2005

Information Quality in the Banking Industries

Lotfollah Najjar
University of Nebraska at Omaha, najjar@unomaha.edu

Ram Bishu
University of Nebraska at Lincoln, rbishu@engux.unl.edu

Information Quality in The Banking Industries

Lotfollah Najjar
Department of Information Systems and Quantitative Analysis
University of Nebraska at Omaha
najjar@unomaha.edu

Ram Bishu
Department of Industrial & Management System Engineering
University of Nebraska-Lincoln
rbishu@engunx.unl.edu

ABSTRACT

Poor information quality is very costly and may cause serious financial problems for industries. Customers, suppliers, stakeholders, and employees through the supply chain are negatively impacted as the result of poor information quality. Data quality problems are exacerbated in large organizational databases and distributed through the multiple data sources. Any decision making based on poor information quality can lead to a poor service, error in the statement, and ultimately loss of the customers.

The objective of this research was to identify underlying dimensions of information quality and to assess the importance of each of these dimensions. Two different large banks were selected, with five branches among them. The IQ (Information Quality) questionnaires were sent to 278 employees of the banks that use information to serve internal or external customers. The result of the information quality analysis showed that all of the information quality dimensions except accessibility were significantly different between banks, but they were the same for branches within each bank. Reputation, believability, value-added, and relevancy had the highest score; completeness, access, and security had the lowest score for all the banks.

KEYWORDS

Information quality, banking industry, data quality.

INTRODUCTION

Poor data quality 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 errors, and inconvenience. Data quality problems are exacerbated in large organizational databases where data are 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 they relate to the context of the data consumers' tasks.

Businesses have implemented programs to improve data quality to enhance competitive advantage. AT&T used its data quality program to suggest opportunities to reengineer their billing system; 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 considerable time ensuring quality of data, but the focus on data quality gradually fades. Without proper data quality processes, the data warehouse will begin to accumulate "dirty data" (Garcia, 1997).

Data quality problems may cause serious financial problems for organizations. Data quality problems recently cost a fiber-optics manufacturer $500,000 when a mislabeled shipment caused the wrong cable to be laid along the bottom of a lake, caused a brokerage firm to lose $500 million when a dealer entered an incorrect exchange rate, and caused the U.S. government to lose over $2 billion in federal loan monies (Firth 1996). Organizations increasingly rely on their information systems to integrate and support their business processes (Wang and Kon 1993). These information systems and the quality of the data they contain affect customer's perceptions of the quality of purchased products and services (Wang and Strong 1996).

Information Quality Dimensions

Wang and Strong (1996) in their previous research determined the essential dimensions of IQ for delivering high quality information (see Table 1). Huang, Lee, and Wang (1999) conducted a series of comprehensive empirical studies and developed a framework with four IQ categories (Table 1).

IQ Category	IQ Dimensions
Intrinsic IQ	Accuracy, objectivity, believability, reputation

Contextual IQ	Relevancy, value-added, timeliness, completeness, Amount of information
Representational IQ	Interpretability, ease of understanding, ease of manipulation, concise Representation, consistent representation
Accessibility IQ	Access, security

Table 1: Category and IQ Dimensions

Intrinsic IQ denotes that information has quality in its own right.

Contextual IQ highlights the requirement that IQ must be considered with the context of the task at hand.

Representational IQ emphasizes the importance of the role of the systems.

Accessibility IQ represents the importance of the role of systems

Definitions of dimensions are shown in Table 2 (Wang and Strong, 1996).

Dimensions	Definitions
Accessibility	the extent to which information is available, or easily and quickly retrievable
Amount of Information	the extent to which the volume of information is appropriate Information for the task at hand
Believability	the extent to which information is regarded as true and credible
Completeness	the extent to which information is not missing and is of sufficient breadth and depth for the task at hand
Concise Representation	the extent to which information is compactly represented
Consistent Representation	the extent to which information is presented in the same format
Ease of Manipulation	the extent to which information is easy to manipulate and apply to different tasks
Free-of-Error (Accuracy)	the extent to which information is correct and reliable
Interpretability	the extent to which information is in appropriate languages, symbols, and units, and the definitions are clear
Objectivity	the extent to which information is unbiased, unprejudiced, and impartial
Relevancy	the extent to which information is applicable and helpful for the task at hand
Reputation	the extent to which information is highly regarded in terms of its source or content
Security	the extent to which access to information is restricted appropriately to maintain its security
Timeliness	the extent to which the information is sufficiently up-to-date for the task at hand
Understandability	the extent to which information is easily comprehended
Value-Added	the extent to which information is beneficial and provides advantages from its use

Table 2: Dimensions of Information Quality Dimensional IQ Assessment

Based on the 16 IQ dimensions (Wang and Strong, 1996) presented in Table 2, a set of questions can be generated to determine the perception of the state of IQ in an organization. Such a questionnaire has been developed based on the cumulative research conducted at MIT's TDQM (Total Data Quality Management) program (CRG, 1997, Lee et al. 2002). Each question is rated using a Likert-type scale on a scale of 0 to 10 where 0 indicates "not at all" and 10, "completely." This questionnaire has been used effectively in both public and private sectors.

Information quality (IQ) is an inexact science in terms of assessment and benchmarks. Although various aspects of quality and information have been investigated, there is still a critical need for methodologies that assesses how well organizations develop information products and delivers information services to consumers. Benchmarks developed from such a methodology can help compare information quality across organizations, and provide a baseline for assessing IQ improvements (Kahn *et al.*, 2002).

Banking Industry and Information Functions

While the overall market continues *to* grow, retail banking in the USA and other countries continue to lose considerable market share (Berger et al, 1995). For example, the highest revenue segment of the retail banking market, those households with over $1 million in investible net worth excluding houses, grew by 13 percent in 1994. However, retail banking's share of that market segment dropped by 33 percent during the same period (Palmer and Scheide, 1995). This statistic is even more unnerving when considered with its counterpart: 50 percent of a bank's value typically comes from the top 3-5 percent of its customers, while 60-80 percent of its costs come from the bottom 20 percent (Bird, 1997). What this means is that not only

are banks losing customers to competitors, but also they are losing their most profitable customers while being left with their most costly customers. Furthermore, as technology continues to drive down costs, it becomes easier for new competitors to enter the market and target the top customers of the banks with better prices (Nelson, 1999).

What are banks to do? To address these issues banks must become better at attracting and retaining top customers. Industry publications from 1970 onward provide little evidence of banks having information systems capable of supporting the needs of marketing. At the same time, other companies or industries have had demonstrable success in developing effective marketing information systems (Blattberg *et al.*, 1994), with Sainsburys and Tesco in the UK and companies such as USAA in the USA being outstanding examples within the financial services industry. The apparent failure of most banks to successfully integrate marketing and information services suggests that bank managers are not effectively managing the interface between the two functional areas (Nelson, 1999).

RESEARCH OBJECTIVES/QUESTIONS

The objective of this research was to identify underlying dimensions of information quality in the banking industries and to assess the importance of each of these dimensions in the banking industries with the following hypothesis .

H01: The mean of each dimension of information quality does not differ across banks.
H02: The mean of each dimension of information quality does not differ among the branches of the same bank.

METHODOLOGY

Two different large regional banks (the same size) in Nebraska were selected for this study. Three branches from bank A and two branches from bank B randomly selected to visit after several contacts with the bank‘s executives. Personal interviews were conducted to get the cooperation of employees who use information to serve internal or external customers. All the employees were approached to solicit responses for the questionnaire. Employees were informed regarding the objective of this research and how it would be beneficial to them and to the organization. The information quality questionnaires (Wang and Strong, 1996) were sent to 278 employees of the banks that use information to serve internal or external customers. The period of the field research lasted from January 2002 until June 2002 and the survey was by the paper. The overall response rate was 236 / 278 = 84% and the following table shows the breakdown of sample sizes and response rates for the banks and the branches:

Banks	Number of Contacted Employees	Number of Respondents	Response Rate
A1	60	53	88%
A2	53	42	79%
A3	42	38	90%
B1	67	55	82%
B2	56	48	85%

Table 3: Employees Response Rate

Each question is rated using a Likert-type scale on a scale of 0 to 10 where 0 indicates “not at all” and 10, “completely.” This questionnaire has been used effectively in both public and private sectors.

Data analysis was conducted using SAS statistical package in two stages:
Descriptive statistics: Descriptive statistics (tabular and graphical) was used for the dimensions of information quality to compare and interpret means and standard deviations as a preliminary analysis.
ANOVA (General Linear Models): ANOVA (General Linear Models) using nested design was conducted for the dimensions of information quality to find the differences of rating between banks and within the branches and also to identify and to assess the dimensions of information quality.

RESULTS

DESCRIPTIVE STATISTICS

As shown in Table 4, reputation, believability, and value-added respectively are the most important dimensions of information quality for all the banks based on the mean values.

Dimensions	Mean	Standard Deviation	Coefficient of Variation.
Reputation.	9.3463	0.4086	0.0437
Believalibility.	9.2542	0.4151	0.0449
Value -Added	9.1915	0.3119	0.0339
Relevency	9.1728	0.3304	*0.0360*
Amount of Information.	9.0233	0.4080	0.0452
Consistent Representation	9.0180	0.4507	0.050
Objectivity	8.9073	0.5031	0.0565
Accuracy	8.7762	0.5523	0.0629
Concise Representation	8.5173	0.4243	0.0498
Ease of Manipulation	8.3161	0.3489	0.0420
Ease of Understanding	8.2618	0.3573	0.0432
Timeliness	8.2237	0.3229	0.0393
Interpretability	8.1762	0.3840	0.4080
Completeness	8.1413	0.4143	0.0470
Accessibility	7.6699	0.5579	0.0727
Security	6.8538	0.5242	0.076

Table 4: Information Quality (All Banks)

NOTE: THE CRITERIA WERE TO CHOOSE THE TOP THREE

THE GENERAL LINEAR MODELS

The general linear models procedure was used to see the differences between dimensions of information quality between banks and among the branches. The branches were nested within the banks. The level of significance was established at the 0.05. The following results based on the ANOVA table were found as shown in Table 5.

Variables	Main Effects	P.Value	Pair wise Comparison of Banks and Branches
Accessibility	Branch (Banks) Banks	0.0001 0.1119	A2=A3, B1=B2
Accuracy	Branch (Banks) Banks	0.7074 0.0001	A>B
Amount of Information	Branch (Banks) Banks	0.7362 0.0001	A> B
Believability	Branch (Banks) Banks	0.3254 0.0001	A> B
Completeness	Branch (Banks) Banks	0.9561 0.0001	A>B
Concise Representation	Branch (Banks) Banks	0.8846 0.0001	A> B
Consistent Representation	Branch (Banks) Banks	0.9480 0.0001	B>A
Ease of Manipulation	Branch (Banks) Banks	0.9909 0.0001	A> B
Ease of Understanding	Branch (Banks) Banks	0.9465 0.0001	A> B
Interpretability	Branch (Banks) Banks	0.8179 0.0001	A> B
Objectivity	Branch (Banks) Banks	0.9666 0.0001	A>B
Relevancy	Branch (Banks) Banks	0.0163 0.0001	A2=A3, B1=B2 A> B

Reputation	Branch (Banks) Banks	0.9575 0.0001	 A> B
Security	Branch (Banks) Banks	0.9887 0.0443	 A>B
Timeliness	Branch (Banks) Banks	0.7437 0.0001	 A> B
Value Added	Branch (Banks) Banks	0.0018 0.0001	A2=A3, B1=B2 A>B

Table 5: ANOVA Table for Information Quality

Note: A>B means bank A has a higher mean value than bank B for a given dimension

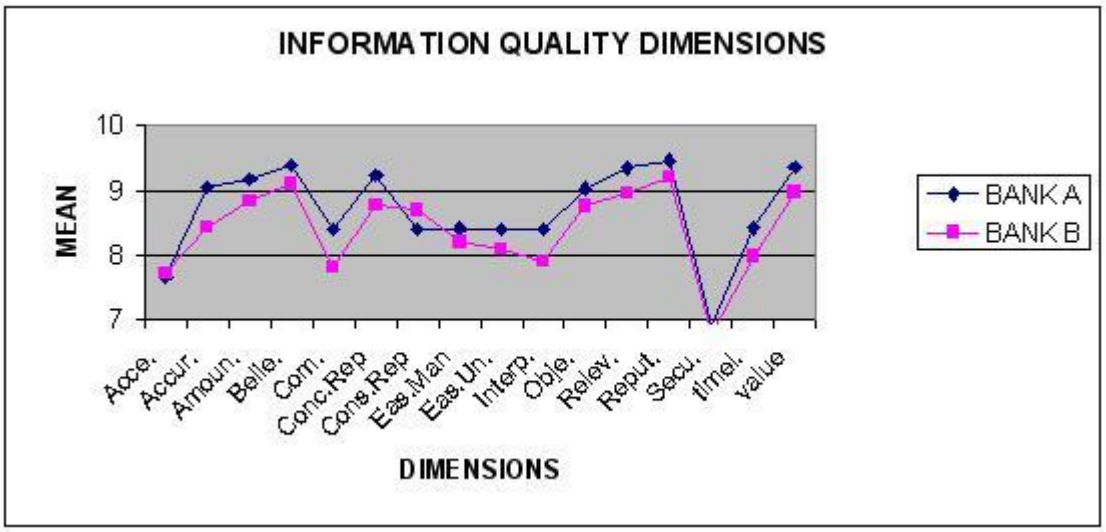

Figure 1. Information Quality Dimension for Banks

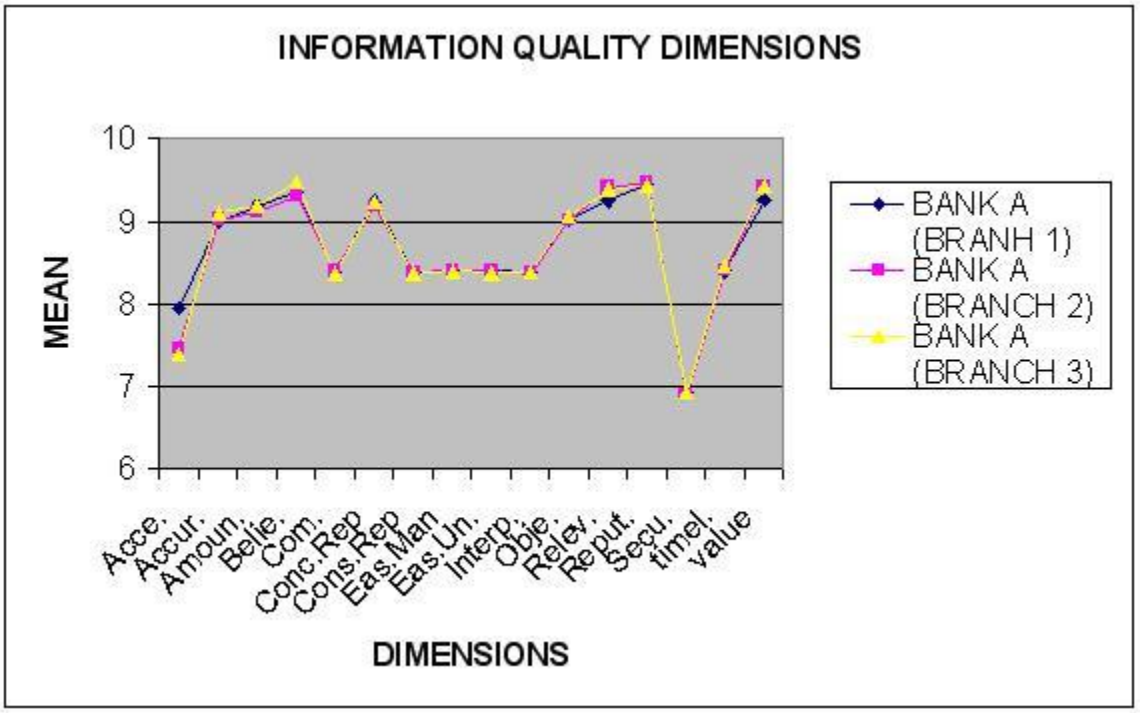

Figure 2. Information Quality Dimension for Bank A

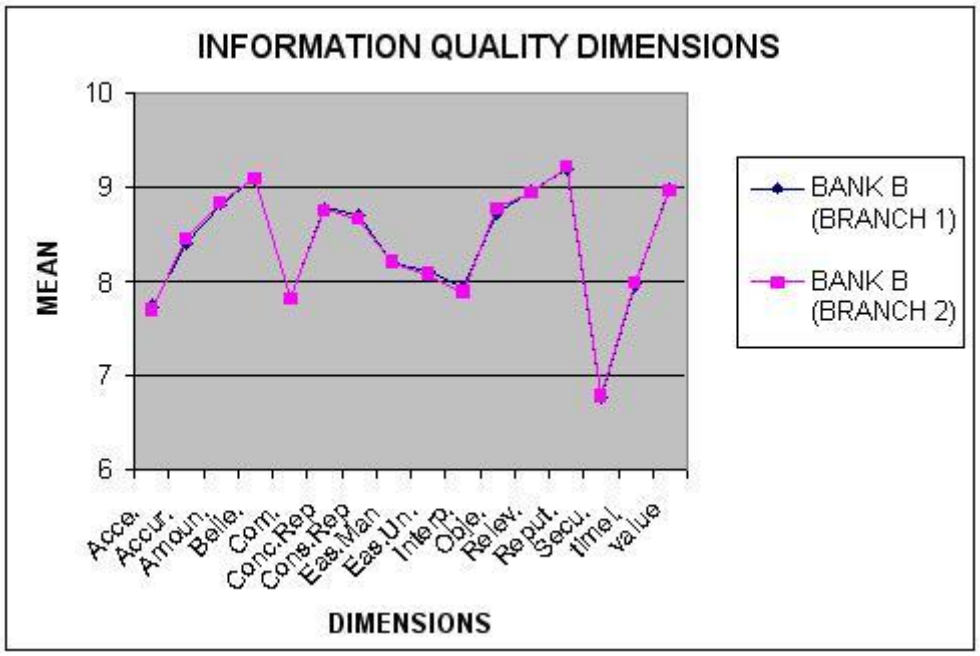

Figure 3. Information Quality Dimension for Bank B

Accessibility: Table 5 shows that the two banks do not differ in this dimension of information quality. Branches 2 and 3 within bank A and branches 1 and 2 within bank B are the same in terms of accessibility dimension. Fig. 1-3 shows that branch1 within bank A has a higher mean value than all the branches within bank A and B

Accuracy: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of accuracy dimension. Also Fig. 1-3shows that bank A has a higher mean value for this dimension than bank B. Branch 3 within bank A has a higher mean value than branches 1 and 2(with branch 1 the lowest). Branch 2 within bank B has higher mean value than branch 1.

Amount of information: Table 1 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of amount of information dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. Branches 1 and 3 within bank A have a higher mean value than branch 2. Branch 2 within bank B has higher mean value than branch 1.

Believability: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of believability dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. Branch 3 within bank A has a higher mean value than branches 1 and 2. Branch 2 within bank B has higher mean value than branch 1.

Completeness: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of completeness dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. Branch 2 within bank A has a higher mean value than branches 1 and 3. Branches 1 and 2 within bank B have almost the same mean value for this dimension.

Concise Representation: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of concise representation dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. Branch 3 within bank A has a higher mean value than branches 1and 2(with branch 2 the lowest). Branches1 and 2 within bank B have almost the same mean value for this dimension.

Consistent Representation: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of consistent representation dimension. Also Fig. 1-3 shows that bank B has a higher mean value for this dimension than bank A. Branch 1 within bank A has a higher mean value than branches 2 and 3. Branch 1 within bank B has a higher mean value than branch B for this dimension.

Ease of Manipulation: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of ease of manipulation dimension. Also Fig. 1-3 show that bank A has a higher mean value for this dimension than bank B. All the branches within bank A and B have almost the same mean value for this dimension.

Ease of Understanding: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of ease of understanding dimension. Also Fig. 1-3 shows that all the branches within bank A and B have almost the same mean value for this dimension. bank A has a higher mean value for this dimension than bank B.

Interpretability: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of interpretability dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. All the branches within bank A have almost the same mean value for this dimension. Branch 1 within bank B has a higher mean value than branch 2 for this dimension.

Objectivity: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of objectivity dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. Branch 3 within bank A has a higher mean value than branches 1 and 2 for this dimension. Branch 2 within bank B has a higher mean value than branch 1 for this dimension.

Relevancy: Table 5 shows that the two banks differ in this dimension of information quality. Branches 2 and 3 within bank A and branches 1 and 2 within bank B are the same in terms of relevancy dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. Branch 2 within bank A has a higher mean value than branches 1 and 3 (with branch 1 the lowest) for this dimension. Branch 1 within bank B has a higher mean value than branch 2 for this dimension.

Reputation: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of reputation dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. Branches 1 and 2 within bank A have a higher mean value than branch 3 for this dimension. Branch 2 within bank B has a higher mean value than branch 1 for this dimension.

Security: Table 5.shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of security dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. All the branches within banks A and B have almost the same mean value for this dimension.

Timeliness: Table 5 shows that the two banks differ in this dimension of information quality, but all the branches within each bank are the same in terms of timeliness dimension. Also Fig.1-3 shows that bank A has a higher mean value for this dimension than bank B. All the branches within banks A and B have almost the same mean value for this dimension.

Value Added: Table 5 shows that the two banks differ in this dimension of information quality. Also branch 2 and 3 within bank A and branches 1 and 2 within bank B are the same in terms of value added dimension. Also Fig. 1-3 shows that bank A has a higher mean value for this dimension than bank B. Branches 2 and 3 within bank A have higher mean value than branch 1 for this dimension. Branch 1 within bank B has a higher mean value than branch 2 for this dimension.

DISCUSSION AND CONCLUSIONS

The results of showed that Wang and Strong's 16 data quality dimensions are important within the banking industry. All of the data quality dimensions except accessibility were significantly different between banks, but they were the same for branches within each bank. In general bank A has higher quality information than bank B. Reputation, believability, value-added, and relevancy had the highest score for all the banks together as a whole, bank A, and bank B. Completeness, access, and security had the lowest mean score for all the banks as a whole and bank B. Interpretability, access, and security have the lowest mean score for bank A. Gendron et al. (2001) examined Wang and Strong's data quality dimensions for three sectors of the healthcare industry. Their statistical analyses indicated that fifteen of Wang and Strong's data quality dimensions are sufficient to define data quality in all sectors of the healthcare industry. Accuracy, accessibility, and security had the highest score and amount of information, consistent representation, and reputation had the lowest mean score for all sectors of the healthcare industry. Huang, K. Lee, Y. Wang, R. (1999) conducted a study to asses the information quality in Appliance Company. Their statistical analyses indicated that believability, reputation, and relevancy had the highest score and ease of manipulation, security, and amount of data had the lowest mean score for Appliance Company. Thus, few dimensions of information quality are common across the different industries; in general different industries view information quality differently.

Wang and Strong's information quality dimensions have been used in all sectors of the healthcare industry, manufacturing industry, and many other service industries. It is suggested that the sixteen dimensions of information quality proposed by Wang and Strong need to be tested and customized to each service setting.

Tracking customers and developing creative strategies to retain them is very profitable. For example, in 1982, Charles Cawley, the president of the credit card company MBNA of America, became increasingly frustrated by numerous complaints from defecting customers and took action. Cawley announced to all MBNA employees that the mission of the company would be to keep every customer. To accomplish this goal, a strategy was implemented to call defecting customers personally and obtain information about the reason for their defection. Chronic problems were determined and prioritized; appropriate changes were implemented. Eight years later, MBNA's defection rate was reduced to just 5 percent, one of the lowest in the industry. Without making any acquisitions, MBNA's industry ranking went from 38 to 4, and profits increased 16-fold (Reichheld and Sasser, 1990). Many studies indicate that it costs eight to ten times less to keep a customer than to develop a new one Thus improving service quality leads to the customer satisfaction and ultimately to customer loyalty.

Poor data quality 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 errors, and inconvenience. Data quality problems may cause serious financial problems for organizations. Data quality problems recently cost a fiber-optics manufacturer $500,000 when a mislabeled shipment caused the wrong cable to be laid along the bottom of a lake, caused a brokerage firm to lose $500 million when a dealer entered an incorrect exchange rate, and caused the U.S. government to lose over $2 billion in federal loan monies (Firth, 1996)

LIMITATIONS

Sample of branches within banks were not sufficient to represent the population of the banks. Employee also had a difficulty of understanding some of the questions.

RECOMMENDATIONS

The result showed that Wang and Strong's 16 data quality dimensions are important within the banking industry and could possibly serve as a meaningful framework for tracking a firm's information quality performance over time and comparing it against the performance of competitors. It is recommended that banks A and B continuously measure and improve the dimensions of information quality especially the lower end dimensions since they are very critical to be competitive in the marketplace. It is recommended that the banking industries must ensure continuously improving these critical dimensions of information quality. Security dimension had the lowest rating since the definition of security was not well defined in Wang and Strong' questionnaires. It is recommended that this questionnaire should be revised with broader information security view of integrity, confidentiality, and availability.

REFERENCES

1. Berger, A.N., Kashyap, A.K. and Scalise, J.M. (l995), "The transformation of the US Banking industry: what a long, strange trip it's been", Brookings Papers on

Economic Activity, The Brookings Institution, Washington, DC, pp. 55-218.

2. Bird, A. (1997), Super Community Banking Strategies: Winning the Ware for the Customer Relationship, Irwin Professional Publishing, Chicago, IL.

3. Blattberg, R.C., Glazer, R. and Little, J.D.C. (1994), *The* Marketing Information Revolution, Harvard Business School Press, Cambridge, MA.

4. Cowels, D. and Crosby, L. (1990), ``Consumer acceptance of interactive media," The Services Industries Journal, Vol. 10 No. 3, pp. 521-40.

5. CRG, (1997) Information Quality Assessment Survey: Administrator's Guide. Cambridge Research Group, Cambridge, MA.

6. Firth**,** C. P.(1996). Data quality in practice: Experience from the frontline. The 1996 Conference on Information Quality, Massachusetts Institute of Technology, October 25-26. (Also see

http://sunflower.singnet.com.sg/~cfirth/dataquality/. Special viewing authorization may be needed.)

7. Garcia, M.R. (1997), "Taking Care of Your Warehouse," Beyond Computing, 64, 32-36.

8. Gendron , Michael S and D'Onofrio, Marianne J.(2001) "Data Quality in the Healthcare Industry." Data Quality. September, Vol. 7, No. 1.

9. Huang, K. Lee, Y. Wang,R.(1999), " Quality Information and Knowledge",Prentice-Hall, Inc.

Journal of Services Marketing, Vol. 3 No. 3, Summer, pp. 43-51.

10. Kahn, B., D. Strong, and R. Wang, "Information quality benchmarks: Product and service performance," *Communications of the ACM*, 45(4), 184-192, 2002.

11. Nelson, Mark, R. (1999), "Bank marketing and information technology: a historical analysis of the post-1970 period," International Journal of Bank Marketing, Vol.17, No. 6,pp 265-273.

12. Palmer, D.H. and Scheide, R.G. (1995),"CEO alert! Banks' most lucrative market is booming, but banks' share dropped 33% in 1994", The Banker's Magazine, March- April, pp. 20-6.

13. Redman, T.C. (1996), *Data Quality for the Information Age*, Norwood, MA: Artech House.

14. Reichheld, F.F, Sasser, W.E, (1990), "*Zero defections: quality comes to services*", Harvard Business Review, 105-11.

15. Strong, D.M., Lee, Y.W., and Wang, R.Y. (1996), "Data Quality in Context," Communications of the ACM, 40, 103-110.

16. Wang, R.Y., and D.M.Strong, (1996), "Beyond Accuracy: What Data Quality Means to Data Consumers", Journal of Management Information Systems (JMIS) 12(4), 1996, pp. 5-34.

17. Wang, R.Y., and H. B. Kon. (1993), Toward total data quality management. In Information technology in action: Trends and perspectives, ed. R.Y. Wang, 179-97. Englewood Cliffs, New Jersey: Prentice Hall.