2003

Measuring the Quality of Data Models: An Empirical Evaluation of the Use of Quality Metrics in Practice

Daniel L. Moody

Charles University, Prague, dmoody@infotech.monash.edu.au

Measuring the Quality of Data Models: An Empirical Evaluation of the Use of Quality Metrics in Practice

Daniel L. Moody
Department of Software Engineering
Charles University
Prague, Czech Republic
moody@ksint.ms.mff.cuni.cz
and
School of Business Systems,
Monash University
Melbourne, Australia
dmoody@infotech.monash.edu.au

Abstract

This paper describes the empirical evaluation of a set of proposed metrics for evaluating the quality of data models. A total of twenty nine candidate metrics were originally proposed, each of which measured a different aspect of quality of a data model. Action research was used to evaluate the usefulness of the metrics in five application development projects in two private sector organisations. Of the metrics originally proposed, only three "survived" the empirical validation process, and two new metrics were discovered. The result was a set of five metrics which participants felt were manageable to apply in practice. An unexpected finding was that subjective ratings of quality and qualitative descriptions of quality issues were perceived to be much more useful than the metrics. While the idea of using metrics to quantify the quality of data models seems good in theory, the results of this study seem to indicate that it is not quite so useful in practice. The conclusion is that using a combination of "hard" and "soft" information (metrics, subjective ratings, qualitative description of issues) provides the most effective solution to the problem of evaluating the quality of data models, and that moves towards increased quantification may be counterproductive.

Keywords

Software quality, requirements analysis, metrics, data model, Entity Relationship Model, action research

1. Introduction

1.1 The Importance of Data Model Quality

The choice of an appropriate representation of data is one of the most crucial tasks in the entire systems development process. Although data modelling represents only a small proportion of the total systems development effort, its impact on the final result is probably greater than any other phase (Witt & Simsion, 2000). The data model is a major determinant of system development costs (ASMA, 1996), system flexibility (Gartner Research, 1992), integration with other systems (Moody & Simsion, 1995) and the ability of the system to meet user requirements (Batini, Ceri & Navathe, 1992). The combination of low cost and

high impact suggests that data modelling represents a priority area for improving the quality of application development (Moody & Shanks, 2003).

Currently however, there are few guidelines for evaluating the quality of data models, and little agreement even among experts as to what makes a "good" model. However for data modelling to progress from a "craft" to an engineering discipline, formal quality criteria and metrics need to be explicitly defined (Wand & Weber, 1990; Poels, Nelson, Genero & Piattini, 2002).

1.2 The Need for Measurement

Evaluating the quality of data models is a discipline which is only just beginning to emerge. Quantitative measurement of quality is almost non-existent (Poels et al, 2002). A number of frameworks for evaluating the quality of data models have been proposed in the literature (Roman, 1985; Mayer, 1989; von Halle, 1991; Batini et al, 1992; Levitin & Redman, 1994; Lindland, Sindre & Sølvberg, 1994; Moody & Shanks, 1994; Simsion, 1994; Kesh, 1995; Krogstie, Lindland & Sindre, 1995; Moody & Shanks, 1998b). A major limitation of such frameworks is that while they all define criteria for evaluating the quality of data models, these must be interpreted and applied in a subjective way. In general, defining quality criteria is not enough to ensure quality in practice, because people will have different interpretations of what these criteria mean. Empirical studies have shown problems applying such frameworks reliably in practice (Moody & Shanks, 2003).

According to the Total Quality Management (TQM) literature, measurable criteria for assessing quality are necessary to avoid "arguments of style" (Deming, 1986). Measurable quality criteria are also essential for implementing the concept of statistical process control. A clear objective for data model quality research should therefore be to replace intuitive notions of model "quality" with formal, quantitative measures to reduce subjectivity and bias in the evaluation process. However developing reliable and objective measures of quality in software development is a difficult task. As Van Vliet (1993) says:

> "The various factors that relate to software quality are hard to define. It is even harder to measure them quantitatively. There are very few quality factors or criteria for which sufficiently sound numeric measures exist."

Two of the frameworks proposed (Kesh, 1995; Moody & Shanks, 1998b) propose metrics for evaluating the quality of data models but no empirical validation has been conducted for these metrics.

1.3 The Need for Empirical Validation

It is essential for IS design methods to be tested in practice—ultimately, the scientific merit of any method is an empirical rather than a theoretical question (Rescher, 1973; Ivari, 1986). However real world validation of methods in IS design research is very poorly done. IS design research tends to emphasise the development of new methods while addressing the use and evaluation of methods in only a limited fashion (Bubenko, 1986; Curtis, 1986; Fitzgerald, 1991; Westrup, 1993; Wynekoop & Russo, 1997; Moody & Shanks, 1998a). Wynekoop and Russo (1997) conducted a review of IS design research published in the leading IS journals over the past three decades. The results of the analysis showed a heavy reliance on normative research, largely focusing on the development of new methods or modifications to existing methods. They concluded that there was a "lack of serious

empirical research into the efficacy of methods in practice" and a "need for validation of methods in organisational contexts using real practitioners".

1.4 Objectives of this Research

A previous paper (Moody, 1998) defined a comprehensive set of metrics for evaluating the quality of data models. This paper evaluates the utility of these metrics in evaluating the quality of data models in practice. So far there has been only one previous empirical study of metrics for data model quality (Genero, Poels & Piattini, 2001). This was a laboratory experiment conducted using university students, which evaluated metrics for one particular aspect of quality (structural complexity). The current paper addresses the broader issue of the practical utility of metrics in data modelling practice.

2. Proposed Metrics

This section describes the metrics proposed for evaluating the quality of data models. This defines the *a priori* theory tested by this research.

2.1 Data Model Quality Framework

The metrics proposed were developed based on the framework for data model evaluation and improvement proposed by Moody and Shanks (1994; 1998b). The framework consists of five primary constructs and is summarised by the Entity Relationship model shown in Figure 1:

- Stakeholders are people involved in developing or using the data model, and therefore have an interest in its quality.
- Quality factors are the characteristics of a data model that contribute to its quality.
- Quality metrics define ways of measuring particular quality factors. There may be multiple measures for each quality factor.
- Weightings define the relative importance of different quality factors and are used to help make trade-offs between them.
- Improvement strategies are techniques for improving the quality of data models with respect to one or more quality factors.
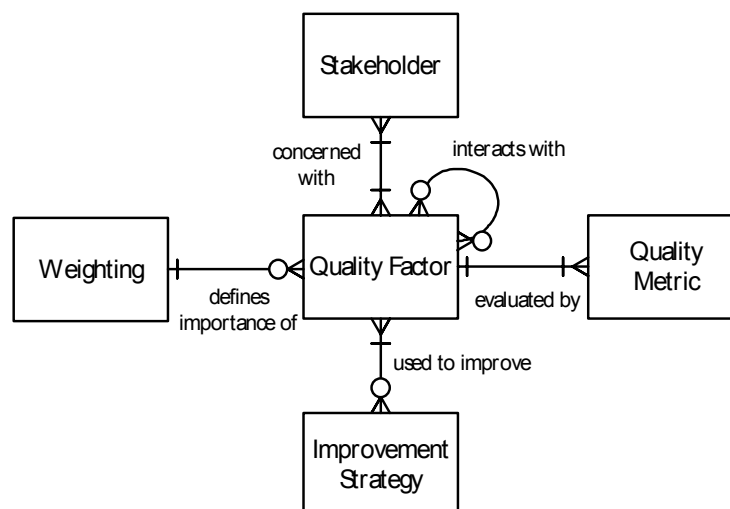

Figure 1. Data Model Quality Evaluation Framework

2.2 Quality Factors

The proposed quality factors are summarised in the "fishbone" diagram in Figure 2. Together the set of quality factors incorporate the needs of all stakeholders, and define a complete picture of data model quality. The quality factors defined may be used as criteria for evaluating the quality of individual data models and comparing alternative representations.
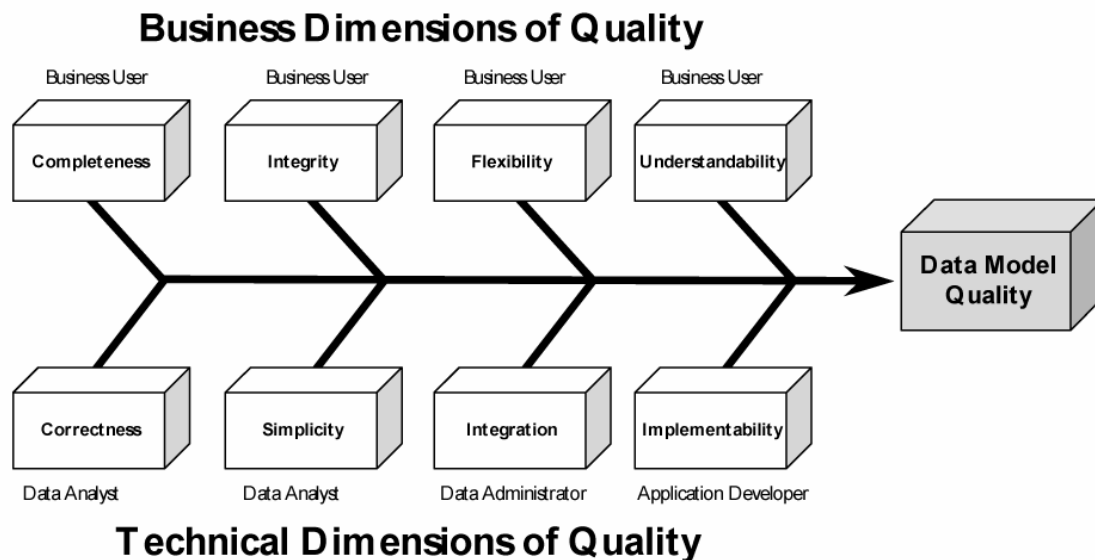

Figure 2. Data Model Quality Factors

The definitions of the quality factors are:

1. Correctness was defined as whether the model conforms to the rules of the data modelling technique (i.e. whether it is a *valid* data model). This includes diagramming conventions, naming rules, definition rules, rules of composition and normalisation.
2. Completeness refers to whether the data model contains all information required to support the required functionality of the system.
3. Integrity is defined as whether the data model defines all business rules which apply to the data.
4. Simplicity means that the data model contains the minimum possible entities and relationships.
5. Flexibility is defined as the ease with which the data model can cope with business and/or regulatory change.
6. Integration is defined as the consistency of the data model with the rest of the organisation's data.
7. Understandability is defined as the ease with which the concepts and structures in the data model can be understood.
8. Implementability is defined as the ease with which the data model can be implemented within the time, budget and technology constraints of the project.

These quality factors have been validated using multiple action research studies and two laboratory experiments, which evaluated their completeness, relevance and independence (Moody & Shanks, 2003).

2.3 Metrics

A total of twenty nine candidate metrics (along with eighteen second level metrics) were proposed, each of which measures one of the quality factors defined (Moody, 1998). The metrics are not discussed in detail here, but are summarised in Table 1:

Table 1. Summary of Proposed Metrics for Data Model Quality

QUALITY FACTOR	PROPOSED METRICS
Correctness	1. Number of violations to data modelling standards 2. Number of instances of entity redundancy 3. Number of instances of relationship redundancy 4. Number of instances of attribute redundancy
Completeness	5. Number of missing requirements (Type I errors) 6. Number of superfluous requirements (Type II errors) 7. Number of inaccurately defined requirements 8. Number of inconsistencies with process model
Integrity	9. Number of missing business rules 10. Number of incorrect business rules 11. Number of business rules inconsistent with process model 12. Number of business rules redundantly defined in process model rules
Flexibility	13. Number of data model elements which are subject to change 14. Probability adjusted cost of change 15. Strategic impact of change
Understandability	16. User rating of understandability 17. User interpretation errors 18. Application developer rating of understandability 19. Subject area-entity ratio 20. Entity-attribute ratio
Simplicity	21. Number of entities (E) 22. System complexity (E+R) 23. Total complexity (aE+bR+cA)
Integration	24. Number of data conflicts with Corporate Data Model 25. Number of data conflicts with existing systems 26. Number of data items duplicated in existing systems or projects 27. Rating of ability to meet corporate needs
Implementability	28. Development cost estimate 29. Technical risk rating

3. Research Methodology

3.1 Action Research

A major barrier to the empirical validation of IS design methods is that it is very difficult to get new approaches, especially those developed in academic environments, accepted and used in practice. Practitioners who have developed familiarity and expertise with existing techniques are reluctant to adopt academic approaches that are theoretically sound but unproven in practice (Bubenko, 1986; Wynekoop & Russo, 1997; Avison, Lau, Myers & Nielsen, 1999).

Action research is an alternative social science research approach which links theory and practice to solve practical problems in the field (McKernan, 1991; Baskerville & Wood-Harper, 1996; Stringer, 1996; Baker, 1998). It has a long history of successful application in other applied disciplines, such as education, psychology and health care (Masters, 1995), and can be applied in field settings where more traditional experimental or quasi-experimental methods cannot easily be applied (Dick, 2000). One of its major advantages is that it can help to overcome the problem of persuading practitioners to adopt new techniques, and overcome the cultural divide that exists between information systems academics and practitioners (Checkland, 1991; Moody & Shanks, 1998a; Avison et al, 1999; Moody, 2000)

Action research is usually carried out in a number of discrete cycles, which function as "mini-experiments" carried out in practice. Through reflection on previous action, a theory of the form "if I do X, then Y will occur" is proposed, which is applied in practice, and then evaluated in a cyclic manner (Oosthuizen, 2000). One of the most widely used approaches to action research is that developed by Kemmis and McTaggart (1988), which emerged from the field of education. Each action research cycle consists of the following steps (Figure 3):

- Plan: Develop a plan of action to improve current practice.
- Act: The participants act together to implement the plan.
- Observe: The action is observed to collect evidence which allows thorough evaluation of outcomes. A variety of data collection methods may be used to evaluate the results of the intervention (Holter & Schwartz-Barcott, 1993; Stringer, 1996).
- Reflect: Participants reflect on what went wrong, what went right and how to improve the idea in the next cycle. Each cycle may lead to improvement of the original idea (M_1), resulting in a sequence of successively refined and improved ideas M_2, M_3...
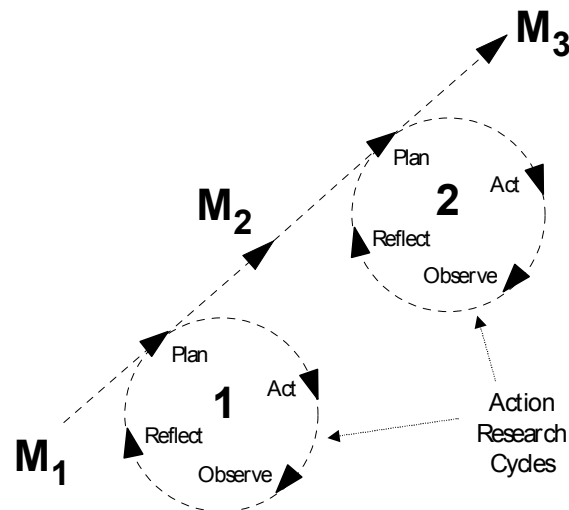
Figure 3. The Action Research Spiral

3.2 Justification for Research Method Selection

Action research is an appropriate research method to use at this stage of the research for three primary reasons:

- It is a *field based* method, which allows evaluation of the metrics "in an organisational context using real practitioners", as recommended by Wynekoop and Russo (1997). The results of a laboratory experiment would provide less convincing evidence about the utility of the metrics because of the artificiality of the laboratory situation (external validity).
- It is an *interventionist* approach, in which the researcher acts as a change agent rather than a passive observer. Traditional field based research methods like case study or surveys are passive approaches, which are only appropriate for studying pre-existing phenomena. Because the metrics represents a change to existing practices, an interventionist approach is required.
- It is a *qualitative* method, and is therefore suitable for exploratory research (Wynekoop & Russo, 1997). As the metrics have never been applied in practice, the research is in its exploratory stages. Action research is particularly suited to exploratory research, as it allows research ideas to evolve as part of an ongoing learning and reflection process (Dick, 2000).

3.3 Research Questions Addressed

The objective of this research was to evaluate the practical usefulness of the proposed metrics in supporting the task of data model quality evaluation in an application development context. The broad research questions addressed by this research are:

- How useful are the metrics in supporting the process of data model quality evaluation?
- How can the metrics be improved?

4. The Intervention (Action)

4.1 Organisational Context

The action research was conducted in five application development projects in two private sector organisations: a large financial institution and a telecommunications company. The metrics were applied as part of data model quality reviews, which were conducted at prescribed points in the requirements analysis phase. Quality reviews were conducted by the researcher in conjunction with members of the data administration/data architecture group in each organisation, who were responsible for quality assurance of data models. Data collection methods used to evaluate the metrics included interviews, observation, facilitated group discussions and direct use of the metrics (participation in the review process).

4.2 Stakeholders

Two main groups of stakeholders were identified in the action research studies:

- Reviewers: These represent *direct users* of the metrics, as they applied the metrics in conducting quality reviews. These were members of the data administration/data architecture groups. As direct users of the metrics, they were involved as participants in the research and in the interpretation of the results (evaluation of outcomes and reflection). Feedback was obtained from reviewers on the usability of the metrics and their usefulness in supporting the review process.
- Reviewees: These represent *indirect users* of the metrics, as they received the metrics information as feedback on their work. These were project analysts working on application development projects, who were responsible for developing data models. Feedback was obtained from reviewees as to the usefulness of the metrics for improving the quality of the model.

Each of these stakeholder groups is likely to have a different perspective on the utility of the metrics.

4.3 Evaluation Process

Each quality review constituted a separate action research cycle. At the end of each cycle, a facilitated group discussion was conducted to obtain perceptions of reviewers and reviewees on the usefulness of the metrics, and to reflect on how they could be improved. Given the qualitative nature of the study, we did not attempt to formally evaluate the reliability or validity of the metrics—this would be better evaluated by quantitative research. Instead, we focused on evaluating their practical usefulness, as perceived by study participants.

5. Results

Rather than reporting the results of each action research cycle, we summarise the results of all five action research studies, as the cycles were conducted in a cumulative manner.

5.1 Number of Metrics

The most common reaction by reviewers was that there were too many metrics. In both organisations, there was only a limited number of people qualified to conduct reviews, and the need to calculate the metrics added significantly to their workload. As one of the reviewers in the first action research study said:

> *"It takes almost as long to calculate the metrics as it does to review the model, which I think is overkill. I am also not entirely sure whether all of the metrics add value to the process. While I can appreciate what you are trying to do, there must be a better way…"*

The large number of metrics also resulted in information overload for the reviewees. In general, participants felt that a small but critical set of metrics was preferable to trying to measure all possible aspects of quality. This was reflected in the fact that in each action research cycle, the number of metrics was *reduced*—the final set of metrics was less than 20% of the number of metrics originally proposed. This may highlight a fundamental philosophical difference between research and practice:

- Research tends to strive for completeness and closure—in this case, to measure all possible aspects of quality.
- Practice tends to focus on what is necessary to get the job done—to measure only those aspects that are most important for improving quality—an "80/20" approach. This reflects a focus on *utility*.

The need for completeness in measuring data model quality must be balanced with the need for the measurement approach to be practical and useable.

5.2 Usage of Metrics

The most common reaction by reviewees was that they were unsure about how the metrics information should be used. In the initial formulation of the metrics, the intended usage of the information was not always clearly thought out. In trying to measure all possible aspects of quality, and to generate as many metrics as possible, the result was often "measuring what you *can* measure". As one participant said:

> *"I really can't see the point of measuring some of these things. While it might be useful for theoretical purposes, I can't see how it is going to help us develop better models, which is the only reason I am interested in doing this…"*

A useful outcome of the action research studies was to clarify *how* different metrics could be used. Three possible uses were identified for metrics:

1. To improve the quality of the model (product quality)
2. To choose between alternative models (model choice)
3. For comparison across projects and over time to improve the data modelling process (process quality).

In general, different metrics are useful for different purposes:

- For example, counting the *number of data items* missing from the model (Metric 5) does not provide useful information for improving the model. To improve the model, we need to know *which* items are missing (not just *how many*) and to incorporate them into the model. It is also not useful for model choice, as in practice, choices are generally required between models which represent the same requirements (semantically

equivalent representations). However this metric is useful for comparison over time, to identify potential problems in the data modelling process (process quality). However it would only be useful for comparison if it was standardised across projects—for example, if it was expressed as a ratio of the size of the model.

- As another example, knowing the *number of entities* in a model (Metric 21) does not provide useful information for improving the quality of the model. As an absolute value, there is no way of knowing whether complexity is "high" or "low" relative to the complexity of the problem. Nor is such a metric useful for comparison over time. However it *is* useful for choosing between alternatives (model choice): when there are two alternative representations of requirements, the simpler one should be chosen.

The identification of how metrics could be used to improve the data modelling process was an important insight from the study. Process quality has been largely ignored in research on data model quality (Poels et al, 2002; Moody & Shanks, 2003) and was not explicitly considered in generating the metrics. However according to Total Quality Management (TQM), sustainable improvements in quality can only be achieved by modifying processes to prevent defects (Deming, 1986).

5.3 Use of Metrics is a Cost-Benefit Decision

Application development projects have strictly limited timelines and budgets, and any activity which does not directly contribute to the delivery of the final system (including quality assurance activities) will be scrutinised very carefully. Quality metrics are costly to collect, and will only be used if their perceived benefits clearly outweigh the effort required to collect them. Discussions with participants during the reflect phases of the action research cycles showed that their perceptions of the usefulness of a particular metric and their consequent willingness to use it is a pragmatic decision based on :

- Cost: Perceived cost or effort of collection (reviewer's viewpoint)
- Benefit: Perceived benefits in terms of improving the quality of models (reviewer and reviewee's viewpoint)

This was used as the primary basis for determining the "validity" of a metric. Effectively this defines the "dependent variable" for the study: the criteria for determining whether each metric was practical useful or not. The concepts of "cost" and "benefit" used by participants correspond roughly to the concepts of Perceived Ease of Use and Perceived Usefulness as defined in the Technology Acceptance Model (TAM) (Davis, Bagozzi & Warshaw, 1989). These are the joint determinants of Intention to Use, which provides a theoretical explanation for participants' decisions about whether they would use a metric or not. It also reflects the principle of *methodological pragmatism*: people will only use a method if helps them to get the job done faster, more cheaply or improves the quality of the result (Rescher, 1977).

5.4 Validation of Metrics

Only three out of the twenty nine metrics proposed survived the "cost-benefit" test—these were the only ones that participants considered to be worth collecting. These were:

- *Metric 22. System Complexity (E+R):* this calculated the complexity of the model as the number of entities plus the number of relationships. This was found to be the most useful of the data model complexity metrics proposed. It was used for choosing between alternative solutions (model choice). As it is a purely objective measure and easy to

calculate (it can be automatically calculated when the model is represented in a CASE tool), it provided a good way of settling arguments about which was the best model when both met requirements. The principle adopted was "all other things being equal, choose the simpler model" (an application of Ockham's Razor).

- *Metric 26. Number of data items duplicated in existing systems or projects.* This measures the level of "waste" in the development process, and has a direct and measurable impact on the development cost of the system. This metric was found to be useful in minimising duplication in a particular data model (product quality) and also for monitoring levels of duplication over time (process quality). Use of this metric helped to expand the focus of quality reviews from individual systems (application-centred view) to the wider organisational context (enterprise view). It also led to major cost savings: in the first action research study, it reduced the size of the model by almost half and led to estimated savings in data entry and maintenance costs of over $100,000 per year. While productivity improvements can be achieved at the level of individual projects, in most large organisations, the greatest productivity gains can be achieved by reducing duplication and overlap *between* projects (Moody & Shanks, 1998b).

- *Metric 28. Development Cost Estimate.* This is a relatively time-consuming metric to calculate, but was considered worthwhile because it corresponds to the "bottom line". A frequent problem with data modelling in practice is that models are produced which are impractical to implement. Often this is only discovered when the model is handed over to the development team, after it has been signed off by users. In one of the organisations, there was considerable animosity between analysts and developers, as a result of wholesale changes to models in physical database design. Calculating this metric is the equivalent of getting a "quote" from the application developer. This was useful for making cost/quality trade-offs to ensure that the model could be realistically implemented (product quality) and choosing between alternative solutions (model choice). It also encouraged involvement of application developers in the analysis stage, which provides advantages of concurrent engineering.

Some of the metrics were considered so difficult to collect or of such marginal benefit that reviewers did not even attempt to collect them. For example, Metric 16 (User Interpretation Errors) requires end users to answer a set of comprehension questions about the model. While this might be possible to do in a laboratory situation, it would be difficult, if not impossible, to get users to do this in practice.

5.5 New Metrics Discovered

Two new metrics were discovered as a result of the action research process.

- *Metric 30. Reuse Level.* This is the inverse or "positive" of the level of duplication metric (Metric 26) and measures the number of existing data items reused as part of the new model[†]. This can be converted into a dollar figure, representing the potential development cost savings as a result of reuse. The main purpose of this metric was to encourage behavioural change on the part of application developers. Typically, resistance by application developers is one of the major barriers to implementing reuse, and changing performance measurement processes can help to encourage behavioural change (Banker & Kauffman, 1991; Moody & Simsion, 1995). In one of the

---

[†] While these metrics are the logical "inverse" of each other, they are independent metrics and cannot be derived from each other.

organisations, average reuse levels more than doubled (from 18.8% to 41.3%) in the two years following introduction of this metric.

- *Metric 31. Number of Issues by Quality Factor.* Each quality issue raised as a result of quality reviews can be classified by the quality factor it relates to. The number of issues raised and their severity by quality factor gives a "defect frequency" which can be used for comparison over time‡. This can be used to identify patterns and frequencies which can be used to rectify problems in the data modelling process (process quality). For example, if a large number of integration issues are being identified as part of quality reviews, this may indicate the need for better education in this area or improve communication between project teams.

In the case of "Completeness", Metric 31 effectively combines together all the Completeness metrics proposed into a single figure, with a consequent loss of precision in tracking the type of completeness error, but reduction in cost of collection.

5.6 Usefulness of Subjective Ratings

A surprising outcome of the study was that overall subjective ratings for each quality factor using a 1—5 scale (1 = poor; 5 = excellent) were considered by those being reviewed to be much more useful than the metrics. Ironically this is what the research was trying to improve upon. Figure 4 shows an example of such an analysis in the form of a Polar or Kiviat chart (this is an example taken from one of the action research studies).
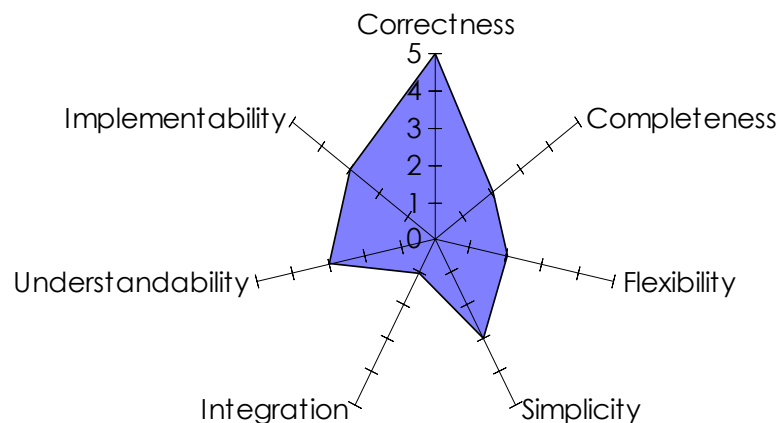

Figure 4. Subjective Ratings of Data Model Quality

One reason for this may be *conceptual manageability*: it is difficult to synthesise a large number of metrics into an overall picture of the quality of the model. The subjective ratings provide a much more *holistic* view of the quality of a model. As one of the analysts said:

> *"The nice thing about this chart is that it shows what is wrong with the model on a single page. This is a lot easier than going through all the numbers and trying to work out which need to be improved. This acts like a "report card" from the review. "*

Another reason may be that many of the metrics have little comparison value. For example, if there are 5 missing user requirements in the data model (Metric 2), is this "good" or "bad"

---

‡ This may be considered as eight metrics rather than a single metric since a separate value is calculated for each quality factor.

compared to acceptable performance? The subjective ratings shown in Figure 4 clearly indicate what the strengths and weaknesses are in the model—ratings below 3 are "bad" and ratings above 3 are "good". This provides feedback to the analyst as to where the model needs to be improved. It is also human nature to want to know how well one has done compared to the average or what is considered "acceptable". Subjective quality ratings are useful for all of the purposes identified earlier: improving the quality of the model, choosing between alternatives and comparison over time to improve processes.

5.7 Qualitative Information is the Most Useful of All

The focus of this research was to develop quantitative measures of data model quality. An unexpected finding was that qualitative information, in the form of textual descriptions of quality issues, was perceived to be the most valuable output of the review process. Table 2 shows an excerpt from a matrix used to record quality issues. Each issue is classified by quality factor and prioritised on a scale of 1 (critical), 2 (important) or 3 (desirable).

Table 2. Quality Issues Matrix

No.	Issue Description	Quality Factor	Priority	Status	Resolution
1.	Definition of Customer does not include external clients, only internal business units	Flexibility	1	Resolved	Definition of Customer expanded and subtypes introduced for internal and external clients
2.	Little information in common between subtypes of Communication Result	Correctness	3	Resolved	Subtypes remodelled as separate entities and super-type removed
3.	Staff training and skills information (e.g. language capabilities) missing	Completeness	1	Open	Will require additional analysis work
…	…	…	…	…	…

The reason why this information was perceived to be so valuable is that it provided the basis for improving the model. Each quality issue corresponds to a "quality defect" (a requirements error) which needs to be corrected. The list of issues thus provides a concrete basis for getting the model to an acceptable level of quality. Many of the metrics proposed represent defect counts of various kinds, which may be useful for process improvement but not for improving the model. As one of the participants said about Metric 9 (Number of Missing Business Rules):

> *"OK, this tells me that there are five integrity rules missing from the model. However what I really need to know is exactly <u>what</u> rules are missing so I can include them in the model—counting them doesn't provide any useful information to me on its own."*

In this case, transforming the qualitative information (what the individual defects are) into the form of a metric (*how many* defects there are) loses the most important information from the analyst's viewpoint.

6. Conclusion

This paper has conducted an empirical evaluation of a set of proposed metrics for evaluating the quality of data models. Action research was used as a way of evaluating their utility in an organisational setting using real practitioners.

6.1 Summary of Findings

The process of originally developing the quality metrics focused on generating as many metrics as possible. The action research studies were used to determine which of the metrics were useful in practice. The studies showed that many of the metrics were very time consuming to collect, and many were perceived to have marginal benefits from a project viewpoint. Only five metrics emerged from the analysis as having a "positive NPV", in terms of perceived costs of collection versus perceived quality benefits. These were:

1. *System complexity (E+R):* this was found to be useful for choosing between alternative models.
2. *Number of data items duplicated in existing systems:* this was found to be useful for reducing development costs and improving the data modelling process.
3. *Development cost estimate:* this was found to be useful for making cost/quality trade-offs, choosing between alternatives and getting application developers involved early.
4. *Reuse percentage:* this was found to be useful for calculating cost savings through sharing of data and encouraging culture change in the organisation.
5. *Number of defects by quality factor:* this was found to be useful for identifying potential problems in the data modelling process and making process improvements.

The first three metrics were part of the original set of proposed metrics and the other two emerged as a result of the action research process.

6.2 A Success or Failure?

On the face of it, the action research studies might be considered to be a failure, as only a small proportion of the metrics (3 out of the 29 originally proposed) were found to be useful. However this serves to underline the critical importance of empirical testing of ideas in practice. Ideas that are good in theory don't always work in practice. In many ways, an action research study is more of a failure if there is no change to the idea proposed—that is, if the result merely confirms the researcher's ideas. A major objective of action research is to learn from the experience and improve the idea.

6.3 Implications for Data Modelling Practice

Surprisingly, both subjective ratings and qualitative information were perceived to be more useful than the metrics for evaluating and improving the quality of data models. This may reflect the fact that requirements analysis is more of an "art" than a science – for example, in physical database design, the quality of the design can be measured quantitatively by things such as storage space, speed of access, CPU requirements etc. It is much more difficult to measure the quality of a logical specification than a finished product (e.g. a working system). The conclusion from this research for data modelling practice is that a combination of "hard" and "soft" information provides the best solution to the problem of evaluating the quality of data models:

- Hard (quantitative measures): The set of five metrics described above is sufficient for practical purposes, and satisfies the need for a small set of critical measures as expressed by participants.
- Medium (subjective ratings): Expert ratings of overall quality by quality factor. These provide useful feedback to the analyst as to the strengths and weaknesses of the model and where it needs to be improved.
- Soft (textual information): Qualitative description of issues. These provide a concrete basis for improving the quality of the model.

References

1.   ASMA (1996): ASMA Project Database Release 7.0, ASMA (Australian Software Metrics Association), P.O. Box 1287, Box Hill, Victoria, Australia, 3128, November.

2.   AVISON, D., LAU, F., MYERS, M. and NIELSEN, P.A. (1999): "Action Research", Communications of the ACM, 42, 1, January.

3.   BAKER, T.L. (1998): Doing Social Research, McGraw-Hill.

4.   BANKER, R.D. and KAUFFMAN, R.J. (1991): "Reuse and Productivity in Integrated Computer Aided Software Engineering: An Empirical Study", MIS Quarterly, 15, 3, pp. 375-401, September.

5.   BASKERVILLE, R.L. and WOOD-HARPER, T. (1996): "A Critical Perspective on Action Research as a Method for Information Systems Research", Journal of Information Technology, 3, 11, pp. 235-246.

6.   BATINI, C., CERI, S. and NAVATHE, S.B. (1992): Conceptual Database Design: An Entity Relationship Approach, Benjamin Cummings, Redwood City, California.

7.   BUBENKO, J.A. (1986): "Information Systems Methodologies - A Research View". Information Systems Design Methodologies: Improving The Practice, T.W. Olle, H.G. Sol and A.A. Verrijn-Stuart, (Eds.), North-Holland.

8.   CHECKLAND, P.B. (1991): "From Framework Through Experience To Learning: The Essential Nature of Action Research", Information Systems Research: Contemporary Approaches And Emergent Traditions, H.E. Nissen, H.K. Klein and R. Hirschheim, (Eds.), North-Holland,

9.   CHOW, S.L. (1988): "Significance Test or Effect Size", Psychological Bulletin, 103, 1.

10.   CURTIS, B. (1986): "By The Way, Did Anyone Study Any Real Programmers?". Empirical Studies of Programmers, E. Soloway and S. Iyengar, (Eds.), Norward, N.J., Ablex.

11.   DAVIS, F.D., BAGOZZI, R.P. and WARSHAW, P.R. (1989): "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models", Management Science, 35, 8, pp. 982-1003.

12.   DEMING, W.E. (1986): Out of the Crisis, MIT Center for Advanced Engineering, Cambridge, MA.

13.   DICK, B. (2000): "A Beginner's Guide to Action Research [On line]". Action Research and Evaluation On Line (AREOL), B. Dick, R. Passfield and P. Wildman, (Eds.), Available on-line at http:\\www.scu.edu.au/schools/gcm/ar/arp/guide.html.

14.   FITZGERALD, G. (1991): "Validating New Information Systems Techniques: A Retrospective Analysis", Information Systems Research: Contemporary Approaches And Emergent Traditions, H.E. Nissen, H.K. Klein and R. Hirschheim, (Eds.), North-Holland,

15.   GARTNER RESEARCH (1992): "Sometimes You Gotta Break the Rules", Gartner Group Strategic Management Series Key Issues, November 23.

16.   GENERO, M., POELS, G. and PIATTINI, M. (2001): "Defining and Validating Measures for Conceptual Data Model Quality", 13th International Conference on Advanced Information Systems Engineering (CAiSE 2001), K.R. Dittrich, A. Geppert and M.C. Norrie, (Eds.), Interlaken, Switzerland, Springer, Lecture Notes in Computer Science 2068, June 4-8.

17.   HOLTER, I.M. and SCHWARTZ-BARCOTT, D. (1993): "Action Research: What is it, how has it been used and how can it be used in nursing?", Journal of Advanced Nursing, 18.

18.   IVARI, J. (1986): "Dimensions Of Information Systems Design: A Framework For A Long Range Research Program", Information Systems Journal, June.

19.   KEMMIS, S. and MCTAGGART, R.E. (1988): The Action Research Planner, Deakin University, Melbourne, Australia.

20.   KESH, S. (1995): "Evaluating The Quality Of Entity Relationship Models", Information And Software Technology, 37, 12.

21.   KROGSTIE, J., LINDLAND, O.I. and SINDRE, G. (1995): "Towards a Deeper Understanding of Quality in Requirements Engineering", Proceedings of the 7th International Conference on Advanced Information Systems Engineering (CAISE), Jyvaskyla, Finland, June.

22.   LEVITIN, A. and REDMAN, T. (1994): "Quality Dimensions of a Conceptual View", Information Processing and Management, 31, 1.

23.   LINDLAND, O.I., SINDRE, G. and SØLVBERG, A. (1994): "Understanding Quality In Conceptual Modelling", IEEE Software, 11, 2, pp. 42-49, March.

24.   MASTERS, J. (1995): "The History of Action Research", Action Research Electronic Reader (on-line), I. Hughes, (Ed.), http://www.behs.cchs.usyd.edu.au/arow/Reader/rmasters.htm, The University of Sydney,

25.   MAYER, R.E. (1989): "Models for Understanding", Review of Educational Research, Spring.

26.   MCKERNAN, J. (1991): Curriculum Action Research: A Handbook of Methods and Resources for the Reflective Practitioner, Kogan Page, London.

27.   MOODY, D.L. (1998): "Metrics for Evaluating the Quality of Entity Relationship Models", Proceedings of the Seventeenth International Conference on Conceptual Modelling (ER '98), T.W. Ling, S. Ram and M.L. Lee, (Eds.), Singapore, Elsevier Lecture Notes in Computer Science, November 16-19.

28.   MOODY, D.L. (2000): "Building Links Between IS Research and Professional Practice: Improving the Relevance and Impact of IS Research", International Conference on Information Systems (ICIS'00), R.A. Weber and B. Glasson, (Eds.), Brisbane, Australia, December 11-13.

29.   MOODY, D.L. and SHANKS, G.G. (1994): "What Makes A Good Data Model? Evaluating the Quality of Entity Relationship Models", Proceedings of the Thirteenth International Conference on the Entity Relationship Approach, P. Loucopolous, (Ed.), Manchester, England, December 14-17.

30.   MOODY, D.L. and SHANKS, G.G. (1998a): "Evaluating and Improving the Quality of Entity Relationship Models: An Action Research Programme", Australian Computer Journal, November.

31.   MOODY, D.L. and SHANKS, G.G. (1998b): "What Makes A Good Data Model? A Framework For Evaluating And Improving The Quality Of Entity Relationship Models", Australian Computer Journal, August.

32.   MOODY, D.L. and SHANKS, G.G. (2003): "Improving the Quality of Data Models: Empirical Validation of a Quality Management Framework", International Journal of Information Systems, forthcoming.

33.   MOODY, D.L. and SIMSION, G.C. (1995): "Justifying Investment In Information Resource Management", Australian Journal Of Information Systems, September.

34.   OOSTHUIZEN, M. (2000): "Action Research". Research Methods for Students and Professionals: Information Management and Systems, K. Williamson, (Ed.), Charles Sturt University, Centre for Information Studies.

35.   POELS, G., NELSON, J., GENERO, M. and PIATTINI, M. (2002): "Quality in Conceptual Modeling - New Research Directions", International Workshop on Conceptual Modeling Quality (IWCMQ'02), G. Poels, J. Nelson, M. Genero and M. Piattini, (Eds.), Tampere, Finland, Springer-Verlag, October 11.

36.   RESCHER, N. (1973): The Primacy of Practice, Basil Blackwell, Oxford.

37.   RESCHER, N. (1977): Methodological Pragmatism: Systems-Theoretic Approach to the Theory of Knowledge, Basil Blackwell, Oxford.

38.   ROMAN, G. (1985): "A Taxonomy of Current Issues in Requirements Engineering", IEEE Computer, April.

39.   SIMSION, G.C. (1994): "Data Management in the 90's (keynote address)", DAMA International Symposium, New York, May.

40.   STRINGER, E.T. (1996): Action Research - A Handbook for Practitioners, SAGE publications, London.

41.   VAN VLIET, J.C. (1993): Software Engineering: Principles and Practice, John Wiley and Sons, Chichester, England.

42.   VON HALLE, B. (1991): "Data: Asset or Liability?", Database Programming and Design, 4, 7, July.

43.   WAND, Y. and WEBER, R.A. (1990): "A Model for Systems Decomposition", Proceedings of the Eleventh International Conference on Information Systems, J.I. DeGross, M. Alavi and H. Oppelland, (Eds.), Copenhagen, Denmark, December.

44.   WESTRUP, C. (1993): "Information Systems Methodologies in Use", Journal of Information Technology, 8.

45.   WITT, G.C. and SIMSION, G.C. (2000): Data Modeling Essentials: Analysis, Design, and Innovation, The Coriolis Group.

46.   WYNEKOOP, J.L. and RUSSO, N.L. (1997): "Studying Systems Development Methodologies: An Examination Of Research Methods", Information Systems Journal, 7, 1, January.