriënboer, "Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing," *Computers & Education*, vol. 59, pp. 274-280, 2012.
- [34] J. H. McMillan and J. Hearn, "Student self-assessment: the key to stronger student motivation and higher achievement," *Educational Horizons*, vol. 87, no. 1, pp. 40-49, 2008.

**Suzanne W. Dietrich** received the Ph.D. degree in computer science from Stony Brook University, Stony Brook, NY, in 1987.

She is a Professor in the School of Mathematical and Natural Sciences, Arizona State University, Phoenix, AZ. She is the author of *Understanding Relational Database Query Languages* (Upper Saddle River, NJ: Prentice Hall, 2001) and coauthor of *Foundations of Object Databases: Object-Oriented and Object-Relational Design* (San Rafael, CA: Morgan & Claypool, 2011). Her research interests include computer science education with an emphasis on databases as well as the incremental monitoring of conditions and evaluation of rules in active environments that respond to events and streaming information.

Dr. Dietrich is a member of the Association for Computing Machinery (ACM) and the ACM special interest groups on the management of data (SIGMOD) and computer science

education (SIGCSE). Dr. Dietrich is recognized by the ACM as a Distinguished Educator for her impact and significant accomplishments in the field of computer science education.

**Don Goelman** received the Ph.D. degree in mathematics from the University of Pennsylvania, Philadelphia, PA, in 1971.

He is an Associate Professor in the Department of Computing Sciences, Villanova University, Villanova, PA. His research areas are computer science education, particularly databases for non-majors, and database theory. In fall 2013 he was invited to be a visiting associate professor at the University of British Columbia, primarily in the computer science directorate of the Carl Wieman Science Education Initiative.

Dr. Goelman is a member of ACM and the ACM special interest groups in computer science education (SIGCSE) and management of data (SIGMOD). He served as program chair of SIGCSE's Technical Symposium (SIGCSE 1994), program co-chair of its conference on Innovations and Technology in Computer Science Education (ITiCSE 2002), birds-of-a-feather chair at SIGCSE 2005 and proceedings co-chair at ITiCSE 2012.

**Connie M. Borrer** received her Ph.D. degree in industrial engineering from Arizona State University, Tempe, AZ, in 1998.

She is a Professor in the School of Mathematical and Natural Sciences, Arizona State University, Phoenix, AZ. She is the co-author of two books: *Probability and Statistics in Engineering* (Hoboken, NJ: John Wiley & Sons, Inc, 2003), and *Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models*. (Philadelphia, PA.: ASA-SIAM Series, 2005). Her research interests include experimental design, response surface methodology, applied statistics, regression analysis, statistical process control, and measurement systems analysis.

Dr. Borrer is a Fellow of the American Statistical Association (ASA) and the American Society for Quality (ASQ) as well as a Senior Member of the Institute of Industrial Engineers (IIE) and the American Society for Engineering Education (ASEE).

**Sharon M. Crook** received a Ph.D. in applied mathematics from the University of Maryland College Park in 1996.

She is an Associate Professor in the School of Mathematical and Statistical Sciences and the School of Life Sciences at Arizona State University in Tempe, Arizona. Her research interests include the use of mathematical models to understand the dynamics of neurons and neural networks and topics in neuroinformatics related to model sharing and model reproducibility.

Dr. Crook is a member of the Board of Directors for the Organization for Computational Neuroscience, and a member of the Society for Neuroscience, the Society for Industrial and Applied Mathematics, the Society for Mathematical Biology, and the Association for Women in Mathematics. Dr. Crook is a 2011 recipient of a Scottish Informatics and Computer Science Alliance Distinguished Visiting Fellowship.