12-31-1997

Quality in Conceptual Modelling: Linking Theory and Practice

Graeme Shanks
Monash University

Peta Darke
Monash University

Follow this and additional works at: http://aisel.aisnet.org/pacis1997

Recommended Citation
http://aisel.aisnet.org/pacis1997/76

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 1997 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Quality in Conceptual Modelling: 
Linking Theory and Practice

Graeme Shanks & Peta Darke
Department of Information Systems, 
Monash University, 
P.O. Box 197, Caulfield East, 
Melbourne, Australia, 3145.

Email: graeme.shanks@is.monash.edu.au  
Telephone: +61 3 9903 2243  
Fax: +61 3 9903 2005

Executive Summary
Conceptual models form the basis of requirements specifications and are a important component of any systems development effort. The quality of those models will impact significantly upon the quality of the system which is eventually delivered. However, the notion of quality in conceptual modelling is not well-understood. In practice, much discussion of quality in conceptual modelling is ad hoc, unstructured and largely unproductive. Attempts to describe quality have concentrated mainly on providing lists of desirable features of conceptual models. Frequently these lists are poorly defined, contain overlapping features and are not based on any sound, underlying theory.

Recently two frameworks for quality in conceptual modelling have been proposed which address quality in a much more systematic and comprehensive way. The framework proposed by Krogsbie, Lindland and Sindre (1995) is soundly based in semiotic theory but lies at a high level of abstraction and is therefore difficult to apply in practice. The framework of Moody and Shanks (1994) supports the evaluation of models in practice but lacks a sound theoretical basis. This paper proposes a composite framework for understanding and evaluating quality in conceptual models which builds on and extends these two frameworks. The composite framework is represented as a meta-model which integrates and formalises the links between the two underlying frameworks. The composite framework assists in understanding how theory and practice inform each other by relating practice-based components to concepts grounded in theory.

The framework has been implemented as a hypertext tool with facilities which support explanation of concepts in the framework and how they are related. The tool also supports the evaluation and comparison of up to three alternative conceptual models. The composite framework and hypertext tool were evaluated by twenty experienced conceptual modelling practitioners and academics. The results indicate that the framework is useful for understanding and evaluating quality in conceptual modelling and the tool is useful for both explaining the framework and for evaluating and comparing the quality of conceptual models.

1. Introduction
Requirements definition is a critical activity in information systems development (Jarke et al. 1993, Greenspan, Mylopoulos and Borgida 1994). It is during this phase that requirements are determined and expressed as a requirements specification. Significant difficulties can arise during subsequent development activities if requirements specifications are incomplete, unclear, or incorrect (Roman 1985, Curtis, Krasner and Iscoe 1988), and difficulties and inadequacies in requirements have been identified as a major factor in information systems failures (Boehm 1981, Martin 1989). The conceptual models which form the basis of a requirements specification are therefore a significant component in the overall development effort, and the quality of those models will impact upon the quality of the system which is eventually delivered. Quality assurance in conceptual modelling needs to address both the process, the task of acquiring and modelling information system requirements, and the product, the resulting conceptual model (Moody and Shanks 1994).

Quality in conceptual modelling is not well understood. Attempts to describe it have mainly concentrated on providing lists and collections of desirable features and properties of conceptual models (Lindland, Sindre and Solvberg 1994). However two recent frameworks have addressed quality in conceptual modelling in a much more systematic and comprehensive way. Lindland, Sindre
and Solvberg (1994) framework is based in semiotic theory and defines quality in terms of the syntactic, semantic and pragmatic quality of models. It identifies quality goals and the means of achieving them as key concepts. The framework has been extended by Krogslet, Lindland and Sindre (1995) to incorporate a social dimension of quality in terms of modelling participants reaching agreement on conceptual models. Moody and Shanks (1994) have also proposed a framework for quality in conceptual models. It identifies the key concepts for supporting the evaluation of models in practice. These include quality factors, evaluation methods for measuring quality factors, strategies for improving values of quality factors, and weightings for indicating the relative importance in a specific development situation of individual quality factors. Although it focuses on data models, Moody and Shanks' framework is readily generalisable to any form of conceptual model.

The first of these two frameworks is soundly based in semiotic theory but "lies on a high level of abstraction" (Krogslet, Lindland and Sindre 1995) and therefore offers little to the practitioner. The second framework provides the components to support the evaluation of models in practice but lacks a sound basis in theory. In discussing the links between theory and practice in conceptual data modelling, Batra and Marakas (1995) note that "... there are indeed wide differences between the academic and the practitioner focus in conceptual data modelling". They argue strongly for greater synergy between theory and practice. Techniques are required that embed sound theoretical principles and yet are pragmatic and useable in practice. The purpose of this paper is to propose a framework for understanding and evaluating the quality of conceptual models which links theory and practice in this way.

The paper first presents an overview of previous research in quality in conceptual modelling. The frameworks of Krogslet, Lindland and Sindre (1995) and Moody and Shanks (1994) are described and their contributions to the notion of quality in conceptual modelling are discussed. The frameworks are integrated to form an extended and comprehensive composite framework. The meta-model of the composite framework formalises the links between the two underlying frameworks. The fourth section of the paper describes a hypertext tool which has been developed to implement the composite framework. The tool supports the use of the framework to address quality in the conceptual modelling process as well as in the evaluation of conceptual models. The results of an empirical study of the use of the tool are presented. Finally, the implications of the composite framework for improving the effectiveness of conceptual modelling practice are discussed, and areas for further research are identified.

2. Approaches to Quality in Conceptual Modelling

Research into quality in conceptual modelling has focused on the quality of the end product. Lists and taxonomies of desirable features and properties of conceptual models have been published (e.g. Roman 1985, Batini, Ceri and Navathe 1992, Simsson 1994, LeVithin and Redman 1995). Although a useful starting point for identifying quality factors for evaluating conceptual models, Lindland, Sindre and Solvberg (1994) note that many existing lists of properties are unstructured and use vague definitions, that properties often overlap, that properties of models are often confused with language and method properties, and that many quality goals are unrealistic. Comprehensive frameworks have recently been proposed which attempt to organise and structure the key concepts and features of quality in conceptual modelling. Lindland, Sindre and Solvberg (1994) have developed a theory-oriented framework for understanding quality in conceptual modelling. Moody and Shanks (1994) have proposed a practice-oriented framework for evaluating the quality of conceptual data models, and Pohl (1994) has described a framework which identifies three dimensions of the requirements engineering process and defines three associated goals for a requirements specification (i.e. complete, agreed, and formally represented). Lindland, Sindre and Solvberg's framework has been extended by Krogslet, Lindland and Sindre (1995) to incorporate the social agreement goal from Pohl's framework. An overview of these frameworks and other approaches to quality in conceptual modelling is presented in Table 1. This classifies and compares them according to the dimensions of purpose, feature, type and focus. None of these approaches to conceptual modelling quality has been tested empirically. A number of empirical studies of conceptual data modelling have compared data modelling formalisms and contrasted expert and novice performance (e.g. John and Naumann 1985, Shoval and Even-Chelme 1987, Batra, Hoffer and Bostrum 1980, Maiden and Sutcliffe 1982, Kim and March 1995). Although these use the quality of the model produced as a basis for comparison, most of them define quality only in terms of one property, i.e. completeness. These are therefore not considered to be studies which focus on the notion of quality in conceptual modelling.
The most comprehensive theory-oriented approach is the framework proposed by Krogstie, Lindland and Sindre (1995), and the most comprehensive practice-oriented approach is the framework proposed by Moody and Shanks (1994). These two frameworks are described in the remainder of this section. The composite framework proposed in section three of this paper builds on and extends these two frameworks.

### Table 1 Approaches to Quality in Conceptual Modelling

<table>
<thead>
<tr>
<th>Approach</th>
<th>Purpose</th>
<th>Features</th>
<th>Type</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poli (1994)</td>
<td>Defining goals and process dimensions for requirements modelling</td>
<td>Specification, representation and agreement dimensions</td>
<td>Framework</td>
<td>Theory</td>
</tr>
<tr>
<td>Moody and Shanks (1994)</td>
<td>Evaluating the quality of entity relationship (ER) models</td>
<td>Quality factors, strategies and evaluation methods</td>
<td>Framework</td>
<td>Practice</td>
</tr>
<tr>
<td>Simson (1994)</td>
<td>Defining quality features in ER models</td>
<td>Design and evaluation of alternative models</td>
<td>List</td>
<td>Practice</td>
</tr>
<tr>
<td>Batini, Carlo and Navathe (1992)</td>
<td>Improving the quality of a database schema</td>
<td>Quality features of a good schema, schema transformations</td>
<td>List</td>
<td>Practice</td>
</tr>
<tr>
<td>Roman (1985)</td>
<td>Defining properties of requirements specifications</td>
<td>Properties linked to their use in the design process</td>
<td>List</td>
<td>Theory</td>
</tr>
<tr>
<td>Levitin and Redman (1985)</td>
<td>Defining quality features in entity relationship models</td>
<td>Quality dimensions, reinforcement and trade-offs</td>
<td>List</td>
<td>Practice</td>
</tr>
</tbody>
</table>

2.1 The Framework of Krogstie, Lindland and Sindre (1995)
Krogstie, Lindland and Sindre's framework is based in semiotic theory and defines a conceptual model as a set of statements made in a language. The framework has five components: the model, language, domain, audience participation, and perceived knowledge. Model quality is defined by relationships between the model and the other four framework components. These relationships use four semiotic levels: syntactic, semantic, pragmatic and social (Stamper 1987). Syntax relates the model to its language, semantics relates the model to its domain, pragmatics relates the model to its interpretation by its intended audience, and the social level concerns the agreement of participants about the meaning of the model. Figure 1 shows the framework components and their relationships.

For each semiotic level the quality goals and the means of achieving those goals are defined. The means concept distinguishes between modelling activities which improve the quality of a model and the properties of a model which enable those activities to be carried out. The concept of feasibility is introduced to qualify the quality goals where it is recognised that a goal cannot be fully achieved and a realistic compromise is necessary.
The *model* is a set of statements made in some language, for example the entity relationship notation. The *language* is the set of all possible statements in that language. *Syntactic quality* is the adherence of a model's statements to the syntax rules of the language. The goal is feasible syntactic correctness and the means of achieving it is syntax checking, which is enabled by the property of a model having a formal syntax. The *domain* is the set of all statements which would be correct and relevant. It represents the ideal solution or model. *Semantic quality* is the correspondence between the model's statements and statements in the domain. However, the domain is an ideal concept which is inaccessible to the audience in reality, thus it is not possible to determine semantic quality. The concept of *participant knowledge* is therefore more relevant.

*Participant knowledge* is the set of all correct and relevant statements about the problem situation according to the participants' knowledge about the domain. *Perceived semantic quality* is the correspondence between statements in the participants' knowledge and the audience's interpretation. The goals are perceived feasible completeness and perceived feasible validly, and the means are consistency checking, statement insertion and deletion, and audience training in knowledge about the domain and the model. The *audience interpretation* is the set of statements that the audience thinks the model contains. Different parts of the model will be of interest to different participants, so the audience interpretation consists of overlapping subsets. *Pragmatic quality* is the correspondence between a part of a model and the relevant actors' interpretation of it, and the goal is feasible comprehension (that the model is understood, not the understandability of the model). Means to achieve this include model properties of expressive economy, structuredness and executability, which are enabled by model visualisation, diagram layout, explanation, animation and simulation activities.

*Social quality* has the goal of feasible agreement between the actors, where inconsistencies between the various actors' interpretations of the model are resolved. Relative agreement (audience interpretations may differ but remain consistent) is more realistic than absolute agreement (all audience interpretations are the same). Means of achieving feasible agreement include model comprehension and conflict modelling, enabled by viewpoint analysis, conflict resolution and model merging.

The framework provides a sound theoretical basis for understanding quality in conceptual modelling. It relates quality goals to the means of achieving them. However, these are defined at an abstract level, and thus are difficult for practitioners to readily understand and utilise.

2.2 The Framework of Moody and Shanks (1994)

Moody and Shanks (1994) propose a framework for evaluating the quality of conceptual data models which is practice-oriented. It is intended to provide practitioners with a coherent approach to resolving two key problems which arise in the practice of data modelling: the need to choose between a
number of alternative data models, and the need to understand and accommodate the different views of the various stakeholders in the data modelling process. Both problems require decisions about the data model based on the notion of model quality.

The first problem focuses on what makes one model of higher quality than another, and the second focuses on how the quality of a data model can be improved to meet the needs and expectations of all stakeholders. These issues are considered to be related, as quality is defined as the features and characteristics of a product that bear on its ability to satisfy given needs, which will vary between stakeholders and over time. Thus quality is a relative concept, and will be defined differently by different stakeholders in different problem situations and at different points in the modelling process, as knowledge and understanding of requirements emerges. Moody and Shanks' framework provides a systematic basis for the evaluation of data models in practice, and includes components which will be seen as useful and usable by practitioners. However, because of its practitioner focus, there is less emphasis on theoretical foundations, and thus the components themselves lack a strong basis in theory. It is recognised that practitioners "choose methods based on whether they are useful rather than if they are theoretically sound" (Moody and Shanks 1994). The strength of the framework lies in its utility in practice. The framework consists of seven main components: model, quality factor, stakeholder, evaluation method, weighting, rating and strategy. These are shown in Figure 2.

![Diagram showing the relationships between model, quality factor, stakeholder, evaluation method, weighting, rating, and strategy.]

Figure 2 Concepts in the Framework of Moody and Shanks (1994)
A model is a conceptual data model which is being developed, represented as an entity relationship model. A quality factor is a desirable property of a data model. The goal of the evaluation process is to maximise the value of the model with respect to these quality factors. Five quality factors are proposed which are based on practical experience and published taxonomies: correctness, completeness, understandability, flexibility and simplicity. These may be augmented as required. A stakeholder is a participant (group) involved in the data modelling process (eg. business users, data analysts, application developers, data administrators). Different stakeholders have different perspectives on the quality of a data model. A strategy is a process or activity which can be used to increase the value of a data model with respect to one or more quality factors. Strategies may involve the use of automated techniques as well as human judgement and insight. An evaluation method or metric is a systematic way of measuring a quality factor. In most cases this is a subjective rating by a relevant stakeholder or external expert, although objective metrics have been defined for some quality factors. A weighting is a value assigned to a quality factor which represents its relative importance in the context of the project. Different stakeholders may assign different weightings to a quality factor, and quality factor weightings may vary from project to project. A rating is a value assigned to a quality factor representing its valuation in a particular model by a stakeholder.

For each quality factor a number of evaluation methods and improvement strategies can be defined. Thus, quality factors and the strategies for achieving them are separated, as are quality goals and means in Krogsdal, Lindland and Sindre's framework. This enables both the process and product dimensions of quality in conceptual modelling to be supported. Moody and Shanks' framework is applied to the evaluation and comparison of conceptual models using the evaluation method, rating and weighting components. The various stakeholders' ratings of a model indicate the extent to which...
the model at its current stage of development is meeting their needs and expectations. Decisions about selecting and applying strategies to improve the model need to take into account another aspect of the framework: that there will be interactions between the quality factors, so that increasing the value of one factor may decrease the value of another. Understanding of these interactions is necessary for tradeoffs between qualities to be made in an informed manner.

The framework provides a systematic basis for evaluating the quality of conceptual models in practice. It identifies the key components necessary to support the evaluation process (stakeholder, quality factor, weighting, evaluation method, rating) and the improvement of conceptual models (quality factor, strategy). These components can all be readily understood and applied by practitioners. The framework can be populated with different quality factors, stakeholders and weightings to suit the requirements of particular modelling situations. Quality factors are linked to stakeholders and weightings, supporting the view that quality is a relative concept which is perceived differently by different stakeholders in different situations.


A composite framework for understanding and evaluating quality in conceptual modelling has been developed by integrating the two frameworks described in the previous section and consolidating their components. A comparative analysis of the two frameworks enables identification of their areas of commonality and the merging of their components. The composite framework formalises the links between the two underlying frameworks.

The two frameworks share several similar components: "audience" and "stakeholder"; "goal", "property" and "quality factor"; and "activity" and "strategy". These are explored as a means of integrating the two frameworks. The concept "model" occurs in both frameworks, and could be identical, compatible or incompatible in the two frameworks. Other components are disjoint, and can be incorporated into the composite framework. Integration of the framework components is discussed below.

**Audience and Stakeholder**. The "audience" component in Krogstie, Lindland and Sindre's framework is more general than the "stakeholder" component in Moody and Shanks' framework as it includes both technical and human actors. "Audience" is therefore used in the composite framework to identify both human and technical actors as participants in the modelling process.

**Goal, Property and Quality Factor**. All six quality factors identified by Moody and Shanks can be mapped onto the goals and properties proposed by Krogstie, Lindland and Sindre. Correctness, completeness and understandability map onto the goals of syntactic correctness, perceived feasible completeness, and feasible comprehension. Simplicity maps onto the property expressive economy, and is thus a means of achieving comprehension. Flexibility is a useful addition to the properties in Krogstie, Lindland and Sindre's framework. It is a means of achieving completeness over time as requirements evolve. Although Krogstie, Lindland and Sindre argue that all specific quality factors can be subsumed by the generic quality goals and properties in their framework, it is clearly much more useful to practitioners in actual modelling situations to be able to discuss alternative models in terms of more specific quality factors. It has therefore been decided to retain goal, property and quality factor as separate components in the composite framework. "Goal" and "property" are thus more abstract components which underpin the practice-oriented component "quality factor" which supports the application of the practice-based part of the composite framework.

**Activity and Strategy**. Both Krogstie, Lindland and Sindre's "activity" and Moody and Shanks' "strategy" define the processes used to achieve the goals and improve quality factors. Both components can reasonably co-exist within the single concept of "activity": in the composite framework, Krogstie, Lindland and Sindre's "means" concept is at a higher level of abstraction, including both the activities for improving quality and the model properties which enable those activities.

**Model**. Both frameworks define a conceptual model as statements in a language. Krogstie, Lindland and Sindre refer to any conceptual model, whereas Moody and Shanks only discuss entity relationship models. However, the concepts of Moody and Shanks' framework are generalisable to the evaluation of quality in any form of conceptual model. There is thus one "model" concept in the composite framework.
The two frameworks have been merged to form the composite framework. The meta-model for the composite framework is shown in Figure 3.

![Diagram of the Composite Framework](image)

**Figure 3 Meta-model of the Composite Framework**

The components which derive from theoretical considerations of quality in conceptual modelling and are grounded in theory are indicated, as are the components deriving from practical considerations and which support the evaluation of quality in practice. Components which are both theory and practice-based are also indicated, so that it is possible to identify areas of overlap between theory and practice. This assists in understanding the links between them, and how each can inform the other. The composite framework can be applied at any stage in the conceptual modelling process, thus supporting quality in both the modelling product and the modelling process.

4. **Tool Support and Empirical Evaluation of the Composite Framework**

The composite framework for understanding and evaluating quality in conceptual modelling offers considerable scope for tool support. The Data Model Quality Advisor (DMQA) is a prototype tool implemented in Visual Basic and Access. It provides a hypertext explanation facility for the components of the composite framework and their interrelationships, and supports the evaluation and comparison of up to three conceptual models.

The hypertext explanation facility provides a graphical view of the composite framework as a user interface. Explanations and examples of any of the framework components can be viewed by selecting the appropriate icon within the graphical model. The user is able to navigate amongst components of the framework using the hypertext links provided. Users can also access the framework components via the meta-model which is implemented as an alternative user interface. The evaluation and comparison facility supports the allocation of weightings for quality factors by a number of stakeholders. Ratings for quality factors for up to three alternative conceptual models may be entered and stored. After all ratings have been entered, a summary of the evaluations with rankings for alternative models can be displayed. Figure 4 shows the ratings summary screen of the DMQA. The user may seek explanation of any component of the framework during evaluation and comparison using the explanation facility.
An empirical study to examine the useability and usefulness of the composite framework and the DMQA has been conducted. The study involved twenty experienced data modelling practitioners and academics each using the DMQA to learn about the framework and to evaluate three alternative data models for a small case example. Each participant then completed a questionnaire about the framework and their use of the tool.

![Ratings Summary Screen of the Data Model Quality Advisor (DMQA)](image)

Figure 4  Ratings Summary Screen of the Data Model Quality Advisor (DMQA)

Some results from the empirical study are shown in Table 2 below. Statements and participant's responses to those statements were obtained using a 5-point Likert scale (from 1 indicating "strongly agree" to 5 indicating "strongly disagree"). Full details of the study can be found in Tan (1995). The table shows the mean and standard deviation for ratings for each statement together with results from a t-test applied to determine if the average rating was significantly different to a normally distributed population with a mean value of 3.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Rating</th>
<th>T-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluating the quality of a conceptual data model is critical to the successful development of an information system</td>
<td>1.55 (0.60)</td>
<td>t(18)=-10.808</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>It is important to consider alternative conceptual data models within systems development</td>
<td>1.70 (0.73)</td>
<td>t(18)=-7.964</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>A framework for evaluating the quality of conceptual data models would constrain the way practitioners prefer to work: they would ignore it</td>
<td>3.15 (0.81)</td>
<td>t(18)=0.828</td>
<td>not significant</td>
</tr>
<tr>
<td>In practice, conceptual data models are evaluated in an ad hoc way based on personal opinion</td>
<td>1.80 (0.70)</td>
<td>t(18)=-7.665</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>The tool is useful when evaluating and comparing conceptual data models</td>
<td>2.25 (0.78)</td>
<td>t(18)=-4.300</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>The ratings summary screen helps to compare conceptual data models</td>
<td>1.80 (0.83)</td>
<td>t(18)=-6.466</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>The tool is irrelevant to understanding and using the quality framework</td>
<td>3.55 (0.83)</td>
<td>t(18)=2.983</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

Results from the empirical study provide strong support for the need to evaluate the quality of conceptual data models and for the use of a framework for understanding and evaluating conceptual data models in practice. The tool was also viewed favourably and found to be useful when evaluating and comparing conceptual data models. The summary screen shown in Figure 4 was considered particularly helpful.
5. Conclusion and Implications for Future Research

The quality of conceptual models can have a significant impact on the quality of the information system which is ultimately implemented. This paper contributes to understanding of the notion of quality in both the theory and practice of conceptual modelling. Two comprehensive frameworks for quality in conceptual modelling, one theory-oriented and one practice-oriented, have been discussed. The links between the two frameworks have been formalised by developing a composite framework which builds on the components within the two frameworks. The composite framework assists in understanding how theory and practice inform each other by relating components based in practice to concepts grounded in theory. This is an important step in linking the theory of quality in conceptual modelling with the evaluation of quality in practice. The composite framework can be applied at any stage in the conceptual modelling process, supporting quality in both product and process.

The composite framework has been implemented as a hypertext tool, the Data Model Quality Advisor (DMQA), with explanation facilities and support for the evaluation and comparison of up to three alternative conceptual models. An empirical study of the use of the composite framework and the DMQA has been conducted. Results of the study indicate that the framework is useful and useable in practice, and that the DMQA improves understanding of the framework and facilitates evaluation and comparison of alternative models. Enhancements to the tool are planned. These include extended support for the application of the framework components as modelling process guidelines and for the evaluation of models, and extensions to the DMQA to provide group support for collaborative conceptual modelling activities.

Further work is required to refine and extend the composite framework. Inclusion within the framework of reference to additional practice-based quality factors would extend the population of the framework. The activities defined within the framework could be incorporated directly into process models of conceptual modelling to help ensure quality within the process. Further empirical studies also need to be undertaken within an organisational setting to investigate and evaluate the use of the composite framework and the DMQA in practice.
References
Greenspan, S., Mylopoulos, J. and Borgida, A. "On Formal Requirements Modeling Languages: RML Revisited" in Proc. 16th Int. Conf. on Software Engineering, Sorrento, May 1994.