Teaching Data Quality in the Undergraduate Database Course

Richard G. Mathieu
University of North Carolina at Wilmington
mathieur@uncwil.edu

Omar Khalil
University of Massachusetts at Dartmouth
okhalil@umassd.edu

Abstract

Current evidence indicates that poor data quality is pervasive and has a significant negative impact on business success. Information system professionals are typically charged with managing the enterprise's data resources, yet few have received training in techniques for improving data quality. As a result there is a clear need to integrate topics related to data quality in a database systems course. This paper describes the importance of data quality and defines data quality in a four dimensional framework. The content of five popular database system textbooks are analyzed for their coverage of data quality concepts. This analysis is used to suggest specific topics that can be covered in an undergraduate course in database systems.

I. Introduction

Poor quality data is pervasive and costly to industry. Redman reports that error rates of 1-5% are typical with an estimated immediate cost of about 10% of revenue (Redman, 1996). Customers, suppliers, distributors and employees are negatively impacted through poor service, billing and inconvenience. Data quality problems are exacerbated by large organizational databases where data is collected from multiple data sources. Strong, Lee and Wang (1996) caution that information system professionals should seek not only to improve data accuracy, but should also consider data accessibility and data relevance as it relates to the context of the data consumers' tasks.

Businesses have implemented programs to improve data quality as a source of competitive advantage (Redman, 1995). AT&T used its data quality program to suggest opportunities to re-engineer their billing system, and as a result billing errors were reduced by

two orders of magnitude (Redman, 1996). Data warehouses are used by organizations to improve customer service and managerial decision making. A major issue in building and maintaining a data warehouse is data quality. Typically organizations will initially spend a lot of time ensuring the quality of the data, but over time the focus on data quality fades. Without proper data quality processes, the data warehouse will begin to accumulate "dirty data" (Garcia, 1997).

While the importance of data quality has been increasing, most information systems curricula do not directly address this issue. Textbooks in database systems, system analysis and design, management information systems and decision support systems typically pay little attention to data quality. While prior model curricula gave little attention to quality issues, the IS'97 Model Curriculum has placed a significant emphasis on quality. However, the primary focus in the IS'97 Model Curriculum is on the "principles of quality improvement" and on "software quality" (Davis, Gorgone, Couger, Feinstein, Longenecker, Jr.,1997). No specific mention of data or information quality is made in the IS'97 Model Curriculum. Data quality is addressed indirectly through such topics as data integrity, EDP Auditing, data dictionaries and software development procedures. The result is that most university information system curricula give little attention to data quality as an important concept. At best, the IS student is exposed to a variety of topics that impact the quality of data, but is not equipped with a broad understanding of the principles behind measuring, tracking and improving data quality in an organization.

The purpose of this paper is to assist IS educators in teaching data quality principles in the undergraduate database course. First, the literature on data quality is surveyed, and from this survey a data quality framework is presented. This framework is used to analyze the coverage given to data quality in a sample of database textbooks. Finally, the implications drawn from the database textbook analysis are used to suggest specific topics that should be taught in the undergraduate database course. These suggestions are made in the form of four teaching modules: (1) Importance of Data Quality and the Role of Business Processes in Data Quality, (2) Role of Data Quality in the Life Cycle of Data, (3) Measure, Assess, Track and Improve Data Quality, and (4) Tips for Improving Data Quality.

II. A Review of the Literature on Data Quality

A. Data Quality: Data Consumer Perspective

Many data quality programs have focused exclusively on improving data accuracy. Today, many researchers point out that data quality is a multi-dimensional issue that needs to be assessed from the data consumer's perspective. Wang and Strong (1996) identified four dimensions of data quality consisting of fifteen measurable attributes:

(1) Intrinsic Data Quality
- Accuracy
- Believability
- Objectivity
- Reputation

(2) Contextual Data Quality
- Value-Added
- Relevancy
- Timeliness
- Completeness
- Appropriate Amount of Data

(3) Representational Data Quality
- Interpretability
- Ease of Understanding
- Representational Consistency
- Concise Representation

(4) Accessibility Data Quality
- Accessibility
- Accessibility Security

These results strongly suggest that information system professionals should strive to deliver more than accurate and objective data to the users. Conventional control techniques such as edit checks, database integrity constraints and EDP auditing have traditionally been used to improve data accuracy. However, IS professionals must give attention to techniques which help data consumers perform their tasks by developing and maintaining systems that are flexible enough to allow easy aggregation and manipulation of data, that are responsive enough to provide relevant and easily interpreted information, and that are secure and robust enough to prohibit accidental or criminal data corruption.

B. Data Quality: The Data Life Cycle

Data should be viewed as a business asset. Organizations hold mangers accountable for business assets like capital, raw materials, machinery and equipment, and employees. According to Redman (1996), organizations that wish to possess high quality data must recognize data as a business asset and define explicit management responsibilities for them. Unlike many other organizational assets, data is highly dynamic. Understanding the data life cycle is very important to understanding the nature of data.

Redman (1996) proposes a data life cycle that includes two cycles: the data acquisition cycle and the data usage cycle. The data acquisition cycle has the ultimate goal of storing data. Thus data modeling and obtaining data values are necessary components of acquiring data. The four steps of the data acquisition cycle are:

(1) Define the View (Logical Data Modeling)
(2) Implement the View (Physical Data Design and Implementation)
(3) Obtain Values (Populating the database)
(4) Update Records (Data Storage and Maintenance)

It is important to note that while data storage and database modeling & design are typically viewed as the province of the Information Technology group, the process of populating the database typically rests with the manager(s) responsible for the business processes that use the data.

The data usage cycle starts with data that has previously been stored. While many business uses of data are routine, like customer billing and paying employees, data is also used for making decisions. Using data in the decision making process typically requires combining large amounts of data, putting them in a new context, and interpreting the output. The four steps of the data use cycle are:

(1) Define the Subview (Query Design)
(2) Retrieve (Query Processing)
(3) Manipulation (Sorting, Aggregating, Reformatting, Analysis)
(4) Present Results (Form, rather then content, of the Results)

While query design and processing is viewed as the province of IT professionals, the manipulation of the content and the presentation of the results is often considered to be in the purview of the user (data consumer).

C. *Data Quality: Root Causes of Poor Data Quality*

According to the U.S. Defense Information System Agency (Cykana, Paul, and Stern, 1996), the root causes of poor data quality can be attributed to four primary areas:

(1) process problems,
(2) system problems,
(3) policy and procedure problems, and
(4) data design problems.

An understanding of the processes that generate, use and store data are essential toward an understanding of data quality. Business processes typically extend horizontally across the organization and process owners should be made responsible for the quality of data that they use or produce. Redman (1996) recommends the use of an information model called functions of information processing (FIP) to help process owners describe information chains. The FIP diagram models how data is created, moved, stored, filtered, queued, and associated in an information chain and is quite useful in identifying the source of data quality problems. Experience in the United States Department of Defense revealed that a majority of data errors can be attributed to process problems (Cykana, Paul, and Stern, 1996). As a result IS analysts are encouraged to examine the existing processes that support data entry, the assignment and execution of data quality responsibilities, and methods used to exchange data.

Data quality problems often stem from system deficiencies that stem from poorly documented modifications, incomplete user training and user manuals, or systems that have been extended beyond their original intent. IS professionals should examine system modifications, user training, user manuals, engineering change requests and problem reports in an effort to improve data quality.

Poor data quality often stems from inadequate organizational policies. Clear management responsibilities should be established within the organization. A data policy should cover: (1)

security, privacy and rules of use, (2) inventory of data assets, (3) data sharing and availability, (4) data architecture, (5) planning, and (6) the role of quality (Redman, 1996).

D. Data Quality: Steps for Improving Data Quality

Improving the quality of data in an organization is often a daunting task. Organizations often have enormous quantities of data spread over many divisions that often employ different technologies. A data quality program is essential for improving data quality within an organization. Redman (1996) states that a good data quality program should:

1) have clear business goals and objectives,
2) properly assign responsibilities for data and ensure that those responsible have the tools needed to succeed,
3) have an operational plan for improvement that specifies which improvement methods are to be applied to which data, and
4) establish a program administration.

A Total Quality Management (TQM) framework for improving data quality has been proposed by Dvir and Evans (1996). Inherent to this process is the need to translate data customer needs into metrics, a team-oriented approach to continuous quality improvement, and benchmarking performance. Many of the classical techniques of statistical quality control, such as pareto charts and control charts, can be applied to the measurement, tracking and improvement of data quality.

III. The Database Systems Course and Data Quality

The traditional undergraduate course in database systems focuses on database modeling, query languages, database design, and database development. These topics dominate the database textbooks currently available. Rollier (1993) stressed the importance of a database project in the undergraduate IS curriculum and emphasized the importance of the implementation experience in understanding concepts such as referential integrity, commit and rollback, and domains. Hossein (1992) describes a project-oriented, database life cycle approach to teaching the undergraduate database course. However, data quality is seldom explicitly mentioned in a database course or textbook. Topics that relate indirectly to data quality, such as data integrity, security, and concurrency control, are widely covered. However, the student does not generally

receive instruction on the overall importance of data quality in the design and implementation of databases.

The undergraduate course on database systems would appear to be an ideal course to introduce the IS student to the principles of data quality. Typically the student in this class will have had previous exposure to PC-based database and spreadsheet software. In addition, students have had exposure to different data structures in a programming class. Students in a MIS program will typically have had an introductory accounting class where students are exposed to the information needs of accounting departments. Given this background, the student is ideally situated to learn about the importance of data quality.

The recent IS 97 Model Curriculum and Guidelines for Undergraduate Degree Programs in Information Systems has put a great deal of emphasis on quality. In particular, the Model Curriculum emphasizes quality in two areas:

(1) principles of quality management, and

(2) software quality.

Noticeably missing from the IS 97 Model Curriculum (Davis, Gorgone, Couger, Feinstein, Longenecker, Jr.,1997) is any explicit mention of data (or information) quality. However, as part of the database course (IS'97.8) in the Model Curriculum there is a call to have the student have the experience of developing a database application where the developer must "ensure that data collection, verification, and control is established" and must "ensure that external audits will establish consistent goals and accomplishments".

Certainly the increased emphasis on the principles of quality management and software quality will greatly aid the IS professional wishing to deal with data quality. Indeed the principles of Total Quality Management have been suggested as a tool to improve classroom teaching (Thomson, 1994). However, it is felt that the omission of data quality as a core IS curriculum topic will leave the student with an incomplete perception of the role of data quality in the information system. This comes at a time when information system professionals are increasingly being held responsible for data quality within the organization.

IV. Conceptual Framework for Analysis

The primary purpose of this paper is to analyze several of the popular database system textbooks to determine their coverage of topics related to data quality. This analysis will then be used to suggest specific topics that can be integrated into a course in database systems that will introduce the IS student to data quality.

There are many different approaches that can be applied to the study of data quality. Wang, Storey and Firth (1995) created a framework for analyzing data quality research based on an analogy between product manufacturing and data manufacturing. Other approaches to data quality relate to the data life cycle, value chain analysis, EDP auditing, and database integrity.

A data quality framework was created that met the following criteria: (1) complete coverage of topics necessary for a student to understand all aspects of data quality, (2) understandable and consistent with the concepts students need to learn, and (3) pedagogically consistent with the existing IS undergraduate curriculum. The ultimate purpose of the data quality framework is to analyze the curriculum topics that relate to data quality. The four dimensions of the framework consist of:

(1) data life cycle: data acquisition
(2) data life cycle: data use
(3) process management, and
(4) data quality control.

Within each of these dimensions, several sub-dimensions exist. Associated with each sub-dimension is a set of specific topical areas. These detailed framework permitted a thorough and consistent analysis of the textbooks. Table 1 provides a detailed description of the topics covered within each dimension of the data quality framework.

Data Quality Dimension	Topics Covered within DQ Dimension
A. Data Life Cycle: Data Acquisition	overall view the data acquisition cycle
A1. Define the View - Logical Database Design	E/R diagrams, normalization, database models
A2. Implement View - Physical Database Design	integrity constraints, indexes, denormalization
A3. Obtain Values - Populating the Database	data entry, data import, downloading
A4. Update Records - Storage, Update and Deletion of Records	data dictionary, metadata, database architectures (distributed, client/server, centralized)
B. Data Life Cycle: Data Use	overall view of data use life cycle
B1. Define Subview	query design, SQL, QBE, other DDL's
B2. Retrieve Data	query performance and optimization, concurrency control (deadlock handling, locking)
B3. Manipulate Data	sort, aggregate, reformat, classify, analyze
B4. Present Results	report design and layout, forms, graphical presentation of results
B5. Use of Data	role of data in decision making, user interface design
C. Process Management & Data Quality	importance of business processes on data quality
C1. Business Processes	the role of data suppliers, data processors and data users in a business process
C2. Process Design and Modeling	data flow diagrams, workflow diagrams, enterprise modeling
C3. Management Issues	accountability for data, legal issues, policies
C4. Issues in Transaction Processing	definition of transaction, rollback, recovery, timestamping
D. Data Quality Control	importance and role of data quality control
D1. Data Quality Metrics	accuracy, timeliness, relevance
D2. Management Issues Related to Data Quality	planning and administration of the DQ function
D3. Quality Control for Data	auditing, sampling, tracking, assessment
D4. Database Security	authorization rules, authentication, encryption

Table 1: Data Quality Dimensions with Specific Topics Covered

IV. Analysis of Textbooks

The data quality framework was used to evaluate several representative database textbooks to determine the extent of their coverage in regards to topics related to data quality. The analysis is not intended to rank the textbooks in order to select a "best" textbook. Rather the analysis is done to demonstrate the coverage given to data quality topics in popular, representative textbooks. Five textbooks suitable for the introductory undergraduate class on database systems were chosen for analysis. The particular textbooks were chosen based partially on popularity (market share), pedagogical emphasis (MIS or computer science), and obvious coverage of topics related to quality.

The textbooks chosen for analysis were:

a) Kronke, D., Database Processing (5th ed.), 1995
b) Date, C. J., An Introduction to Database Systems (6th ed.), 1995
c) McFadden, F. and Hoffer, J., Modern Database Management (4th ed.), 1994
d) Watson, R. T., Data Management, 1996
e) Silberscharz, A., Korth, H. & Sudarhan, A., Database System Concepts (3rd), 1997

The Kronke (1995), McFadden and Hoffer (1994) and Watson (1996) texts would generally be considered textbooks suited for MIS curriculums while the Date (1995) and Silberscharz, Korth & Sudarhan (1997) texts are geared more towards a computer science curriculum. The Watson (1996) text was chosen for analysis because it is the only known database textbook with explicit material devoted to quality improvement.

In order to evaluate the coverage that a textbook gives to a particular data quality dimension, book content was evaluated on whether the material presented was theoretical or definitional in nature and whether or not an example application was presented. Table 2 shows the legend for the textbook evaluations.

Legend	Description
T	Theoretical - provides a theoretical framework of the topic with detailed explanation
D	Definition - provides a definition of the topic with a brief explanation
A	Applied - shows an application or an example related to the topic
T/A	Theoretical with Applied application or example
D/A	Definition with Applied application or example
N	None - topic not covered

Table 2: Legend for Textbook Evaluations

Table 3 shows the analysis of the five database textbooks across the data acquisition phases of the data life cycle. As expected all of the texts provided detailed theoretical coverage with examples related to logical and physical database design and the issues related to data storage. However all texts were weak on material related to populating the database with data. Only the McFadden and Hoffer text provided extensive coverage on the overall database design process as it relates to system analysis, design and implementation.

	Kronke (5th)	Date (6th)	McFadden & Hoffer (4th)	Watson	Silberscharz, Korth & Sudarhan (3rd)
A. Overall View of Data Life Cycle: Data Acquisition	D/A	D	T/A	D	D
A1. Database Design (logical)	T/A	T/A	T/A	T/A	T/A
A2. Database Design (physical)	T/A	T/A	T/A	T/A	T/A
A3. Populating the Database	D	D	D/A	D	D
A4. Storage, Update & Deletion of Records	T/A	T/A	T/A	T/A	T/A

Table 3: Topical Coverage of Data Acquisition Concepts in Database Textbooks

Table 4 shows the analysis of the five database textbooks across the data use phases of the data life cycle. None of the texts provides a detailed explanation of the data usage life cycle, although McFadden & Hoffer do present an example which demonstrates all stages of data in a database application. All texts provided detailed coverage of query languages with a strong emphasis on SQL. No material was presented on techniques for planning a query (e.g. concept mapping). Only the more technically focused books provide coverage of issues related to query performance and optimization (retrieve data). While all the textbooks provide strong coverage of data manipulation, this coverage is limited to manipulation within the query itself (data manipulation). Only the text by Silberscharz, Korth & Sudarhan covers material related to human manipulation of the data as a pre or post query activity. Little coverage is given on how data is presented to the user (report and form design) and on how the information is used in organizational decision making.

	Kronke (5th)	Date (6th)	McFadden & Hoffer (4th)	Watson	Silberscharz, Korth & Sudarhan (3rd)
B. Overall View of Data Life Cycle: Data Use	N	N	D/A	D	N
B1. Define Subview	T/A	T/A	T/A	T/A	T/A
B2. Retrieve Data	D/A	T/A	T/A	D/A	T/A
B3. Manipulate Data	T/A	T/A	T/A	T/A	T/A
B4. Present Results	A	N	D/A	N	D
B5. Use of Data	N	N	D	D	D/A

Table 4: Topical Coverage of Data Use Concepts in Database Textbooks

Table 5 shows the analysis of the five database textbooks as they relate to business process concepts. None of the texts provides good overall coverage on the role of business processes in data quality. McFadden & Hoffer do provide an example application that shows the role of data in a business process. All of the texts provide detailed coverage of issues related to transaction processing. The text by Silberscharz, Korth & Sudarhan provides the most detailed technical coverage in this area.

Table 6 shows the analysis of the five database textbooks as they relate to data quality control concepts. All of the textbooks give detailed coverage on database security. Only the Watson text provides detailed coverage on the overall importance of data quality and management issues related to data quality. However, even the Watson text does not cover the need for a program that involves measuring, tracking, assessing and improving data quality.

	Kronke (5th)	Date (6th)	McFadden & Hoffer (4th)	Watson	Silberscharz, Korth & Sudarhan (3rd)
C. Overall Importance of Process Management on Data Quality	N	N	D	D	N
C1. Business Processes	D	N	D/A	D	N
C2. Process Design and Modeling	N	N	T/A	N	N
C3. Management Issues (policies, legal, accountability)	T	N	T	T	D
C4. Issues in Transaction Processing	T/A	T/A	T/A	T	T/A

Table 5: Topical Coverage of Business Process Concepts in Database Textbooks

	Kronke (5th)	Date (6th)	McFadden & Hoffer (4th)	Watson	Silberscharz, Korth & Sudarhan (3rd)
D. Overall Importance of Data Quality Control	N	N	N	T	N
D1. Quality Metrics	N	N	N	D	N
D2. Management Issues Related to Data Quality	D	N	D	T	N
D3. Quality Control for Data	N	N	N	N	N
D4. Database Security	T/A	T/A	T/A	T/A	T/A

Table 6: Topical Coverage of Data Quality Control Concepts in Database Textbooks

V. Implications for the Database Course

We conclude that data quality is not explicitly covered in database system textbooks. However, database design issues that relate to data quality are typically covered. These topics include: logical and physical database design, data storage, defining a subview (designing a query), manipulating data (in a query), transaction processing and database security. These topics are the ones that are more technical in nature.

The organizational component of the database system is not typically emphasized in the database class. More specifically, the role of the data user in data quality is not covered in the undergraduate database course. These topics, among others, include report and form design, the role of data in business process design, populating a database, and system quality. While many of these system issues should be covered in a course on system analysis and design, the database course is the ideal place for the student see the relationship between the system issues and the database design issues as they relate to the quality of data.

The context analysis of our small sample of database textbooks shows that none of texts gave coverage on techniques for assessing, tracking and improving data quality in an organization. In order to raise the awareness of data quality in the database system class, the following classroom "modules" are suggested for use by the IS instructor. While these modules can be used in the database course, they may also be integrated into a course on systems analysis and design, information resource management or the introductory course in management information systems.

Module 1: Importance of Data Quality and the Role of Business Processes in Data Quality

Through the use of examples, demonstrate to the class the importance of data quality. Redman (1996) provides many examples taken from the popular press. However, perhaps the most convincing argument given in his book is a personal example in which a hotel chain incorrectly recorded a hotel reservation made by Redman and his wife. The result was an estimated loss of over $5,000 to the organization. Demonstrate how the business process is the number one root cause of data quality problems. Won't every student in your class have a personal example where inaccurate data resulted in lost time and/or money?

Module 2: The Role of Data Quality in the Life Cycle of Data

Show students both the Data Acquisition and Data Use Life Cycle. Emphasize that data is a dynamic entity that changes over time. Demonstrate that the quality of data is determined in part by the creation of an effective database design, but is also determined by the end-user's perception of that data. It is important that IS professionals be able to work with end user's to deliver high quality reports and queries. Give examples of a database with highly accurate data,

but with lousy response times and poorly written reports. Who will be responsible for this low quality database?

Module 3: Measure, Assess, Track and Improve Data Quality

Give students a hands-on project where they are asked to measure, assess, track and improve data quality. Ideally, this would be a project that uses a database that is already deployed. Student teams would develop metrics for assessing data quality, develop diagrams to assess existing business processes, develop techniques for tracking data quality metrics over time and develop solutions for improving data quality. Notice that this project goes beyond a data audit in that it analyzes the existing system and generates solutions. Solutions may be in the form of improvements to the business process, the database design, the system architecture or the existing policies and procedures.

Module 4: Tips for Improving Data Quality

Data quality can be emphasized throughout the database course by providing students with tips for improving data quality. Following is a list of Database Design Tips (Table 7) and Tips for Database Testing (Table 8) for that can be used by information system professionals and IS faculty teaching the database course.

VI. Conclusions

Data quality is an increasingly important topic in business and research. However, most information system professionals have not received training in the skills necessary to identify and improve data quality. The IS'97 Model Curriculum has placed a great deal of emphasis on "quality" , but the emphasis is primarily on quality management principles and software quality as opposed to data quality.

In addition, textbooks suitable for an undergraduate class in database systems do not explicitly address the issue of data quality. While many of the database design issues associated with data quality are covered in the database textbooks, the system issues that involve the interface between the users, the database and the business processes are typically ignored. Some of these topics are addressed in analysis and design course, but usually not within the context of

data quality. A series of four modules are presented as a means of bringing data quality into the undergraduate database course. These modules can be used to raise the awareness of the student towards the topic of data quality and to emphasize the relationship between data quality and the database design techniques that are the traditional domain of the undergraduate database course.

The authors acknowledge that the issues associated with teaching data quality concepts extend well beyond the scope of the undergraduate database course. Students majoring in IS or computer science should be exposed to data quality concepts in a wide spectrum courses including classes in systems analysis and design, software engineering, and traditional computer programming courses. In addition, non IS majors need to be sensitized to the importance of data quality in introductory MIS courses. Ultimately a holistic approach is needed in designing a comprehensive curriculum that includes data quality concepts.

DATABASE DESIGN TIP	DESCRIPTION OF DESIGN TIP
1. Create a data value as few times as possible	Inconsistencies between multiple values often go unnoticed until they are the source of a problem
2. Store data in as few databases as possible.	Multiple storage makes it difficult to maintain consistency, especially when data change.
3. Put data in machine-readable form as early in the business process as possible.	Computers and scanners are better than people at activities such as reading and inputting data. However, do not assume that computerized data collection is 100% accurate.
4. Minimize data format changes within the business process.	If format changes are necessary, used computers, not people, to make format changes.
5. When obtaining data for the first time, it is best to obtain them just before the time they are first needed.	Existing data values change rapidly. Capture changes to data values as soon as possible after they change.
6. Discontinue gathering and storing data that are no longer useful.	Plan for periodic review of data needs. When data are no longer useful they need not be destroyed, simply moved to secondary storage.
7. Employ codes that are easy for data creators and users to understand.	Avoid long, numeric, meaningless coding conventions in favor of short, meaningful words or abbreviations.
8. Place edits as near as possible to data creation or modification.	Use edits as input criteria to a database, as opposed to exit criteria from a database to an application.
9. Employ single-fact data wherever possible.	Single-fact data help reduce code complexity and simplify operator's jobs.

Table 7 : Database Design Tips to Improve Data Quality (from Redman, 1996).

Database Testing Tip	Description of Database Testing Tip
1. Create test data	Create sample files that are small enough to be manageable, but large enough to be representative. Use mostly 'real' data. Include a full set of matching records from each group of related tables.
2. Run the new database in parallel with the existing system.	This helps to uncover gaps in the application such as missing reports, transaction posting or archiving procedures. Run through a full testing cycle.
3. Check for duplicate values and errors in data.	Where errors are found make sure that all necessary integrity constraints are implemented.
4. Verify Contents of Database Through Report and Query Generation	Use record counts, batch totals, hash totals and cross-footing tests to verify data. Prime trouble spots are end-of-month and end-of-year procedures.
5. Develop Auditing Procedures as Part of a Data Quality Program	Regular procedures should be developed to test the accuracy of the database. This should be part of an overall program to improve data quality in the organization.

Table 8: Tips for Testing a Database (from Liskin, 1990).

VII. References

Cykana, P., Paul, A., and Stern, M., "DOD Guidelines on Data Quality Management", Proceedings of the 1996 Conference on Information Quality, Cambridge, MA, 1996, pp. 154-171.

Date, C. J., An Introduction to Database Systems (6th edition), Addison-Wesley, Reading, MA, 1995.

Davis, G. B., Gorgone, J.T., Couger, J. D., Feinstein, D.L. , Longenecker, Jr., H. E., IS'97 Model Curriculum and Guidelines for Undergraduate Degree Programs in Information Systems, developed by Association for Computing Machinery (ACM) Association for Information Systems (AIS) Association of Information Technology Professionals (AITP) (formerly DPMA), 1997.

Dvir, R. and Evans, S., "A TQM Approach to the Improvement of Information Quality", Proceedings of the 1996 Conference on Information Quality, Cambridge, MA, 1996, pp. 207-220.

Garcia, Mary Ryan, "Taking Care of Your Warehouse", Beyond Computing, Vol. 6, No. 4, May 7, 1997, pp. 32-36.

Hossein, S., "Guidelines for a Practical Approach to the Database Management Systems Course", Journal of Information Systems Education, Vol. 4, No. 1, 1992.

Kronke, D., Database Processing: Fundamentals, Design and Implementation (5th edition), Prentice-Hall, Englewood Cliffs, NJ, 1995.

Liskin, M., "Can You Trust Your Database?", Personal Computing, June 29, 1990, pp. 129-134.

McFadden, F. and Hoffer, J., Modern Database Management (4th edition), Benjamin/Cummings, Redwood City, CA, 1994.

Redman, T. C. , "Opinion: Improve Data Quality for Competitive Advantage", Sloan Management Review, Vol. 36, No. 2, Winter 1995, pp. 99-107.

Redman, T. C., Data Quality for the Information Age, Artech House, Norwood, MA, 1996.

Rollier, B., "A Database Project: Maximizing Its Value", Journal of Information Systems Education, Vol. 5, No. 2, Summer 1993, pp. 36-41.

Silberscharz, A., Korth, H. & Sudarhan, A., Database System Concepts (3rd edition), McGraw-Hill, New York, NY, 1997.

Strong, D. M., Lee, Y.W. and Wang, R. Y., "Data Quality in Context", Communications of the ACM, August 1996.

Thomson, Nancy S., "Using TQM Principles to Teach Current Topics in Information Systems", Journal of Information Systems Education, Vol. 6, No. 2, 1994, pp. 65-69.

Wang, R., Storey, V., and Firth, C., "A Framework for Analysis of Data Quality Research", IEEE Transactions on Knowledge and Data Engineering, Vol. 7, No. 4, August 1995, pp. 623- 640.

Wang, R. Y. and Strong, D. M., "Beyond Accuracy: What Data Quality Means to Data Consumer", Journal of Management Information Systems, Vol. 12, No. 4, Spring 1996, pp. 5-34.

Watson, R. T., Data Management: An Organizational Perspective, J. Wiley & Sons, New York, NY, 1996.