Perceived vs. Measured Quality of Conceptual Schemas: An Experimental Comparison

Samira Si-said Cherfi* Jacky Akoka Isabelle Comyn-Wattiau*****

* CEDRIC-CNAM, 292 Rue Saint Martin, F-75141 Paris Cedex 03,
** CEDRIC-CNAM and INT, *** CEDRIC-CNAM and ESSEC

sisaid@cnam.fr, akoka@cnam.fr, wattiau@cnam.fr

Abstract

Frequently the behaviour of an information system is functionally correct, but it does not meet some quality criteria, such as completeness, consistency, and usability. One way to enhance the capability of an information system is to consider its conceptual model quality as well as its functional behaviour. Conceptual model quality can be defined as a set of perceivable characteristics expressed with quantifiable parameters that may be objective and/or subjective. The aim of this empirical investigation is to evaluate and compare perceived and measured quality of different conceptual model versions of the same universe of discourse. This paper describes: a) a set of metrics (clarity, simplicity, expressiveness, minimality) applied to different versions of ER conceptual schemas, b) a framework enabling a comprehensive comparison of the conceptual schemas, c) an experimentation leading to the evaluation of the same schemas by information system (IS) stakeholders such as designers, end-users, and students, d) a comparison of the objective and subjective evaluations based on a sample of about 120 observations using different statistical methods. First results indicate that there exists a strong relationship between perceived and measured quality. A second result reveals a significant difference between groups of respondents in their ways to perceive conceptual schemas quality. Based on our experiment, we are able to identify quality criteria relevant to different groups of stakeholders, depending on several dimensions, such as their professional experience, and/or their specialization degree.

Keywords: quality, metrics, conceptual schema, experimentation, perceived quality, measured quality, quality criterion.

1 Introduction

The subject of conceptual schema quality evaluation has occupied a substantial part of the effort devoted to conceptual modelling. The impact of conceptual schema quality is of central concern to computer scientists, as well as to end-users, and more generally to those who seek to evaluate software quality. In this paper, we take a highly integrative approach towards the subject of conceptual model quality evaluation using two lenses: information systems lifecycle phases (design, implementation, and use) and perceived versus measured quality. Taking into account these two viewpoints enables a thorough analysis of the complementary issues surrounding conceptual models and quality. This paper fills a gap in the existing literature by proposing conceptual model quality evaluation from IS stakeholders' perspectives, confronting perceived vs. measured quality. The human factors of information systems (IS) stakeholders' perception lend conceptual modelling and its associated quality debates a further twist. This paper highlights the debates between two competing perspectives, respectively perceived and measured, in conceptual model quality evaluation.

As for the design phase, perceived quality by IS stakeholders, and more precisely by end-users, is shaped by numerous factors. Effective user participation in information system design is limited since conceptual modellers will seek wherever possible to incorporate their own views and understanding of the universe of discourse. Although a consultation with end-users can take place over the selection of a conceptual schema, there is nevertheless a lack of common language between conceptual modellers and end-users. It is likely that conceptual modellers' dependency on end-user expertise will provide the most compelling reason for granting end-users a degree of influence over the IS design phase. A key constraint for end-user participation in the design phase is the lack of necessary technical expertise allowing them to evaluate if system specifications were correctly translated into the conceptual model. This situation highlights the importance of evaluation of the resulting conceptual model. This is a significant way in determining the effectiveness of involvement in the design phase. This is also a rare opportunity for end-users to scrutinize design proposals at an early stage.

Regarding the implementation phase, the design of an IS implies the ability to bring it into successful operation.

User participation in the implementation phase of an IS is crucial, although it remains fragmented. This phase is generally carried out by IS professionals only. Given the uncertainties generated by the technological aspects of the solution, companies tend to delegate responsibilities to computer specialists who would then impose their solutions. End-users are understandably reluctant to participate in this implementation phase. This lack of involvement prevents end-users from effectively evaluating the solution chosen to be implemented.

Finally, during the use phase several data quality problems result from a non effective use of the IS. We argue that the perceived quality of the conceptual model by IS stakeholders is a predictor of the ease-of-use of the IS. The more the conceptual model is appreciated by end-users, the more its implementation will lead to an intuitive representation of data, processes and business rules, then enabling an effective use of the IS.

We argue that the meaning of IS is fundamentally linked to the relationship between subjectivity and objectivity. The subjective elements include the perception of an IS by its stakeholders. The objective elements include the way in which designers and implementers control the technologies. The latter are viewed as a mean to reduce uncertainty. An IS is characterized by tools, techniques, and languages, which are seen as highly objective elements of the system. On the contrary, end-users and more generally IS stakeholders are seen as having an ambivalent view of the IS, taking into account the wide socio-cultural context. Their perception is shaped by experience as well as by social relations.

The aim of this paper is precisely to compare perceived and measured quality of conceptual models integrating IS lifecycle phases. We propose a framework enabling a thorough comparison of perceived vs. measured model quality, using relevant criteria for each phase.

The rest of the paper is organized as follows. In Section 2, we survey the related work. Section 3 is devoted to the presentation of our framework. In Section 4, we provide a description of the survey undertaken to seek a better understanding of how some IS stakeholders and more precisely computer scientists, managers, and students, perceive the quality of a set of eight conceptual schemas of the same universe of discourse. Finally Section 5 summarizes the contribution and provides directions for further research.

2 Related work

The relevance of quality approaches devoted to conceptual design is increasingly recognized (Moody, 2005). Indeed, conceptual models are produced within the early stages of information systems development. Erroneous or bad conceptual schemas heavily impact the development of an IS, and the quality of the resulting system (Lausen and Vinter, 2001). Work related to our study can broadly be differentiated into three streams :

a) In the area of theoretical quality framework validation, some authors propose theoretical foundations for the broader field of software engineering measurement. A theoretical validation provides a mean to verifying whether a measure is valid with respect to some predefined properties. One of the directions was the usage of theoretical approaches as a mean to evaluate quality metrics proposed in the literature. The objective was to provide practitioners and researchers with a formal framework to construct better evaluation methods. There are mainly two approaches: property-based (Briand et al., 1996) and measurement theory-based approaches (Zuse et al., 1989, Poels and Dedene, 2000).

b) Regarding empirical quality framework validation, there are several types of validation being used for the assessment and the improvement of quality frameworks. Several empirical studies (Bodart et al., 2001, Gemino and Wand, 2005) resort to questionnaires gathering participants opinions on modeling methods. Gemino (2004), Genero et al. (2002), Maes and Poels, (2006) explore sets of independent variables (like complexity or modeling technique) and their effect on dependent variables such as maintainability or understandability.

c) Besides these two established research streams, it is worthwhile to mention related work on the evaluation of conceptual modeling quality in practice. Mendonça and Basili (2000) propose an approach for measurement frameworks improvement. The approach combines a knowledge discovery technique with the GQM paradigm. The validation was based on an industrial case study. Maier (2001) reported three experiments on data modeling and data management in an organizational context. Six organizational dimensions are proposed as key factors influencing quality of data modeling. Moody (2005) discussed the problem of lack of knowledge about practices in conceptual model quality.

While the three streams described above are very active research orientations, to the best of our knowledge, none of these works actually provide a framework for comparing perceived and measured quality of conceptual schemas for different types of stakeholders. The research described in this paper is an attempt to achieve this objective.

3 The validation framework

The purpose of the current section is to define a validation framework for conceptual schemas evaluation comparing perceived and measured quality. In order to establish this framework, we draw on the generally acknowledged objective of quality evaluation using metrics. Our framework for evaluating conceptual schemas quality is based on IS lifecycle phases allowing us to perform a comparison between quality computed values and its evaluation as perceived by different stakeholders. We first define the concepts of perceived and measured quality. We then propose a refinement of past IS stakeholders typology. Finally we propose a validation framework meta-model.

3.1 Measured quality

Numerous authors have proposed and justified several metrics for conceptual modeling quality (Genero et al., 2000; Poels and Dedene, 2000). We have proposed a framework for assessing IS quality based on conceptual schema early evaluation (Si-said et al., 2002). In this framework, several metrics have been defined, such as minimality, clarity, simplicity, and expressiveness.

Clarity measures the ease with which a schema could be read. The metric we proposed in order to measure clarity is based on the heuristics stating that a schema containing N edges can have reasonably at most N crossings. In reality, it will have much less crossings as modelers make efforts on readability. **Minimality** measures the lack of factorization in conceptual schemas. **Expressiveness** measures the richness of a schema (Batini et al., 1992). We distinguish between concept and schema expressiveness. Concept expressiveness measures whether the concepts are expressive enough to capture the main aspects of the reality. Thus we propose to associate weights with the different concepts involved. Schema expressiveness measures the expressiveness of the schema as a whole. **Simplicity**. A schema is said to be simple if it constructed upon simple concepts. Our measure of simplicity is based on the assumption that the complexity of a conceptual schema grows with the number of relationships (including inheritance and aggregation links). Similar considerations can be found in (Genero et al., 2000). We argue that these metrics can be seen as objective measures of conceptual model quality in the sense that they are independent of the characteristics of the stakeholders and of the context. Another reason for looking at them as objective measures is that they can be computed automatically. Moreover, they can be incorporated into a CASE tool. Finally, they can serve as a guidance tool available to IS designers. They represent our measured quality.

3.2 Perceived quality

IS end-users, and more generally stakeholders, have their perception shaped by their own experience, as well as by socio-cultural factors. Lindland's framework suggests evaluating the quality of a model using syntactic quality as well as semantic and pragmatic quality (Lindland et al., 1994). It is generally agreed that IS stakeholders rely on their perception of reality in order to evaluate semantic quality. Krogstie et al. (1995) have extended the Lindland framework by adding perceived semantic quality, described as the correspondence between the elements that the model should contain and the elements that the model contains according to the stakeholders' viewpoint. Maes & Poels (2006) have presented a research model for the evaluation of conceptual model quality based on user perception. In this paper, we build upon past approaches in order to compare perceived quality and measured quality. It is a cross validation of both measures. If perceived quality is confirmed by measured quality, stakeholders' perception can be considered as less subjective allowing designers to rely on their intuitive evaluation. If measured quality is in line with perceived quality, we argue that the proposed metrics are empirically validated.

3.3 Stakeholders

Several types of stakeholders have been identified by Preiss and Wegmann (2001). In the literature, IS stakeholders fall into three groups: end-users, management and IS professionals. This typology has been criticized since it does not take into account organizational and inter-organizational environments (Pouloudi, 1999). Moreover, this typology seems to be static and does not reflect the current multifaceted concerns of IS development. In this paper, we differentiate between end-users and IS professionals. Then we refine this classification by distinguishing between different IS professionals based on their specialization domains (networks, systems, support, project management, conceptual design, etc.). Moreover, we integrate other characteristics such as sex, professional experience, etc.

3.4 The validation framework meta-model

Figure 1 describes the framework proposed for the validation purpose. This meta-model has four main concepts: quality characteristics, life cycle phases, stakeholders, conceptual schemas. It is built upon ISO9126 quality model which recursively defines quality characteristics, and basically composed of quality attributes associated with metrics (this is represented by the grey part of Figure 1). ISO9126 is considered as poor for effective quality assessment. Therefore, we extend it by integrating the concept of lifecycle phases. IS quality is defined by a set of characteristics. However the same quality attribute, for instance understandability, does not have the same meaning depending on the IS lifecycle phase (design, development, use, etc.). The consequence is that different metrics are used at each phase. The ISO9126 quality model is further enriched by integrating the stakeholders and the context. Thus, we can represent the perception of each stakeholder on each conceptual schema based on a specific quality characteristic. This perception leads to an evaluation that we propose to compare to automatically computed metrics, characterizing the same quality attributes.

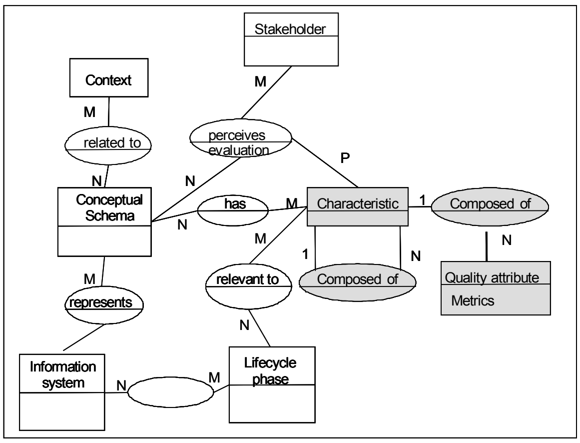

Figure 1:The validation framework metamodel

This framework has been experimented using eight different versions of a conceptual schema representing the same universe of discourse.

4 Empirical validation using the framework

In this section, we use the validation framework to establish a comparison between perceived and measured conceptual model quality for different types of IS stakeholders. Let us remind that there is not a single quality criterion aggregating properly different characteristics in order to build a unique quality viewpoint that can be shared by different stakeholders. As a consequence, we rely on several quality criteria. In our case, we take into account in our experiment the four quality criteria mentioned above. To test our framework, we conducted one experiment involving 113 stakeholders. We describe below this experimentation. We first describe the participants. We then successively present the experimental tasks, the operational procedure, the measures obtained for the validation, and finally the research findings.

4.1 Participants

The 113 stakeholders represent both IS professionals (87) and end-users (26). The group of IS professionals is composed of different specialists. Each IS professional had the possibility to mention one or two specialization domains (computer networks:13, operating systems:15, design and development:48, help desk:18, project management:23). IS professionals were also asked to characterize their experience in several computer areas: database (72), programming (68), information system (59), and conceptual modeling (50). In the same way, end-users were asked for their IS type of use (internet: 18, office automation: 19, business software:8). 88 persons are males whereas 25 are females. We can distinguish between students without any experience (26) and participants having several years of experience (87). The average age of the sample is 31 with a range between 21 and 57. In this experiment, we consider that it is important that the groups should not be homogeneous, with respect to their professional and educational background, as well as working experience. Otherwise, global comparison between perceived and measured quality would not be significant. Moreover, the richness of the sample allows us to discriminate between subgroups using different criteria (sex, industrial experience, profile, age, education, etc.). Let us mention that not all the previous criteria were relevant. However, they all have been evaluated, leading to interesting results.

4.2 Operational procedure

In the experiment, each participant received 8 ER diagrams, representing the same universe of discourse. They were asked to rank the 8 models according to four criteria: clarity, simplicity, expressiveness, and minimality. More precisely, the criteria were explained to the participants as follows:

- Clarity: A model is said to be clear if it's easy to understand,

- Simplicity: A model is said to be simple if it can be understood quickly,

- Expressiveness: A model is said to be expressive if it describes the reality in details,

- Minimality: A model is said to be minimal if it contains no redundant concepts.

Besides ranking the schemas, the experiment consisted of a series of questions related to sex, age, company, educational background, working experience, and IS use. All the questions were adapted from a questionnaire which has been validated previously to the experiment.

4.3 Measures

The questionnaire was sent to 500 participants from 400 companies. The same questionnaire was given to 100 students from a business school and a university computer science department. Students' participation was voluntary. Let us remind that the final sample is composed of 87 IS professionals and 26 management and computer science students. Using our metrics, we computed the ranking for the eight models.

4.4 Research Findings

We performed two types of validation: the first validation was performed on the whole sample leading to a global evaluation of each metric compared to the most frequent participant answers. The second validation, called subgroups validation, was performed on subsets of the data based on the general characteristics of the participants, such as sex, educational background, professional experience, etc. In order to perform these validations, we first computed the ranking of the eight models using our metrics. They represent the measured quality. Finally we performed a correspondence analysis in order to picture the association that may exist between the characteristics of the participants and the issue of perceived versus measured quality values.

4.4.1 Global validation results

In order to evaluate the difference between our measures (considered to be objective) and the perceived rankings by the stakeholders, we proceeded in the following manner:

a) For each quality criterion, we transformed the participant ranking into marks from one to eight.

b) In the same way, we assigned marks one to eight to the models based on our own measures.

c) We computed the total gap between participant ranking and our ranking .

There are several reasons that can justify using the frequent perceived values rather than the averages. The most important one is due to the fact that the standard deviation is significantly high. A second reason is related to the difficulties encountered by the participants to use the full ranking scale. For these reasons we decided to consider the most frequent values assigned by the participants to each model. It has the advantages to discard questionnaires that have not been filled in a rigorous manner.

4.4.2 Subgroups validation results

In this section, we perform a comparison of perceived and measured quality values for several subgroups of the sample. To form these subgroups, discriminant variables have been used such as sex, educational background, professional background, IS experience, etc…In order to perform the comparisons, we proceeded as follows:

For each participant and for each criterion, we only considered the extreme models, namely the best and the worst. The values obtained represent the perceived values.

We computed our metrics. For each metric, we only considered the two most extreme models. The values obtained represent the measured values.

Then, we performed a comparison between the perceived and the measured values.

The reason why we only considered extreme models is due to the fact that respondents had difficulties ranking eight different models. It seems easier to them to rank extreme models, namely the best and the worst for each criterion, than ranking all the models. Participants found more meaningful and easier for them to group models in two or three classes rather than using the full ranking scale. Taking into account this factor, we decided to consider only extreme models for each criterion.

4.4.3 Correspondence analysis results

We then performed a correspondence analysis, which is a statistical method to analyze a two-way table. It allows us to analyze: (i) the row profiles composed of the participants characteristics (sex, educational background, professional background, etc…), (ii) the column profiles representing our models and metrics, (iii) both the row and the column profiles allowing us to picture the association between the levels of the two dimensions of the contingency table. The aim is to have a global view of the data that is useful for analyzing the association between perceived and measured quality values.

4.5 Research Findings

In this section we present and discuss the main results obtained by performing a global validation, subgroups validation, and finally a correspondence analysis.

a) Global validation results

We have compared 8 models according to 4 criteria on a scale varying from 1 to 8. The difference between the perceived quality by the respondents and our own results is equals to 40 for the 32 evaluations performed. The value of this cumulative deviation represents an interesting result. Given the range of the scale (7), the average deviation is expected to be equal to 3.5 for one evaluation, and for 32 evaluations to 112. The difference is significant. One first conclusion is that our metrics are perceived to be relevant. The respondents behave in their evaluation process the same way as we do. There is no real difference between the perceived values and the measured values. As a consequence, it is reasonable to think that our metrics are valid.

If we take into account only the two extreme models (i.e. the worst and the best for each criterion), the difference equals to 4. This is an excellent result if we compare it to the average deviation which is equals to 28 (8*3.5). As a first conclusion, we have strong reasons to believe that our metrics are valid. They are considered as relevant by the respondents. The latter tend to behave in their quality evaluation process the same way as indicated by our own metrics. As a consequence we can assert that there is no significant difference between perceived and measured quality based on our set of metrics.

b) Subgroups validation results

Due to space limitations, we do not include the table which is equivalent to the contingency table but expressed in percentages. However, our analysis will take into account those percentages. The best score is obtained by the model having the best clarity. The worst model in terms of expressiveness is model 1 for 87% of the sample. 96% of developers and 73% of end users tend to consider this model as the less expressive. The worst minimal model is model 1 for 85% of the respondents. This number is equal to 98% of the developers. It remains difficult to explain why the models having the worst clarity and the worst simplicity are not the same for the respondents. 49% of the sample consider model 4 as the worst in terms of simplicity. Although this result is not as high as the other results, nevertheless our metrics are still valid. The lowest scores are those related to expressiveness. It seems to us that expressiveness and minimality rely more on conceptual schema semantics understanding. They are harder to understand. Their underlying concepts do not lead to a consensus. With this relative limitation our metrics proved once again to be relevant and can globally be validated.

c) Correspondence analysis results

Let us remind that the aim of this exploratory technique is to analyze the columns, the rows, and their association. By analyzing this contingency table, we obtain the following results, first by comparing the rows (i.e. the characteristics of the respondents). The characteristics (sex, education background, etc.) of the participants are independent from the metrics used to evaluate the conceptual models quality. Two dimensions can explain about 81% of the variance. These dimensions are extracted in order to maximize the distances between the row points (or the columns points) of the contingency table. The first dimension can explain 50.62% of the Chi2 value. The cumulative percentage explained by the two dimensions is equal to 80.95%. The contribution of the third dimension is marginal. It is therefore discarded. End users and IS professionals have different perceptions of the conceptual schemas quality evaluation. Within the first group, we find mainly computer scientists and developers with a strong programming experience. The second group is mainly composed with respondents having an experience in office automation, and internet use. These two groups behave differently in terms of quality evaluation from a third group composed of conceptual modellers, project managers, information systems specialists, and surprisingly students. Let us notice that there is a significant difference between males and females in their perception of models quality. Finally, it is interesting to mention that computer networks and operating systems specialists, although opposed in their quality evaluation, seem to behave in a singular manner.

If we compare the columns (i.e. the metrics used to evaluate the quality), we find the following results. Conceptual models having the best simplicity and minimality seem to be opposed to models having the worst clarity and simplicity. Conceptual schemas with the worst expressiveness and minimality appear to be in opposition to those characterized by the best clarity and expressiveness. Two other differences are worth to mention: Models considered to have the best simplicity are significantly opposed to those having the best expressiveness. This is the case for those having the worst minimality and the worst clarity.

5 Conclusion and further research

This paper proposed a framework for conceptual model quality validation using an ER meta-model based on ISO9126, enriched with lifecycle design phases and an IS stakeholders typology. This framework has been experimented to validate four metrics using a questionnaire filled by 113 persons aware of conceptual modelling, including IS professionals as well as end-users. Several statistical analysis were conducted to validate the adequacy between perceived conceptual modelling quality and measured quality. Moreover, a correspondence analysis allowed us to discriminate between several homogeneous or heterogeneous groups of respondents. This leads to the conclusion that the validation of a conceptual schema must be adapted depending on the stakeholders. The survey confirms also that not all computer scientists are conceptual modellers. Further research will be conducted to validate other metrics and to perform regression analysis allowing us to build predictive models.

6 References

Batini, C., Ceri, S. and Navathe, S.B. (1992) *Conceptual database design : An entity relationship approach*, Redwood City, California, Benjamin Cummings.

Bodart, F., Patel, A., Sim, M., Weber, R. (2001) Should Optional Properties Be Used in Conceptual Modelling? A Theory and Three Empirical Tests. *Information Systems Research* **12**: 384-405.

Briand L. C., Morasca S., Basili V. R. (1996) Property-Based Software Engineering Measurement. *IEEE Trans. Software Eng.,* **22**: 68-86.

Gemino, A. Wand, Y. (2005) Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties *Data & Knowledge Engineering,* **55**(3):301-326.

Gemino, A. (2004) Empirical comparisons of animation and narration in requirements validation. *Requir. Eng.* **9**(3):153-168.

Genero, M., Jiménez, L., and Piattini, M. (2000): Measuring the quality of entity relationship diagrams. *Proceedings ER2000 Conference on Conceptual Modeling*, LNCS **1920**:513-526, Springer Verlag.

Genero, M., Jiménez, L., Piattini, M. (2002) A Controlled Experiment for Validating Class Diagram Structural Complexity Metrics*, Proceedings of 8th Object-Oriented. Information Systems, Montpellier, France.*

Krogstie, J., Lindland, O. I., and Sindre, G. (1995): Defining quality aspects for conceptual models. ISCO 1995: 216-231.

Lausen, S., Vinter, O. (2001) Preventing Requirement Defects: An Experiment in Process Improvement*, Requirements Engineering* **6**:37-50.

Lindland, O.I., Sindre, G., and Sølvberg, A. (1994): Understanding quality in conceptual modelling. *IEEE Software*, **11**(2):42-49.

Maes, A., Poels, G. (2006) Evaluating Quality of Conceptual Models Based on User Perceptions. *Proceedings of 25th Int. Conf. ER 2006,* LNCS 4215:54-67.

Maier, R. (2001) Organizational concepts and measures for the evaluation of data modeling in *Developing quality complex database systems: practices, techniques and technologies*, 1–27.

Mendonca, M.G., Basili, V.R. (2000) Validation of an Approach for Improving Existing Measurement Frameworks; *IEEE Transactions on Software Engineering,* 26(6):484–499.

Moody D.L. (2005) Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. DKE. **55**(3): 243-276.

Poels, G., and Dedene, G. (2000): Measures for dynamic aspects of object-oriented conceptual schemes. *Proceedings ER2000 Conference on Conceptual Modeling*, LNCS **1920**: 499-512, Springer Verlag.

Pouloudi, A (1999). Aspects of the Stakeholder Concept and their Implications for Information Systems Development. *Proc. of the 32nd Hawaii Int. Conf. on System Sciences.*pp. 1-17.

Preiss, O, and Wegmann, A. (2001). Stakeholder Discovery and Classification Based on Systems Science Principles. *APAQS* 2001: 194-198.

Si-Said Cherfi S., Akoka, J., and Comyn-Wattiau, I. (2002) Conceptual Modeling Quality - From EER To UML Schemas Evaluation" *Proceedings ER2002 Conference on Conceptual Modeling*, LNCS **2503**: 414-428, Springer Verlag.

Zuse, H. Bollmann, P. (1989): Software Metrics - Using Measurement Theory to Describe the Properties and Scales of Static Software Complexity Metrics, *SIGPLAN Notices* **24**(8):23:33, New York.