

A College Course:

Data Quality in Information Systems

Key Topics from Research

Measurement

Impacts

TQM

Data Entry

Policies

Error Detection

Dimensions

Change Processes

User Requirements

Information Overload

DQ Audits

Statistics

Data Mining

Data Warehouse

Analytic Models

Relational Algebra

DQ cost/benefit

Software tools

Three types of Capabilities

Interpretative—Self Reflective & social contexts

Change Mgmt, Impacts, Cost/Benefit, Error Detection

Adaptive—Significant interchange with environment

Measurements, TQM, Data entry improvement, User requirements

Technical—Mechanical Predetermined behaviors

Data Mining, Analytic Models, Data Warehouse, Relational Algebra, Statistics

Which Skills are most important?

- ∞ Interpretative—Self Reflective & social contexts
Professor, executives & managers rated these skills as most important.
- ∞ Adaptive—Significant interchange with environment
Consultants, Project Managers & analysts rated these skills as most important.
- ∞ Technical—Mechanical / Predetermined behaviors
All rated technical skills as lowest importance

Importance of Skill Type varies by type of job.

Undergraduate curricula may focus most on adaptive and technical to identify user needs and measure user satisfaction and data quality.

Graduate Programs and Executive seminars may focus mostly on Interpretative so as to assess organizational implications and policies.

Course Objectives

- Understand DQ/IQ impacts and implications of poor data quality in technological projects,
- Understand TQM/TDQM, Information Products, Data quality dimensions & Process Control concepts,
- Recognize patterns of data and design deficiencies,
- Suggest DQ and IQ improvement plans,
- ROI / Business benefits of DQ Improvement plans for databases / warehouses,
- Role and importance of DQ in Decision Making.

Course Approach

REGULAR CLASSES:

- Study text / journal articles prior to classes
- Discuss w/ leading questions and etc.

FOUR PROJECTS:

- Total Quality Management (TQM),
- Use of DQI,
- Data Warehouse cleaning, and
- Information Quality Assessments.

Information Quality Assessment

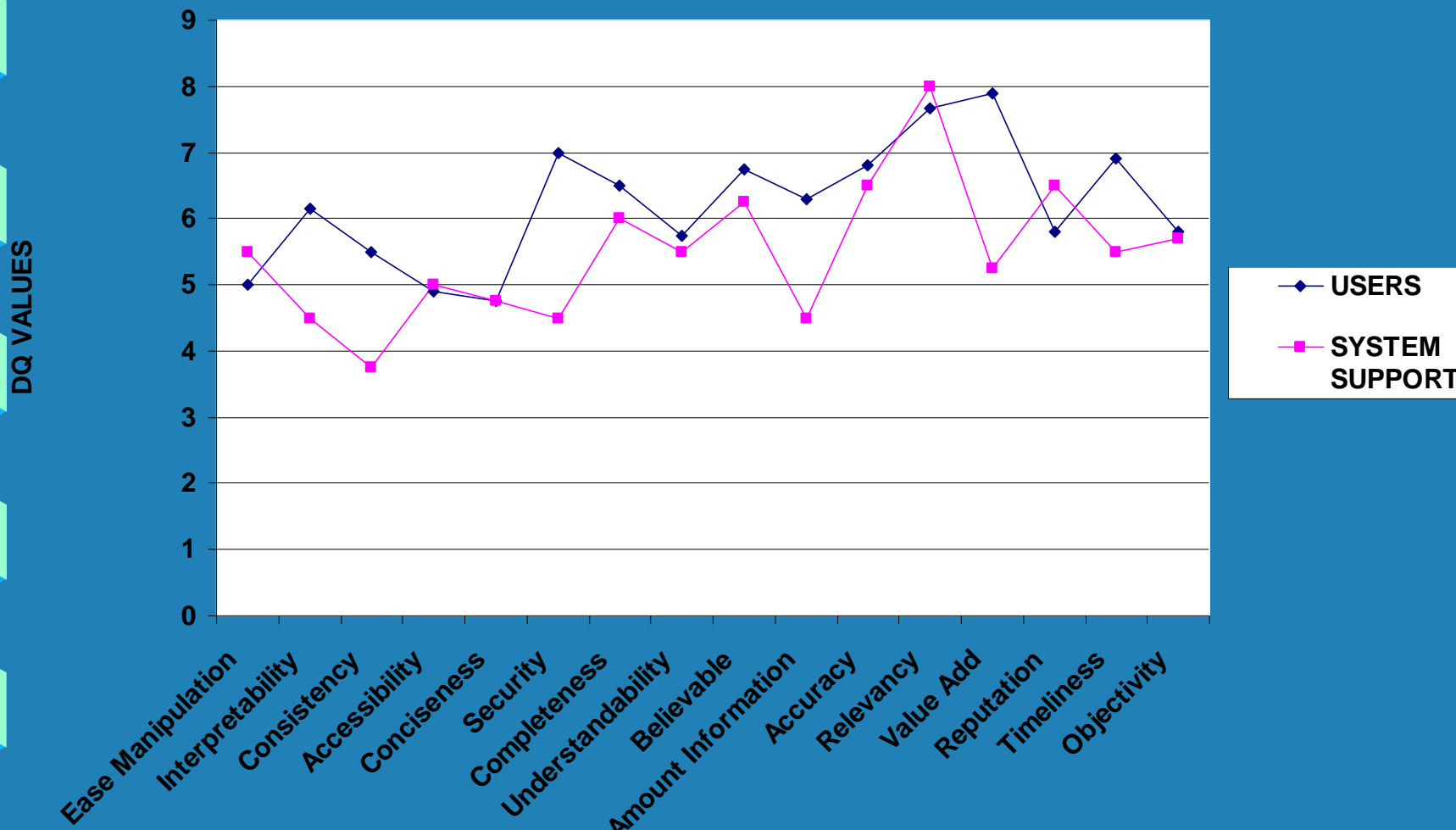
Objective: Perform an assessment of Information Quality in certain areas of XYZ Organization.

Systems: Various IS Applications / Systems.

Personnel: Management and staff from the User areas and the I T department.

ABC Databases: Sys Support v. Users

REGISTRATION DATA

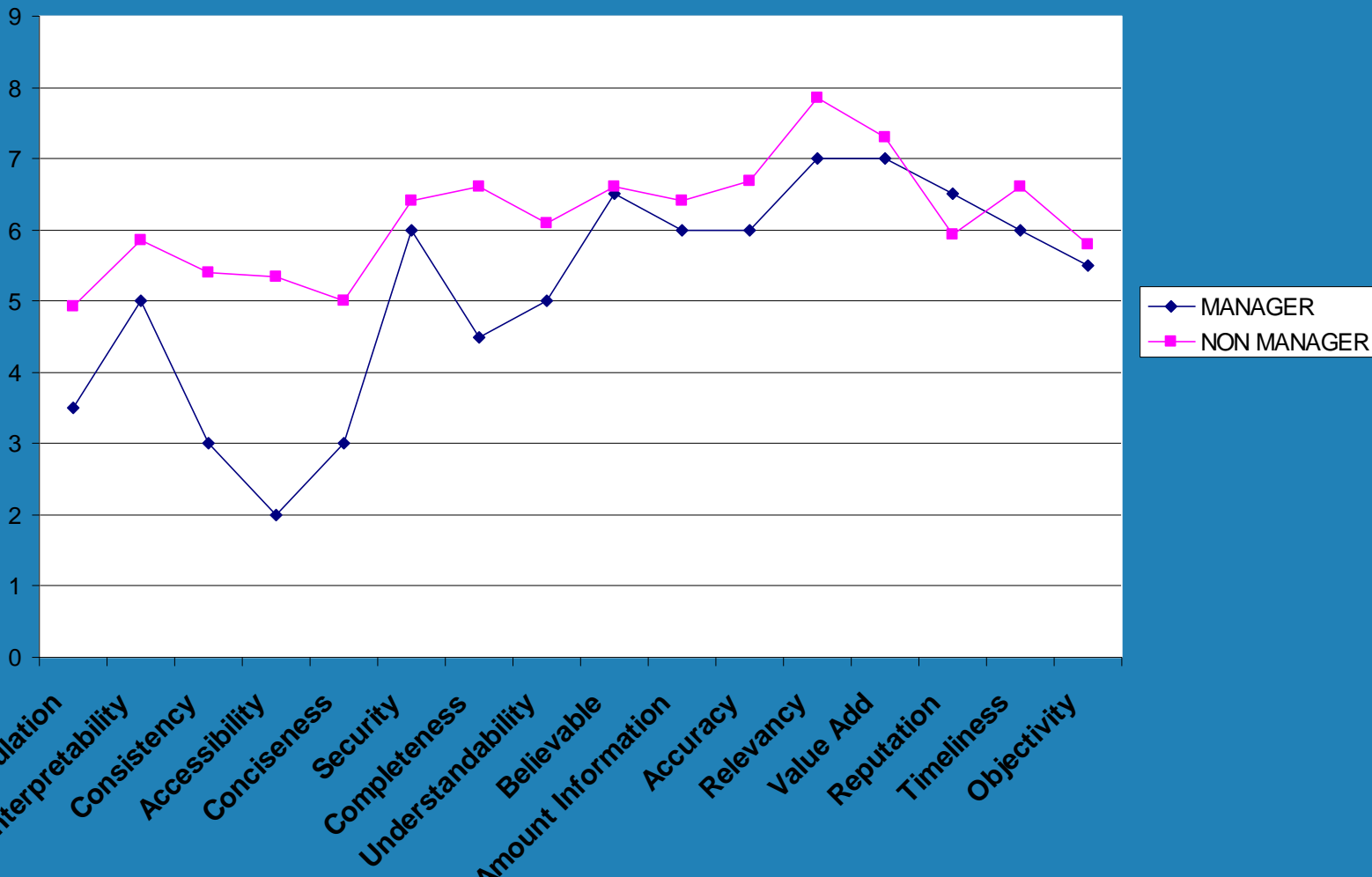


DQ VALUES

User Data: mgr v NonMgr

REGISTRATION DATA

DQ VALUES



Marist College Programs
(Related to IS, MIS, IT, TM & IQ)

Master of Science in Technology Management

Interdisciplinary program offered by School of Computer Science and Math with the School of Management. A course in information quality is required: MSTM730 *Data and Information Quality for the Information Executive*.

Master of Science in Information Systems

This program is a leading edge IS program with two tracks and is offered by the School of Computer Science and Mathematics. MSIS557 *Data Quality in Information Systems* is a very popular elective for both tracks.

Bachelor's Degree in Information and Technology Systems (ITS).

This program has 2 tracks: IT and IS. The ITS degree is new and begins with a freshman class in the fall 2007 semester. The first five semesters are identical providing a broad base in programming, analysis, design, database, data communication, web programming and networking. The IS track focuses on Information Policy, Information Quality, Decision making, Data Mining and Business Intelligence.

The IS Track requires students to take ITS428 *Data Quality in Information Systems*.

Our BS in Information Systems degree is winding down now and is merged with IT into the new ITS degree. IS428 was an elective in that IS program since 2002.

Typical Questions (collected by Liz Pierce)

Question: Why a specialized course or program in information quality? Why not simply include data quality in existing courses, e.g. testing in programming courses, db normalization and dictionaries in data management courses, inspections, systems, and acceptance testing in systems analysis & design courses?

Comments: The reason I posed this question is that it was the most common question that my peers and colleagues at Marist College asked when I brought forward the idea of teaching a course about data quality. Note that our programs above are in the School of Computer Science and Mathematics. The CS people have been teaching programming and related topics for years and may have felt slightly threatened by the new course or simply did not see the need for anything above and beyond the testing, normalization, referential integrity and so forth that were already being taught.

The answer comes in a few main areas:

- a broad description of data and information quality,
- discussion of status of databases,
- impacts of poor quality,
- growth of multi-user databases,
- downsizing,

- ubiquitous database development by users,
- increased need for integration and sharing,
- and global issues.

Question: If indeed a specialized course or program is needed, where is it best housed within the College or University structure? (e.g. CS/IT within an Engineering/Technical School, MIS within Business School, etc.).

Comments: There is no right or wrong answer. The important factor is to be sure that the students receive a fair exposure to the most important Knowledge, Skills and Abilities (KSAs). A good place to start is to determine what KSAs need to be covered by reading recent papers in the literature such as the Chung, Fisher and Wang articles that describe a data quality literature review and a survey conducted among leading information quality researchers at the International Conference on Information Quality [1, 2]. Then match up the KSAs to the rest of the program being offered. For example, a MIS or IS degree may have several more courses on systems analysis than a computer science program or information technology program. Similarly, a Business major may receive more education on controlling the processes of production than either CS or IS does. However the Business major may not receive in-depth training on normalization and testing as compared to the more technical degrees.

Question: What do we see as the viable forms of IQ education? (One course, or if more than once course should we be focusing our attention on a track (say, three courses), a certificate (say, 5 courses), a B.S. degree, a M.S. degree, or a Ph.D. degree).

Question: What background (i.e. prerequisites) should students have to pursue the various forms of IQ education?

Comments: This is similar to my comment to the question as to within what school should the program be housed. See the articles that discuss the important skills that matter and redefining the scope of data quality [1, 2].

Question: Part A) How much of the education should be hands-on? How much theory?

Part B) What types of real customer/client interactions can be brought into the classroom?

Comments: Part A) My thoughts here are based on experience and knowledge of pedagogy. First, I can barely remember any lectures that I attended in college. However, I can remember many term papers and many projects that I did. Students today may not remember what their teacher said in a class two days ago but will remember much of what they themselves, e.g. the students, said during participation opportunities. So give as many participation opportunities as possible. Simultaneously they should not be asked to work in a vacuum and not left to reach wrong conclusions. They do need theoretical grounding.

Ideally, a fine blend of hands-on and theory should thoroughly complement each other motivating an ultimate learning experience. For example, to get the most out of a

theoretical lecture about Codd's integrity rules it should be followed up with an assignment using some tool such as CRG's *Integrity Analyzer*[®] (IA) [3]. A requirement to analyze the findings of an Integrity Analyzer study in light of Codd's rules would be extremely valuable.

Part B) One of the most exciting activities that the students in an introduction to data quality class can do is to perform an *Information Quality Assessment*[®] (IQA) [4] at a client site. I have taken my classes to a customer every year for 6 years to perform an IQA. We have been to IBM twice, Office for the Aging twice, United Way and Marist College. The clients have always been impressed with how much we can learn about their site and issues within just a few days. In turn, the students become, if they weren't already, extremely enthusiastic about what they have learned and the interest generated by the client! Example assignments can be found in the recent book, *Introduction to Information Quality* [5].

References

1. Chung, W.Y., C.W. Fisher, and R. Wang. *What Skills Matter in Data Quality*. in *The Seventh International Conference on Information Quality*. 2002. Cambridge, MA: MIT TDQM Program.
2. Chung, W., C. Fisher, and R.Y. Wang, *Redefining the Scope and Focus of Information-Quality Work*, in *Information Quality*, R.Y. Wang, et al., Editors. 2005, M. E. Sharpe: Armonk, NY. p. 265.
3. CRG, *Integrity Analyzer: A Software Tool for TDQM*. 1997, MIT: Cambridge, MA.
4. CRG, *Information Quality Assessment Survey: Administrators Guide*. 1997, MIT: Cambridge, MA.
5. Fisher, C., E. Lauria, I. Chengalur-Smith, and R. Wang, *Introduction to Information Quality*. 2006, Cambridge, MA: MITIQ Press. 206.