# A College Course: Data Quality in Information Systems

Craig W. Fisher
Marist College
Poughkeepsie, NY 12601
(845) 575-3000 x2621
Craig.Fisher@Marist.edu

**Abstract:** Information Systems (IS) college students are not prepared for the demands of improving information quality in our databases and data warehouses. Information Quality (IQ)/Data Quality (DQ) in IS curricula tends to be subject to individual faculty preferences. Given the significance and pervasiveness of DQ and the lack of focused attention to DQ/IQ in college curricula, I created a *Data Quality in Information Systems* course. The purpose of the course is to alert our would-be-IS-professionals to the pervasiveness and criticality of data quality problems. The secondary agenda is to begin to arm the students with approaches and the commitment to overcome these problems.

Four major exercises were quite powerful in helping the students understand and assimilate concepts of Data and Information Quality. The cases included Total Quality Management analysis, Data Warehouse Cleansing, Data Quality Information (data tags), exercise and an Information Quality Assessment project. The purpose of this paper is to share the essence of the new course.

## Introduction

Khalil et al (1999) states that Information Systems (IS) college students are " . . . not equipped with a broad understanding of the principles behind measuring, analyzing, and improving IQ in an organization." Their finding that Information Quality (IQ)/Data Quality (DQ) in IS curricula tends to be subject to individual faculty preferences seems to hold true at Marist College. At Marist College, our IS program contains a number of courses that touch on various aspects of data quality. Data quality is a topic of at least one complete (75 minutes) lecture in half of our ten IS courses, and is mentioned in all ten courses. However, the five courses that mention data quality do not test on data quality beyond simple definitions. The other five courses that lecture on data quality may or may not directly test on data quality. Typical test questions might indirectly reference data quality in topics such as software engineering, data validation methods, types of data errors, anomalies, testing, requirements documents, normalization and referential integrity.

Given the significance and pervasiveness of DQ problems (Redman, 1996, 1998, Tayi and Ballou, 1998; Orr, 1998) and the lack of focused attention to DQ/IQ in college curricula (Khalil, et al., 1999), I created a *Data Quality and Information Systems* course at Marist College. The purpose of the course is to alert our would-be-IS-professionals to the pervasiveness and criticality of data quality problems. The secondary agenda is to begin to arm the students with approaches and the commitment to overcome these problems. The purpose of this paper is to share the essence of the new course.

## The Students

The course is aimed at first semester seniors majoring in Information Systems who have completed eight of the ten required IS courses, including systems analysis, systems design, data management, data communication, and problem solving and programming. These students should be able to analyze an end-user procedural or process problem, design a system solution, and develop, test, and implement a solution. They have the knowledge and skills to clean up anomalies and ensure *third normal form* of databases. They are aware of, but not experts in, data warehouses. Ten seniors took the course in the Fall 2000 semester and fifteen are registered for the Fall 2001 semester.

## The Course

### *Course Objectives*

There were six basic objectives to the course.

1. Develop in-depth understanding of Data and Information Quality (DQ and IQ).
2. Understand DQ and IQ Concepts in Information Systems projects.
3. Recognize various patterns of data and design deficiencies in systems.
4. Suggest DQ and IQ improvement plans in light of known deficiencies in systems.
5. Understand of the role and importance of DQ and IQ in data warehouses.
6. Discuss the role and importance of DQ in Decision Support Systems.

### *Course Approach*

The class met once a week for 2 hours and 45 minutes including a 15-minute break. Students were asked to read and study text and journal articles to learn the fundamental concepts of Data and Information Quality. There were four key exercises to explore Total Quality Management (TQM), use of data tags, data warehouse cleaning, and information quality assessments. These four IS and DQ exercises were designed to engage the students' new knowledge of the concepts with their analytical abilities so that they can develop meaningful solutions to real problems.

Evaluation of students consisted of two major exams, the midterm and final exams, with each exam counting for one quarter of their grade. The four exercises counted at 10% each, for a total of 40% of the grade. Classroom participation accounted for the final 10%.

The required textbook was Quality Information and Knowledge by Huang, Lee, and Wang, published in 1999 by Prentice Hall. The complete list of recommended journal articles appears in the bibliography of this paper.

### *Semester Plan*

| | |
|---|---|
| **SESSION A**<br>Classes 1,2 & 3 | Motivation & Concepts<br>Manage Information as Product<br>TQM Case Study |
| **SESSION B**<br>Classes 4 & 5 | Measure, Analyze, Improve IQ<br>But Who Will Use Measurements?<br>DQI & DSS Exercise |

| | |
|---|---|
| **SESSION C**<br>Class 6 | Midterm |
| **SESSION D**<br>Classes 7,8,<br>9 & 10 | DQ & Data Warehouse (DW)<br>DW Exercise<br>Knowledge as Assets |
| **SESSION E**<br>Classes 11,12,<br>13 & 14 | IQA Study<br>"Sell" Clients, Conduct Surveys,<br>Analyze Data & Debate Results<br>Present & Discuss with Clients |
| **SESSION F**<br>Class 15 | Future Organizational Knowledge<br>Conclusions, Feedback |
| **SESSION G**<br>Class 16 | Final Exam |

### *Elaboration and Feedback of Sessions*

## Session A: Introduction to Quality

This session was scheduled to require three weeks of class, but consumed four weeks.

**Class 1** cannot be pro forma since it is a 2 hour and 45 minute time slot and should not be wasted; yet the students have not prepared any readings. I gave more pure lecture than is normally desirable. References to a variety of specific data quality issues as raised by Redman (1996, 1998), Kingma (1996), Wilson (1992), Tayi and Ballou, (1998), Orr (1998), Huang, Lee & Wang (1999), and Celko (1995) provided much of the motivational statement. In addition, we spent at least half an hour on the Space Shuttle *Challenger* and another half an hour on the USS *Vincennes* shootdown of a passenger jet (Fisher, 2001). A key aspect of this introduction was asking the students to provide examples of data quality problems from their daily lives. They were most enthusiastic about credit card charges being in error, their names being misspelled, college registration errors, and the like. The Class 1 lecture proved to be successful as the students were astounded at the significance and pervasiveness of data quality problems. They seemed motivated to study further.

**Class 2** focused on Chapters 1 and 2 of the text (Huang, Lee & Wang, 1999) in detail and covered *The Malcolm Baldridge National Quality Improvement Act of 1987 - Application Guidelines 1988* (Public Law 100 – 107, 1987). Students readily grasped Huang's two basic

propositions[1] and several concepts such as "best practice" and "core competency." The students struggled to grasp the concept of managing information as a product due to limited experience in that area. We spent a profitable amount of time discussing the text section *Establish an Information Quality Program* (p. 27 - 28), and we debated whether or not "information quality is 'everybody's responsibility'" (p. 28).

The TQM Case Exercise[2] describes a corporation that is having trouble integrating information systems properly. However, the basis of the problem may be the company's approach to, or lack of an approach to, TQM. Critical ingredients included environment, goals, standards, responsibilities, manufacturing processes & steps, and relationships with customers and vendors. The assignment was to read and analyze the case (16 pages), answer a series of questions, develop and analyze alternative solutions, and recommend a solution. The students enjoyed discussing the facts, symptoms, and problems of the case in Class 2, and the alternative solutions in Class 3.

In **Class 3**, we continued the discussion of the TQM case for close to one hour, about 30 minutes longer than I planned. We established two teams of three and one team of four to provide informal presentations to the class. For the second half of the class we discussed Chapter 3 of the text. I underestimated the time needed for this useful chapter. The students learned the concept of recognizing data, design, and operational deficiencies in mapping real world scenarios to information systems. While some examples are given in the text, I believe that more complete examples or a full case study would be useful.

I next introduced the 16 dimensions of data quality. These are grouped into four categories as follows:

- *Intrinsic IQ* – Accuracy, Objectivity, Believability, Reputation
- *Contextual IQ* – Relevancy, Value Added, Timeliness, Completeness, Amount of Information
- *Accessibility IQ* – Access, Security
- *Representational IQ* – Interpretability, Ease of Understanding, Ease of Manipulation, Concise Representation, Consistent Representation

In addition to these groupings, I may consider additional groupings in the future. For example, the model of product and service presented by Khalil et al (1999) groups the dimensions by specification and expectations. I believe the students would gain more insight by analyzing the two different groupings. The students also read Strong et al's 1997 article on *Data Quality in Context.*

## Session B: Measure, Analyze and Improve Data Quality

This session was planned to cover two weeks but consumed three weeks.

The concept of placing additional fields (data tags) in a database to carry data quality information (DQI) was foreign to the students. Thus the discussion and elaboration of DQI concepts consumed more time than planned. Recent research has demonstrated that simply telling people the quality of their data doesn't predict whether they will use that information

[1] Proposition 1: Firms must create a reservoir of quality information. Proposition 2: Firms must create a wealth of organizational knowledge. (Huang, et al. 1999. p. 4).

[2] Found in the casebook for McLeod, R J., *Management Information Systems: A Study of Computer-Based Information Systems*. 6 ed. 1995, Englewood Cliffs, NJ: Prentice Hall.

about their data quality (Chengalur-Smith, et al, 1999; Fisher, 1999). The students were asked to complete the Apartment Selection Task (Appendix A) to determine the amount of use, if any, that a person makes of data quality information[3] (DQI) when it is provided (Chengalur-Smith, et al, 1999; Fisher, 1999).

I gave the task with DQI to half the class and the task without DQI to the other half of the class. In a discussion before the task activity, many students said that they believed that if people were told the quality of their data that it would influence their decisions. Thus, it was very surprising to them that they did not use the reliability data in the task. I presented results from other studies that indicate experience, manager status, and time pressure influence the amount and type of use made of DQI. The students' first hand experiences with the task made it easy for them to follow.

Most of the students were now aware of the problem that some people may not use DQI if provided, and those who do use DQI may vary in their usage from others who use it. There are many implications here that we alluded to but did not investigate. Some examples of these implications are the "business case" for DQI data, training employees as to the purpose and uses of DQI, format of DQI, and the need to explore characteristics of decision makers and their familiarity with the data.

The three classes of metrics (perception, application independent, application dependent) were clear and readily grasped by the students. We continued studying the 16 dimensions of data quality (Huang, et al., 1999). I found that I made too many assumptions about how much the students would grasp. A more thoroughly planned set of examples and preplanned questions would have made this session more productive.

## Session C: Midterm

## Session D: Data Quality and Data Warehouses

Since data warehouses are becoming critical in corporations and government and are so fraught with data quality problems, the next logical step in the course was to walk through a data warehouse cleaning exercise. The students read Fayyad's *Data Mining and Knowledge Discovery* (1996); Golfarelli and Rizzi's *A Methodological Framework for Data Warehouse Design* (1999); and Ballou and Tayi's *Enhancing Data Quality in Data Warehouses* (1999).

Pamela Neely (2000) has been doing research on applying financial auditing processes to processes for cleaning data warehouses. She agreed to participate in two of my classes. At the first class, she presented a data warehouse problem with which she was involved. In the second class, the students built a dictionary and merged individual files into a data warehouse.

This exercise is an actual current project in a major city in New York state where a number of homeless shelters are receiving financial support. Each homeless shelter collects its own data, but recently the shelters have been asked to provide data for a comprehensive data warehouse. We gave the students definitions of five different databases that were designed to collect data on people[4] who frequent specific homeless shelters. The assignment was to build a common data dictionary and then build an integrated data warehouse. We asked the students to identify and resolve problems and issues in building the data dictionary. Finally, the students were asked to document errors found in the merging of the databases into the data warehouse (Neely, 2000).

[3] Data tags
[4] All names were changed for these exercises

The reality of the possible errors that occurred and the trouble that the students had in merging first the dictionaries and then the data was extremely educational. This is a must-do exercise for a course in data quality.

The article by Golfarelli and Rizzi (1999) proved to be too technical and abstract for our purposes. However, the Ballou article was particularly well received by the class.

## Session D: Real Client Information Quality Assessment (IQA)

The students were asked to work in teams to perform an Information Quality Assessment study at Marist College (Appendix B). I prearranged these studies with the Vice President of Business/Financial Affairs, the Registrar, and the Director of Information Technology. I had three teams and I asked each team to give a kick-off presentation to one of the three stakeholders. The presentation described the value that the organization would receive from an IQA study and what it would require to complete the study. I underestimated the effort to use the IQA system, so the students prepared hardcopy survey forms. They then conducted the IQA Survey, collected the data, put it into an Excel database, analyzed it, and prepared a final report for presentation to the client (Huang, et al., 1999). We used two class periods to analyze, discuss, and debate various findings, and a third class for the final presentation to the clients.

The students gave a final presentation to the clients. The students compared and contrasted IQ perceptions from each of the different stakeholder organizations and from managerial versus worker perspectives. Finally, the students presented their conclusions and recommendations.

Each client agreed to come to a class session for the final presentation. Thus all students saw all teams present and discuss the project with their clients. These sessions were the high points of the course. The clients were interested in the results and also elaborated on various problems that they experienced with data flows and information quality. For example, the VP of Business/Financial Affairs explained the data dependencies between Housing, Registration, and Billing.

While the purpose of this paper is not to elaborate on specific conclusions, I mention some of the specific findings to give the reader a flavor of the students' results.

The IQA study included 38 Information Technology people, 14 Business Office people and 12 Registrar people. Chart 1 illustrates a comparison of the perceptions of all employees of the IT function versus all employees of the Business Office for a specific set of business databases. High numbers represent more satisfaction while low numbers represent more dissatisfaction. It is clear from Chart 1 that the IT people have a much more positive view of the quality of the system than do the Business Office people. IT people rated timeliness a very high 8.2, while the Business Office people rated timeliness at a mediocre 5.5. Differences were also found in ease-of-manipulation, consistency, accuracy, and value-add. Our class study identified potential problem areas and opened up discussion between the two organizations. The students recommended that SQL projects be started to address manipulation and timeliness. They recommended Quality Circles be started to address consistency, accuracy, and value-add.

Chart 1 is just one example of our findings. We also discussed managers versus non-managers, all users-all systems versus all IT-all systems, and all registrar versus all IT. Each group for each set of application databases ranked the dimensions on importance. These findings, coupled with the user executives' feedback, were the highlight of the course.

## *Summary/Conclusion*

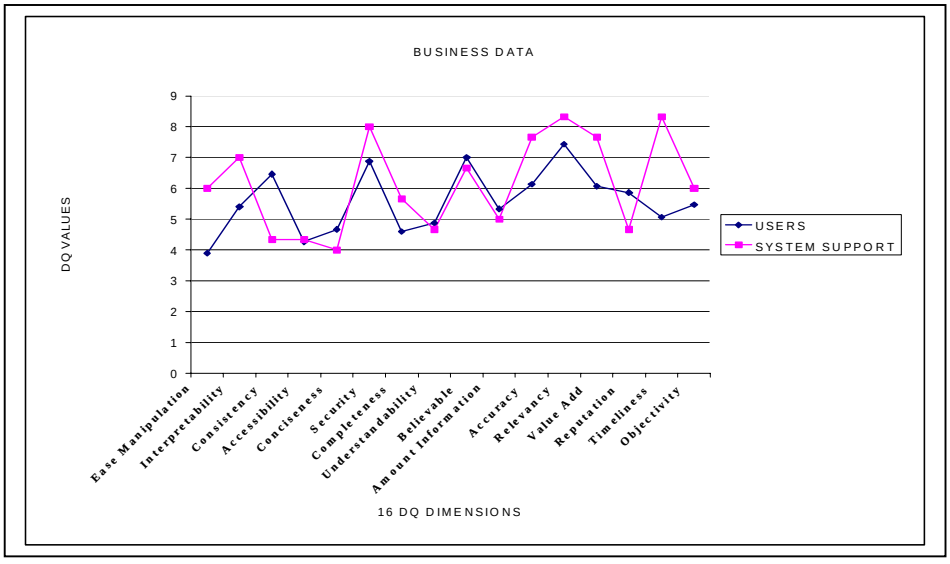
This first course in *Data Quality and Information Systems* met five of its six objectives. The combination of the text, the articles Strong et al (1997), Ballou & Tayi (1999), Redman (1998), Orr (1998), and the four major exercises were quite powerful in helping the students understand and assimilate concepts of Data and Information Quality. The first TQM case set the stage for understanding quality in context of information systems. Professor Neely's class exercise building a real dictionary and data warehouse from five fragmented and disperse databases was very exciting and informative. The IQA work was the highlight of the course. I cannot overstate the value of having the students develop and give the presentation that "sells" the value of doing an IQA study to the user executive. This was followed by conducting the survey, deciding how to enter data, analyzing the data, and presenting findings to the user with subsequent discussion.

The course partially met the last objective, which was to be able to discuss the role and importance of DQ in Decision Support Systems. We completed an exercise and had an interesting discussion about the use of DQI in decision-making. However, this is such a small part of DSS that I cannot claim we made much of an impact in this arena. We did not discuss decision theory, methods of decision-making, group decision-making, nor other factors in decision-making.

Of the ten students that took the course, nine completed the anonymous course/instructor evaluation form at the end of the semester. In the handwritten comment section, five of the students said that they appreciated the real-life exercises. They especially enjoyed the executives coming to class to discuss data quality issues. There was only one negative comment, and that related to the Knowledge Management section of the text. I feel that this was more an issue of time than a problem with the text; we fell behind in the schedule, but I still required that the students read all chapters. The students read the last two assignments without benefit of introduction from their teacher.

While the primary objectives were met in this course the real issue is whether they were the right objectives. This should be a subject for further research and discussion in IQ200x.

CHART 1



### *Recommended Readings*


1. Ballou, D. P., and Pazer, H. L. (1985). "Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems." *Management Science, 31*(2), 150 - 162.
2. Ballou, D. P., & Pazer, H. L. (1995). "Designing Information Systems to Optimize the Accuracy-timeliness Tradeoff." *Information Systems Research, 6*(1), 51-72.
3. Ballou, D. P., and Tayi, Giri K. (1999). "Enhancing Data Quality in Data Warehouse Environments." *Communications of the ACM, 42*(1), 73 - 78.
4. Bontempo and Zagelow, (1998), "The IBM Data Warehouse Architecture." *Communications of the ACM, 41*(9), 38 - 48.
5. Brachman, R., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, and Simoudis, E. (1996). "Mining Business Databases." *Communications of the ACM, 39*(11), 42 - 48.
6. Celko, J., and McDonald, J. (1995). "Don't Warehouse Dirty Data." *Datamation, Oct.15*.
7. Chengalur-Smith, I., Ballou, D. P., Pazer, H. (1999). "The Impact of Data Quality Information on Decision-Making: An Exploratory Analysis." *IEEE Transactions on Knowledge and Data Engineering*.
8. Davenport, T. H. (1997). *Information Ecology*. N. Y., NY: Oxford University Press.
9. Fayyad, U., and Uthurusamy, R. (1996). "Data Mining and Knowledge Discovery in Databases." *Communications of the ACM, 39*(11), 24 - 26.
10. Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). "The KDD Process for Extracting Useful Knowledge from Volumes of Data." *Communications of the ACM, 39*(11), 27 - 34.
11. Fisher, C. W. & Kingma, B. (2001). "Criticality of Data Quality as Exemplified in Two Disasters." *Information & Management*. Elsiever, Netherlands. (Note: paper accepted; proofs completed; publication date pending.)
12. Fisher, C.W., (1999). *An Empirically Based Exploration of the Interaction of Time Constraints and Experience Levels on the Data Quality Information (DQI) Factor in Decision-Making*, University at Albany: Albany, NY. 222.
13. Golfarelli and Rizzi. (1999). "A Methodological Framework for Data Warehouse Design." *DOLAP - ACM 1999, 3 - 9*.
14. Huang, Lee, and Wang. (1999). *Quality Information and Knowledge.* Englewood Cliffs, NJ: Prentice Hall.
15. Kaplan, D., Krishnan, R., Padman, R. and Peters, J. (1998). "Assessing Data Quality in Accounting Information Systems." *Communications of the ACM, 41*(2), 72 - 78.
16. Klein, B. D., Goodhue, D.L., and Davis, G.B. (1997). "Can Humans Detect Errors in Data? Impact of Base Rates, Incentives and Goals." *MIS Quarterly, 21*(June), 169 - 194.
17. Klein, B. D. (1998). *User Detection of Errors in Data: Learning through Direct and Indirect Experience.* Paper presented at the AIS98.
18. Khalil, O. E. M., Strong, D. M., Kahn, B. K., and Pipino, L. L. (1999). "Teaching Information Quality in Information systems Undergraduate Education." *Informing Science*, 2(3). p. 53 - 59.
19. McLeod, R. J. (1995). *Management Information Systems: A Study of Computer-Based Information Systems*. 6 ed. Englewood Cliffs, NJ: Prentice Hall.
20. Morrissey, J. M. (1990). "Imprecise Information and Uncertainty in Information Systems." *ACM Transactions on Information Systems, 8*(2), 159 - 180.
21. Motro, A., and Smets, P. (Ed.). (1996). *Uncertainty Management in Information Systems: From Needs to Solutions.* Kluwer Academic Publishers.{{{missing address info (see

Prentice Hall in # 19}}
22. Neely, M. Pamela. (2000). "A Process for Auditing Source Data Quality in an integrated Data Repository: Development and Testing." *Dissertation Proposal,* University at Albany. Albany, NY.
23. Orr, K. (1998). "Data Quality and Systems Theory." *Communications of the ACM, 41*(2), 66 - 71.
24. Redman, T. C. (1995). "Improve Data Quality for Competitive Advantage." *Sloan Management Review* (Winter), 99-107.
25. Redman, T. C. (1996). *Data Quality for the Information Age*. Norwood, MA: Artech House, Inc.
26. Redman, T. C. (1998). "The Impact of Poor Data Quality on the Typical Enterprise." *Communications of the ACM, 41*(2), 79 - 82.
27. Sanbonmatsu, D., M., Kardes, Frank R., and Herr, Paul M. (1992). "The Role of Prior Knowledge and Missing Information in Multiattribute Evaluation." *Organizational Behavior and Human Decision Processes, 51*(1), 76 - 91.
28. Simpson, C., and Prusak, Laurence. (1995). "Troubles with Information Overload—Moving from Quantity to Quality in Information Provision." *International Journal of Information Management, 15*(6), 413 - 425.
29. Strong, D. M., Lee, Y. W., and Wang, R. Y. (1997). "Data Quality in Context." *Communications of the ACM, 40*(5), 103 - 110.
30. Tayi, G., and Ballou, D. P. (1998). "Examining Data Quality." *Communications of the ACM, 41*(2), 54 - 57.
31. Wand, Y., and Wang, Richard Y. (1996). "Anchoring Data Quality Dimensions in Ontological Foundations." *Communications of the ACM, 39*(11), 86 - 95.
32. Wang, R. Y., Storey, Veda C., and Firth, Christopher P. (1995). "A Framework for Analysis of Data Quality Research." *IEEE Transactions on Knowledge and Data Engineering,, 7*(4), 623 - 639.
33. Wang, R. Y., & Strong, D. (1996). "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems, 12*(4), 5 - 34.
34. Wang, R. P. (1998). "A Product Perspective on Total Data Quality Management." *Communications of the ACM, 41*(2), 58 - 65.
35. Watson, R. T., Pitt, Leyland F., Kavan, C. Bruce. (1998). "Measuring Information Systems Service Quality: Lessons From Two Longitudinal Case Studies." *MIS Quarterly, 22*(1), 61
36. Public Law 100 - 107. (1987). The Malcolm Baldridge National Quality Improvement Act of 1987. Application Guidelines 1988.

## APPENDIX A

### Apartment Selection Task

**Number: _____**
**Task**

Your job requires you to move to a new city. You have a friend who lives there and you request her help in locating an apartment. You provide a list of criteria that are important to you and you have weights in mind for each criteria, which reflect their relative importance to you. Your friend gathers information about four potential apartment complexes and passes it along to you. She scores each apartment complex on each factor on a 50-point scale, such that a higher number is always more desirable For example, a rating of 40 for rent expense is more desirable than a rating of 30.

Your objective is to choose the complex that overall performs the best. However, you realize that the data she obtained may not be completely accurate. For instance, she estimated commuting time by looking at the map. Also, the complex managers indicated that the rent quoted could increase at any time.

You decide to incorporate this uncertainty into your decision-making process by using a reliability measure where a score of 100 indicates perfectly reliable data and 0 scores completely unreliable data. The following table displays the reliability of the information provided about each criterion and the weights you assigned to each criterion.

| *Criterion* | *Reliability* | *Weight* |
|---|---|---|
| Parking facilities | 57 | 1 |
| Commuting time to work | 23 | 2.5 |
| Floor space | 76 | 2 |
| Number of bedrooms | 68 | 1.5 |
| Rent expense | 44 | 3 |

After multiplying the ratings by the weights for each criterion, you obtained the following results. The weighted scores are shown on the next page (for example, a score of 70 for commuting time is the result of multiplying its rating of 28 by its weight of 2.5).

Given the weighted scores, the objective is to choose the apartment complex that overall is the best (has the highest overall sum). Rank the apartment complexes in order of preference with 1 corresponding to the complex you would most prefer and 4 to the one you would least prefer. (Use all the information given to break any ties.) Recall that the reliability refers to the data and not to the weights. Next to each apartment complex write its rank, along with a brief explanation of how you arrived at the rank.

*A*

| *Criterion* | *Reliability* | *Rating* | *Weight* | *Weighted scores* |
|---|---|---|---|---|
| Parking facilities | 57 | 22 | 1 | 22 |
| Commuting time | 23 | 28 | 2.5 | 70 |
| Floor space | 76 | 20 | 2 | 40 |
| # of bedrooms | 68 | 32 | 1.5 | 48 |
| Rent expense | 44 | 40 | 3 | 120 |

***Rank*** **= ___**
*Explanation*:

*B*

| *Criterion* | *Reliability* | *Rating* | *Weight* | *Weighted scores* |
|---|---|---|---|---|
| Parking facilities | 57 | 25 | 1 | 25 |
| Commuting time | 23 | 32 | 2.5 | 80 |
| Floor space | 76 | 31 | 2 | 62 |
| # of bedrooms | 68 | 36 | 1.5 | 54 |
| Rent expense | 44 | 36 | 3 | 108 |

***Rank*** **= ___**
*Explanation*:

*C*

| *Criterion* | *Reliability* | *Rating* | *Weight* | *Weighted scores* |
|---|---|---|---|---|
| Parking facilities | 57 | 28 | 1 | 28 |
| Commuting time | 23 | 27 | 2.5 | 67.5 |
| Floor space | 76 | 33 | 2 | 66 |
| # of bedrooms | 68 | 29 | 1.5 | 43.5 |
| Rent expense | 44 | 26 | 3 | 78 |

***Rank*** **= ___**
*Explanation*:

*D*

| *Criterion* | *Reliability* | *Rating* | *Weight* | *Weighted scores* |
|---|---|---|---|---|
| Parking facilities | 57 | 27 | 1 | 27 |
| Commuting time | 23 | 25 | 2.5 | 62.5 |
| Floor space | 76 | 35 | 2 | 70 |
| # of bedrooms | 68 | 38 | 1.5 | 57 |
| Rent expense | 44 | 37 | 3 | 111 |

***Rank*** **= ___**
*Explanation*:

## APPENDIX B

**Assignment: Information Quality Assessment**

**Objective:** Perform an assessment of Information Quality in certain systems at Marist College.

**Systems:** Various Business and Registration Systems.

**Personnel:** Management and staff from the Business Office, Registration Office and the entire Information Technology department.

**General Statement:**

Study and apply IQ Assessment concepts; key references include:
Text: *Quality Information and Knowledge* (Huang, Lee and Wang, 1999); and
Article: "Data Quality in Context" (Strong, Lee, & Wang, 1996).

**Teams:** You will work in three teams of three or four members. However, you will all need all of the data. I recommend that you divide the data entry and then simply merge the data into one large file and make copies of the file for each team to use.

**Steps:**

1. Prepare and administer IQA Survey.
2. Design Excel or Access DB to store the data. You must agree on the database design and I have supplied a suggested one. Feel free to modify it, but to arrive at one complete report you will need one database.
3. Populate your data storage with the actual data results.
4. Design and apply statistical evaluations of the data.
   For example (but not limited to):
   - Averages, ranges etc for entire population on the 16 DQ dimensions
   - Rank the dimensions
   - Averages, ranges by subset of the population on the 16 DQ dimensions where subsets are organized by system and by type of respondent:
     *System:* Registration, Billing, etc.
     *Respondent:* Consumers, Custodians, Providers, Managers, etc.
     Note: An analysis of the data may lead to different groupings, etc.
   - Rank the dimensions by subset(s)
   - Statistically compare various subsets (e.g., t-tests)
   - Use correlations to determine strength of relationships
   - Perform IQ Context Assessment (Sec 3 of IQA Survey)
5. Develop graphs to illustrate your findings.
6. Develop recommendations where possible from IQA and IQ context.
7. Reach conclusions and prepare professional write-up of those conclusions.
8. Prepare presentations of conclusions.
9. Give presentations to the "stakeholder" groups (or management).