

Evaluating the Quality of Information Models: Empirical Testing of a Conceptual Model Quality Framework

Daniel L. Moody^{1,2}
Guttorm Sindre^{3,4}
Terje Brasethvik³
Arne Sølvsberg³

¹Department of Software Engineering, Charles University, Prague, Czech Republic

²School of Business Systems, Monash University, Melbourne, Australia

³Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

⁴Department of Management Science and Information Systems, University of Auckland, New Zealand
email: dmoody@infotech.monash.edu.au; {guttors, brase, asolvberg}@idi.ntnu.no

Abstract

This paper conducts an empirical analysis of a semiotics-based quality framework for quality assuring information models. 192 participants were trained in the concepts of the quality framework, and used it to evaluate models represented in an extended Entity Relationship (ER) language. A randomised, double-blind design was used, in which each participant independently reviewed multiple models and each model was evaluated by multiple reviewers. A combination of quantitative and qualitative analysis techniques were used to evaluate the results, including reliability analysis, validity analysis, interaction analysis, influence analysis, defect pattern analysis and task accuracy analysis. An analysis was also conducted of the framework's likelihood of adoption in practice. The study provides strong support for the validity of the framework and suggests that it is likely to be adopted in practice, but raises questions about its reliability and the ability of participants to use it to accurately identify defects. The research findings provide clear directions for improvement of the framework. The research methodology used provides a general approach to empirical validation of quality frameworks.

1. INTRODUCTION

1.1 The Importance of Conceptual Model Quality

The traditional thrust of software quality assurance has been to use “brute force” testing at the end of the development process [56]. However empirical studies show that more than half the errors which occur during systems development are the result of inaccurate or incomplete requirements [30, 34]. Requirements errors are also the most common cause of failure in systems development projects [54, 55].

This suggests that substantially more effort should be spent to quality assure conceptual models, in order to catch requirements defects as soon as they occur, or to prevent them from occurring altogether [64]. Empirical studies have shown that moving quality assurance effort up to the early phases of development can be 33 times more cost effective than testing done at the end of development [58]. However, it is during requirements analysis that the notion of software development as a craft rather than an engineering discipline is strongest, and quality is therefore most difficult to assess. There are few guidelines for evaluating the quality of conceptual models, and little agreement even among experts as to what makes a “good” model. As a result, the quality of conceptual models in practice is almost entirely dependent on the competence of the analyst [28, 35].

1.2 Previous Research on Conceptual Model Quality

There have been a number of frameworks proposed in the literature for evaluating the quality of conceptual models [e.g. 3, 27, 28, 32, 33, 37, 57, 62]. The most serious deficiency in the existing literature is the lack of empirical justification: most frameworks are either justified based on theory or the author(s)' experience. Theoretical justification is limited because methods have no "truth" value—the validity of a method is an empirical rather than a theoretical question [24, 48]. Also, software engineering is replete with examples of methods which are theoretically sound but do not work in practice, and vice versa. Experiential justification is limited because personal experience is subject to bias. Also, a method which works well when used by one person may not work for another [63].

1.3 Objectives of this Paper

This paper conducts an empirical analysis of the quality framework proposed by Lindland et al [33] for evaluating information models. The broad research questions addressed are:

- Does the framework provide a reliable and valid basis for evaluating the quality of information models? Reliability and validity are important characteristics of any quality evaluation method [16, 17].
- Is the framework likely to be adopted in practice? This is an issue that is rarely addressed in IS design research, despite the fact that this is critical for research to have an impact on practice.

This is the first quantitative study of the use of a quality framework to quality assure information models.

2. CONCEPTUAL MODEL QUALITY FRAMEWORK

In this section, we define the conceptual model quality framework developed by Lindland et al [33]. This represents the *a priori* theory being tested by this research.

2.1 Overview of the Framework

The conceptual model quality framework tested in this research study is based on semiotic theory (the theory of signs), and has been developed for evaluating conceptual models of any type (e.g. data models, process models, interaction models). The quality of a conceptual model is defined with respect to four components or *sets*:

- The model (M) is the set of statements that has been represented.

- The language (L) is the set of all statements that are possible to make according to the grammar of the language.
- The domain (D) is the set of statements that would be correct and relevant to make about the problem domain, often called the Universe of Discourse (UoD), the part of the world to be modelled.
- The interpretation (I) is what the audience of relevant stakeholders think is expressed in the model.

2.2 Quality Categories

The framework defines three *quality categories* and separates between *quality goals* and *means* to reach the goals. The quality categories are based on the three levels of semiotic theory [41], and are defined by equations on the sets described above:

- Syntactic quality is a relationship between the model and the language. The goal is *syntactic correctness*, which can be described as $M \setminus L = \emptyset$, i.e., there are no statements included in the model that are not part of the language.
- Semantic quality is a relationship between the model and the domain. Here two goals can be identified:
 - *Validity*: $M \setminus D = \emptyset$, i.e., there are no statements in the model that are not correct and relevant about the domain
 - *Completeness*: $D \setminus M = \emptyset$, i.e., there are no statements that are correct and relevant about the domain, but are not included in the model
- Pragmatic quality is the relationship between the model and the audience's interpretation. For each stakeholder, the goal is *comprehension*, which can be defined as $M \setminus I_i = \emptyset$ & $I_i \setminus M = \emptyset$, i.e., there are no statements in the model which are not in the stakeholder's interpretation of the model, and vice versa.

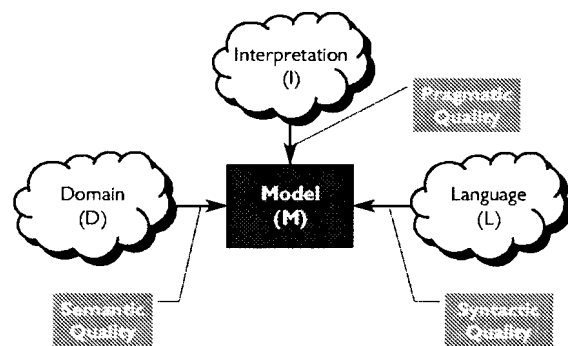


Figure 1. Conceptual Model Quality Framework

3. RESEARCH METHODOLOGY

3.1 Validation of IS Design Methods

The question of how to validate IS design methods has been a longstanding issue in the IS field [e.g. 18, 24, 44, 45, 46, 60, 63]. There are inherent problems evaluating any design method since there is typically no theory, no hypotheses, no experimental design and no data analysis to which traditional evaluation criteria can be applied [60]. As a result, IS design research tends to emphasise the development of new design methods while addressing the evaluation of methods in only a limited fashion [8, 13, 18, 36, 61, 63]. The Method Evaluation Model [39] is a theoretical model for evaluating IS design methods, which incorporates two aspects of method “success”: actual efficacy and adoption in practice. It combines Methodological Pragmatism, a theory for validating methodological knowledge [48], and the Technology Acceptance Model (TAM), a theoretical model for explaining and predicting user acceptance of information technology [15]. The constructs of the model and the relationships between them are summarised in Figure 2. This model has been validated using both laboratory and field experiments [39].

The definitions of the constructs are:

- Actual Efficiency: the extent to which the method reduces the effort required to perform the task;
- Actual Effectiveness: the extent to which the method improves the quality of the result;
- Perceived Ease of Use: the extent to which a person believes that using the method would be free of effort;
- Perceived Usefulness: the extent to which a person believes that the method would be useful;
- Intention to Use: the extent to which a person intends to use the method;
- Actual Usage: the extent to which the method is actually used.

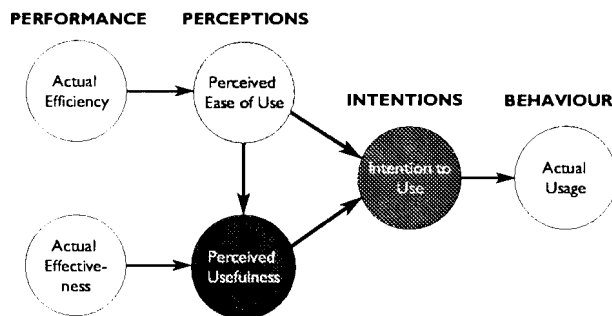


Figure 2. Method Evaluation Model

Actual Efficiency and Actual Effectiveness measure whether the method actually improves task performance, while Perceived Ease of Use and Perceived Usefulness represent perceptions of the method's efficiency and effectiveness.

3.2 Research Design

The research design is summarized in Figure 3. This shows the experimental treatment, experimental tasks, materials and dependent variables. A randomised, double-blind design was used, in which each participant reviewed multiple models and each model was evaluated by multiple reviewers.

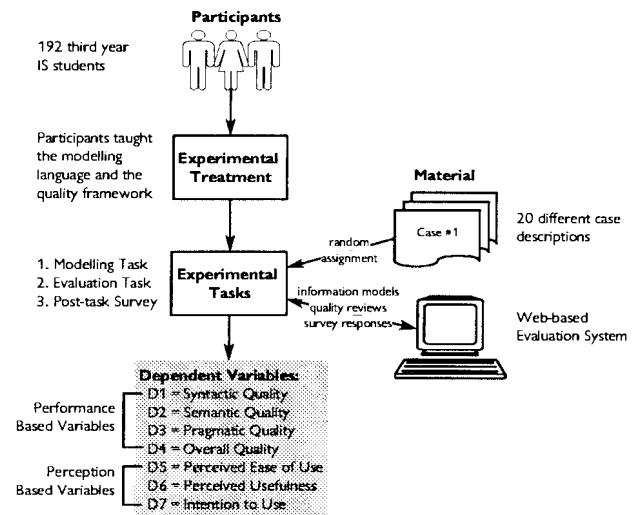


Figure 3. Research Design

3.3 Participants

There were 192 participants in this study, all 3rd year Information Systems students at the Norwegian University of Science and Technology. Computer Science or Information Systems students have been used in most previous experimental studies of conceptual modelling [e.g. 4, 5, 6, 21, 31, 42, 51, 52, 59].

3.4 Experimental Treatment

To prepare the participants for the evaluation task, they were given the original paper defining the framework [33], two 45 minute lectures explaining the framework and a 45 minute lecture explaining the evaluation exercise itself.

3.5 Experimental Tasks

Modelling Task. A total of twenty different case descriptions were prepared for use in the experiment. This was done to minimise the chance of collaboration between participants. Each participant had two weeks to develop an information model to meet the requirements of the case they were assigned, and submit it using a web-based system. The modelling language used was an extended Entity Relationship (ER) notation, called the Referent Modelling Language [7, 20, 53].

Evaluation Task. Following completion of the information modelling task, three reviewers were assigned to evaluate each model. Models were randomly assigned to reviewers, and reviews were double blind (reviewers did not know whose models they were evaluating, and modelers did not know who their reviewers were). Each participant was assigned cases which were different to each other and to the one they had modelled themselves. Participants had one week to complete the evaluation task, and record the results of their reviews using the web-based evaluation system.

Post Task Survey. Finally, participants were required to answer a post task survey, which measured their perceptions of the framework, also using the web-based system. The survey consisted of 16 closed questions and 3 open questions. The closed questions consisted of 6 items used to measure Perceived Ease of Use, 8 items to measure Perceived Usefulness and 3 items to measure Intention to Use. Each item was measured on a 5-point Likert scale, using the opposing-statements format. The order of the items was randomised to avoid monotonous responses [23].

3.6 Dependent Variables

We distinguish between two types of dependent variables:

- Performance based measures: How well did subjects perform the evaluation task?
- Perception based measures: How effective did subjects perceive the framework to be?

These types of measures represent the difference between actual efficacy and perceived efficacy as defined in the Method Evaluation Model.

Performance Based Variables. Four dependent variables were used to evaluate performance on the evaluation task: Syntactic Quality, Semantic Quality, Pragmatic Quality, and Overall Quality. A total of 550 ratings of each type (syntactic, semantic, pragmatic, overall) were collected as part of the study (an average of three per model). These were used to assess the reliability and validity of the evaluation framework (actual efficacy). All four quality

ratings were given on a 7 point Likert scale, from 1 (poor) to 7 (excellent). As well as rating the models, reviewers were required to identify all defects in the models, classified by quality category.

Perception Based Variables. We define three perception-based dependent variables for evaluating the framework. These are all constructs defined in the Method Evaluation Model.

- Perceived Ease of Use: this was measured using five items on the post-task survey (Questions 1, 10, 11, 14 and 16).
- Perceived Usefulness: this was measured using eight items on the post-task survey (Questions 2, 5, 6, 7, 9, 12, 13 and 15).
- Intention to Use: this was measured using three items on the post-task survey (Questions 3, 4 and 8).

Items used to measure these constructs were based on the items defined for the corresponding constructs in the Technology Acceptance Model, which have been shown to have high levels of reliability and validity [9, 14, 23].

3.7 Theoretical Model

The relationship between all dependent variables and their empirical indicators are shown in Figure 4. In the diagram, circles indicate latent variables, rectangles indicate observed variables (empirical indicators), dotted lines indicate measurement relationships, while solid lines indicate causal relationships. The empirical indicators associated with the performance based variables represent the different reviewer ratings, while those for the perception based variables are the items on the post task survey.

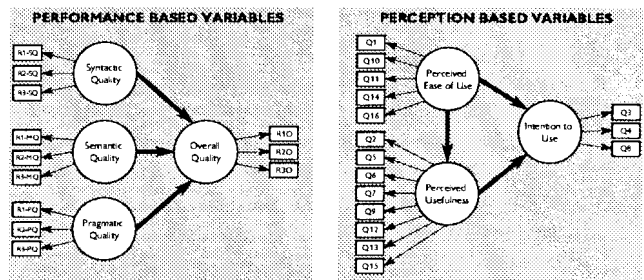


Figure 4. Theoretical Model

4. RESULTS AND DISCUSSION

4.1 Summary of Analyses

The results were evaluated by a combination of quantitative and qualitative analysis techniques. Each of these analyses addresses a sub-question of one of the research questions defined at the beginning of this paper:

1. Reliability analysis: How consistently were reviewers able to apply the framework?
2. Validity analysis: Are the quality categories necessary, sufficient and independent determinants of information model quality?
3. Influence analysis (weight estimation): What was the relative influence of each quality category on decisions about the overall quality of a model?
4. Interaction analysis: What are the relationships between the quality categories?
5. Defect pattern analysis: What was the frequency of errors in each quality category?
6. Task accuracy: How accurately were participants able to identify quality defects using the quality framework?
7. Perceived ease of use: How easy did participants find the framework to use?
8. Perceived usefulness: How useful did participants find the framework?
9. Intention to use: Were the participants likely to use the framework in the future?

Analyses 1-6 evaluate the framework's *actual efficacy* in performing the task. Analyses 7-9 evaluate the framework's *perceived efficacy*. Analyses 1-4 are quantitative analyses, analyses 5-6 are qualitative analyses, while analyses 7-9 involve a combination of the two (both closed and open questions were included in the post-task survey). Mixing qualitative and quantitative research methods can lead to a more comprehensive understanding of a phenomenon [25].

4.2 Reliability Analysis

Reliability was evaluated by measuring the inter-rater reliability between different reviewers of the same model. The results are summarised in Table 1. While there is no definitive standard for reliability, alphas of 0.7 or above are generally considered to be acceptable in the literature [43].

Table 1. Reliability Levels for Each Quality Category

CONSTRUCT	CRONBACH'S α
Syntactic Quality	.6508
Semantic Quality	.6617
Pragmatic Quality	.7043
Overall Quality	.7293

The observed levels of reliability for two of the quality categories (syntactic and semantic) are clearly lower than acceptable, which suggests that the definition of these categories needs to be refined. This was confirmed by the

qualitative analysis, which showed that participants confused the two categories. The most encouraging result of this analysis is that overall quality assessments were at an acceptable level of reliability. The likely explanation for this is that consideration of quality within each category led to an overall increase in reliability, even if there was some ambiguity in the classification of defects.

4.3 Validity Analysis

To assess the validity of the quality framework, we need to address the following questions [1, 40]:

- *Completeness (sufficient)*: Do the set of quality categories include all aspects of quality of information models? Are there any aspects that have been left out?
- *Parsimony (necessary)*: Are all the quality categories necessary for evaluating information models? Are they all relevant determinants of information model quality?
- *Independence (orthogonal)*: Are the quality categories independent of each other? Do they overlap?

A regression analysis was carried out using the quality categories (syntactic, semantic, pragmatic) as independent variables and Overall Quality as the dependent variable (Figure 5). The results of the regression analysis were:

- Adjusted $r^2 = 0.877$
- Significance (p) = 0.000
- Pearson correlation coefficient (r) = .937

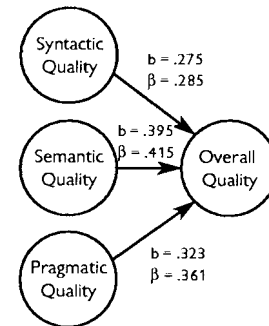


Figure 5. Regression Model for Validity Analysis

Completeness of Quality Categories. The r^2 statistic measures the proportion of the variance of the dependent variable explained by the independent variables. In this case, the quality categories account for almost 90% of the variance in Overall Quality. Ratings of overall quality are almost entirely explained by variations in ratings for the individual quality categories. This provides strong evidence that the model is fully specified (i.e. that there are no missing independent variables) and therefore that the set of quality categories is complete.

Parsimony of Quality Categories. Table 2 shows the size and significance of the separate effects of each independent variable on the dependent variable. All independent variables (quality categories) were found to have highly significant effects ($\alpha < .01$) on the dependent variable (Overall Quality). *Effect size* (ES) has been suggested by a number of researchers as an index of practical meaningfulness [10, 22, 49, 50]. In multiple regression, the standardised regression coefficient (β) is most commonly used as a measure of effect size [47]. As shown in Table 2, Syntactic Quality has a “small” effect ($\beta > .1$) while Semantic and Pragmatic Quality have “medium” effects ($\beta > .3$), following Cohen’s guidelines for effect size [10]. This means that all relationships are practically meaningful as well as statistically significant, which provides strong evidence that all quality categories are relevant determinants of the quality of an information model.

Table 2. Multiple Regression Results

QUALITY CATEGORY	STATISTICAL SIGNIFICANCE (p)	EFFECT SIZE (b)	TOLERANCE
Syntactic	.000	.285	.478
Semantic	.000	.415	.455
Pragmatic	.000	.361	.475

Independence of Quality Categories. Collinearity analysis was also conducted to evaluate the independence of the quality categories. As shown in Table 2, Tolerance values are well above .2 for all quality categories, which means there is no evidence of multicollinearity between the quality categories. This suggests that they are all independent determinants of Overall Quality.

4.4 Influence Analysis

Regression analysis was used to evaluate the separate effects of each quality category on judgements of overall quality. The regression equation which resulted was:

Equation 1. $Overall\ Quality = .275\ Syntactic\ Quality + .395\ Semantic\ Quality + .323\ Pragmatic\ Quality + .021$

The coefficients for each variable in the regression equation measure the effect of each quality category on judgements about the Overall Quality of a model. For example, an increase of one unit in Syntactic Quality is associated with an increase of .275 in Overall Quality. However the relative size of the standardised regression coefficients (β) for the variables indicates their relative influence in determining overall quality. This shows that Syntactic Quality contributes 27% to the total variance explained, Semantic Quality 39% and Pragmatic Quality 34%. As expected, Semantic Quality had the greatest

effect on decisions about overall quality of a model—these represent mismatches between the model and user requirements, so are the most costly and visible type of errors.

4.5 Interaction Analysis

Bi-variate correlation analysis was conducted to test for interactions between the different quality categories. Strong positive correlations were found between all quality categories:

- Syntactic and Semantic Quality ($r = .669, p = .000$)
- Syntactic and Pragmatic Quality ($r = .652, p = .000$)
- Semantic and Pragmatic Quality ($r = .672, p = .000$)

Correlations are non-directional, but our theoretical explanation for the observed correlations is that they represent causal relationships (shown in Figure 6). Improving syntactic quality is expected to improve semantic quality, as a model that is not syntactically correct will have ambiguous semantics. Improving syntactic quality is expected to improve pragmatic quality because a model that is not syntactically correct will be difficult to interpret. Improving pragmatic quality is expected to improve semantic quality, as a model that is difficult to interpret will be difficult to match against the real world.

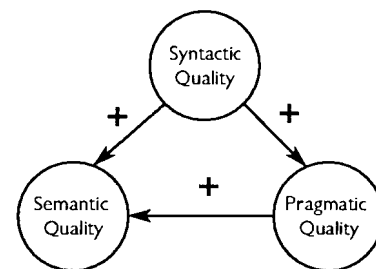


Figure 6. Interactions Between Quality Categories

Three alternative multiple regression models were developed to test this explanation (these represent all possible causal combinations of the variables):

- Model 1: Syntactic Quality + Semantic Quality → Pragmatic Quality
- Model 2: Syntactic Quality + Pragmatic Quality → Semantic Quality
- Model 3: Semantic Quality + Pragmatic Quality → Syntactic Quality

The results of the analysis are summarised in Table 3. While all regression models were statistically significant, Model 2 (our *a priori* theory) provides the best fit (as evidenced by the higher variance explained).

Table 3. Alternative Regression Models

MODEL	ADJUSTED r^2	SIGNIFICANCE
1	.524	.000
2	.543	.000
3	.521	.000

4.6 Defect Pattern Analysis

Rather than evaluate all 192 models and 550 reviews, a random sample of 32 models and 96 reviews were selected. The most frequent defects were (defined in terms of average number of defects per model):

- Cardinality errors on relationships (4.5).
- Missing attributes (3.6).
- Missing names for relationships (2.6).
- Missing entity types (1.3).
- Erroneous use of aggregation, membership and generalisation connectors (1.0).
- Missing relationships (0.5).
- System components as entity types (0.4), e.g., "System" or "Database".
- Activities modelled as entity types (0.4).

Entities without attributes and relationship types without names fall into the category of syntactic errors (given that the notation requires that all entities should have attributes, and all relationships should be named). Otherwise, most of the others fall into the semantic category, and are mostly incompleteness defects (something mentioned in the case description but not covered by the model).

The high level of errors in cardinality of relationships indicates a systematic problem in the information modelling notation and/or the teaching of it. Relationship cardinalities are a basic element of information modelling syntax and one would not expect such high error rates for such a basic syntactic error. Further analysis showed that participants found the conventions used for cardinalities confusing, as they were the reverse of notations they had learnt in previous courses. This explains the high level of errors and also the high level of underreporting of errors. This illustrates how the quality framework can provide useful diagnostic information about the effectiveness of modelling notations and teaching methods used.

4.7 Task Accuracy

Task accuracy was measured by inspecting the sample reviews and comparing them to expert reviews:

- Ratio of Type I errors (false negatives): where defects exist in the model but are not identified by reviewers. On average, 75% of the defects went unreported.

- Ratio of Type II errors (false positives): when defects are identified which are not defects at all. Of the defects identified by the participants, 92% were genuine defects, so 8% were Type II errors.
- Ratio of Type III errors (classification errors): where defects are correctly identified but classified in the wrong category. About 20% of the errors were wrongly categorised. Most of these were confusion between syntactic and semantic quality, which explains the lower reliability for these quality categories.

This analysis shows that participants were not able to apply the framework accurately to identify defects, with very high levels of underreporting. While this could be partly attributed to their lack of experience in information modelling, it also suggests a possible deficiency in the framework. A clear weakness of the framework is that quality categories are defined at such an abstract level. Decomposing the framework to a lower level of detail would help to improve identification of defects and would also improve the reliability of evaluations.

4.8 Likelihood of Adoption in Practice

Measurement Validity. To evaluate the results for Perceived Ease of Use, Perceived Usefulness and Intention to Use, it is necessary first to evaluate the validity and reliability of the survey items used to measure them.

Factor analysis was conducted on the items used in the post-task survey using the principal components extraction method and varimax rotation. The initial factor analysis resulted in four factors being extracted using the Kaiser criterion. Q5 (Perceived Ease of Use) and Q14 (Perceived Usefulness) formed a fourth factor, and were therefore eliminated from the analysis. Factor weights derived from the analysis were then used to estimate the value of the underlying theoretical constructs.

Reliability analysis was conducted on the items used to measure Perceived Ease of Use, Perceived Usefulness and Intention to Use (excluding Q5 and Q14). As shown in Table 4, high levels of reliability were found for all constructs, with Cronbach's alpha > .7 in all cases, which is considered acceptable [43].

Table 4. Item Reliabilities for each Construct

CONSTRUCT	CRONBACH'S α
Perceived Ease of Use	.7055
Perceived Usefulness	.8661
Intention to Use	.9036

Significance Testing. One-sample t-tests were conducted for each construct to see if they were significantly different to 3 (the "zero point" of the Likert scale use). Table 5

summarises the results of the significance testing. Overall, participants found the framework easy to use, useful, and intended to use it in the future ($\alpha < .01$).

Table 5. Significance of Responses

CONSTRUCT	MEAN	STDEV	SIGNIFICANCE	Y/N
Perceived Ease of Use	3.77	.570	.000	YES
Perceived Usefulness	3.72	.665	.000	YES
Intention to Use	4.04	.665	.000	YES

4.9 Analysis of Causal Relationships

In the Method Evaluation Model (Figure 2), a number of causal relationships are hypothesised:

- Perceived Ease of Use $\rightarrow$ Perceived Usefulness
- Perceived Ease of Use + Perceived Usefulness $\rightarrow$ Intention to Use

These relationships were evaluated using regression analysis.

Perceived Ease of Use $\rightarrow$ Perceived Usefulness. The results for this regression analysis were:

- $r^2 = .647$
- $\alpha = .000$
- $r = .804$

The Pearson correlation coefficient (r) is greater than .5, which corresponds to a “large” effect (r is used as a measure of effect size in simple regression) [11]. This means that the relationship between Perceived Ease of Use and Perceived Usefulness is both statistically significant ($\alpha < .01$) and practically meaningful. The resulting regression equation was:

Equation 2. *Perceived Usefulness = .937 Perceived Ease of Use + .182*

Perceived Ease of Use + Perceived Usefulness $\rightarrow$ Intention to Use. The results for this regression analysis were:

- Adjusted $r^2 = .322$
- $\alpha = .003$
- $r = .610$

The Pearson correlation coefficient is greater than .5, which corresponds to a “large” effect [11]. This means that the relationship between the variables is both statistically significant ($\alpha < .01$) and practically meaningful. The resulting regression equation was:

Equation 3. *Intention to Use = .035 Perceived Ease of Use + .682 Perceived Usefulness + 1.333*

The significance levels and effect sizes for each independent variable in Equation 3 are shown in Table 6 below.

Table 6. Multiple Regression Results

VARIABLE	SIGNIFICANCE (p)	EFFECT SIZE (b)
Perceived Ease of Use	(.897)	.035
Perceived Usefulness	.035	.582

The relationship between Perceived Usefulness and Intention to Use was found to be significant ($\alpha < .05$) with a large effect size ($\beta > .5$), while the relationship between Perceived Ease of Use and Intention to Use was neither statistically significant ($\alpha < .05$) nor practically meaningful ($\beta < .1$). This is consistent with many of the studies of the Technology Acceptance Model (on which the Method Evaluation Model is based), which have found that the relationship between Perceived Ease of Use and Intention to Use is not significant after controlling for the effects of Perceived Usefulness [9, 15].

5. CONCLUSION

5.1 Summary of Findings

This paper has conducted an empirical analysis of the conceptual model quality framework proposed by Lindland et al [33]. The validation approach combines both quantitative and qualitative research approaches, providing triangulation of method. While the quantitative analyses suggested that participants were able to successfully apply the framework (based on reliability and validity analysis), the detailed qualitative analysis showed that their evaluations included high levels of underreporting and misclassification of defects. This suggests that use of quantitative analysis techniques on their own can lead to deceptive results.

Overall, the quality framework was found to be valid, with the quality categories found to be necessary, sufficient and independent. It was perceived to be both easy to use and useful in evaluating information models and participants intended to use it in the future. The research findings provide clear direction for future research, which will focus on refining the framework to address the limitations identified, particularly in terms of accurate identification of defects, as highlighted in the qualitative analysis. Finally, the paper conducts a further empirical test of the Method Evaluation Model, which showed the measurement instrument had a high level of reliability and validity and validated two of the three causal relationships in the model.

The results for each of the research questions defined at the beginning of the paper are summarised in Table 7:

Table 7. Summary of Findings

RESEARCH QUESTION	RESULT
1. Task Performance (Actual Efficacy)	
1.1 Is the framework reliable?	Partially - two quality categories had reliability levels lower than acceptable but overall quality was acceptable
1.2 Is the framework valid?	Yes – the quality categories were found to be necessary, sufficient and independent
1.3 What is the relative influence of each quality category on decisions about overall quality?	Syntactic Quality 27% Semantic Quality 39% Pragmatic Quality 34%
1.4 What are the interactions between the quality categories?	Syntactic → Semantic Syntactic → Pragmatic Pragmatic → Semantic
1.6 How accurately were participants able to identify defects?	75% Type I errors 8% Type II errors 20% Type III errors
2. Adoption In Practice (Perceived Efficacy)	
2.1 Did participants find the framework easy to use?	Yes
2.2 Did participants find the framework useful?	Yes
2.3 Are participants likely to use the framework in practice?	Yes

5.2 Theoretical Contributions

The major theoretical contribution of this paper is that it conducts the first quantitative analysis of a conceptual model quality framework. The lack of empirical validation is a major weakness in the conceptual model quality literature, and this paper demonstrates a systematic approach to testing and improving such frameworks. It could also be used to compare the effectiveness of different frameworks, using a multiple group design. The validation approach addresses issues of reliability and validity, which are important requirements of any quality evaluation process, as well as the issue of adoption in practice, which has been largely ignored in IS design research. The results of the analyses provide clear direction for future research, as well as identifying problems in the information modelling technique and teaching methods used. More generally, the validation approach described

in this paper could be applied to validate and improve any quality framework. There is nothing about the approach that is specific to conceptual modelling. It could just as easily be used to evaluate a framework for software quality, a framework for user interface design quality or a framework for website design quality.

5.3 Limitations and Further Research

The major limitation of this study, and it is a significant one, is the nature of sample population used. In general, the population from which one selects subjects for a research study should be representative of the population to which the researcher wishes to generalise results [12]. Computer Science or Information Systems students have been used in most previous experimental studies of conceptual modelling [e.g. 4, 5, 6, 21, 31, 42, 51, 52, 59]. However external validity is a significant problem in most laboratory experiments involving undergraduate students, and has been identified as a major issue in IS research in terms of its relevance to practice [19, 26, 38]. It is also one of the major limitations of experimental research in the social sciences generally [2].

The question therefore remains: would similar results have been found if practitioners (which is the population to which we wish to generalise) had been used instead? The argument could be made that undergraduate students have little or no practical experience in information modelling, and therefore do not have the knowledge or expertise to properly evaluate the quality of information models. Lacking such experience, students may have been tempted to score overall quality simply as an average of their scores for syntactic, semantic, and pragmatic quality. Such averaging would inflate the validity of the quality framework. Similarly, both validity and reliability might be inflated by defensive reviewing (i.e. assigning scores of 3-5 rather than 1-2 or 6-7). These are valid criticisms and difficult to argue against convincingly.

For this reason, we do not see this study as the end of the validation process but only the *beginning*. This study has provided useful empirical data about the framework and pilot-tested the validation approach, but the results must be interpreted with caution. The generalisability of the results to practice is questionable because of the nature of the participants used. For this reason, a similar study is planned using experienced practitioners, possibly with a refined version of the quality framework. Some refinement of the framework has already taken place [29], but the results of this study suggest the need for other refinements.

REFERENCES

- [1] Alexander, C. (1968): *Notes On The Synthesis Of Form*, Boston: Harvard University Press.
- [2] Babbie, E.R. (1998): *The Practice of Social Research*, Belmont, California: Wadsworth Publishing.
- [3] Batini, C., S. Ceri, and S.B. Navathe (1992): *Conceptual Database Design: An Entity Relationship Approach*, Redwood City, California: Benjamin Cummings.
- [4] Batra, D., J.A. Hoffer, and R.P. Bostrom (1990): Comparing Representations with Relational and EER Models, *Communications of the ACM*, **33** (2), February: p. 126-139.
- [5] Bock, D.B. and T. Ryan (1993): Accuracy in Modelling with Extended Entity Relationship and Object Oriented Data Model, *Journal Of Database Management*, **4** (4): p. 30-39.
- [6] Bodart, F., A. Patel, M. Sim, and R. Weber (2001): Should the Optional Property Construct be used in Conceptual Modelling: A Theory and Three Empirical Tests, *Information Systems Research*, **12** (4).
- [7] Brasethvik, T. and J.A. Gulla (2002): A Conceptual Modeling Approach to Semantic Document Retrieval, In A. Banks Pidduck (Ed.) *14th International Conference on Advanced Information Systems Engineering (CAiSE'02)*, Toronto: Springer Verlag, May 27-31.
- [8] Bubenko, J.A. (1986): Information Systems Methodologies - A Research View, In *Information Systems Design Methodologies: Improving The Practice*, T.W. Olle, H.G. Sol, and A.A. Verrijn-Stuart (Eds.), North-Holland.
- [9] Chau, P.Y.K. (1996): An Empirical Assessment of a Modified Technology Acceptance Model, *Journal of Management Information Systems*, **13** (2).
- [10] Cohen, J. (1988): *Statistical Power Analysis for the Behavioural Sciences (2nd ed.)*, 2nd ed, Hillsdale, NJ: Lawrence Earlbaum and Associates.
- [11] Cohen, P.A. (1981): Student Ratings of Instruction and Student Achievement: A Meta-Analysis of Multi-Section Validity Studies, *Review of Educational Research*, **51**: p. 281-309.
- [12] Cooper, D.R. and P.S. Schindler (1998): *Business Research Methods (6th edition)*, 6th ed, Singapore: McGraw-Hill International.
- [13] Curtis, B. (1986): By The Way, Did Anyone Study Any Real Programmers?, In *Empirical Studies of Programmers*, E. Soloway and S. Iyengar (Eds.), Ablex: Norward, N.J.
- [14] Davis, F.D. (1989): Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology, *MIS Quarterly*, **13** (3).
- [15] Davis, F.D., R.P. Bagozzi, and P.R. Warshaw (1989): User Acceptance of Computer Technology: A Comparison of Two Theoretical Models, *Management Science*, **35** (8): p. 982-1003.
- [16] Deming, W.E. (1986): *Out of the Crisis*, Cambridge, MA: MIT Center for Advanced Engineering.
- [17] Evans, J.R. and W.M. Lindsay (2002): *The Management and Control of Quality (5th Ed.)*, Cincinnati, Ohio: South-Western (Thomson Learning).
- [18] Fitzgerald, G. (1991): Validating New Information Systems Techniques: A Retrospective Analysis, In H.E. Nissen, H.K. Klein, and R. Hirschheim (Eds.), *Information Systems Research: Contemporary Approaches And Emergent Traditions*: North-Holland,
- [19] Galliers, R.D. (1994): Relevance and Rigour in Information Systems Research: Some Personal Reflections on Issues Facing the Information Systems Research Community, In *Proceedings of the IFIP TC8 Conference on Business Process Reengineering: Information Systems and Challenges*, Gold Coast, Australia,
- [20] Gulla, J.A. and T. Brasethvik (2000): On the Challenges of Business Modelling in Large-scale Reengineering Projects, In *4th International Conference on Requirements Engineering (ICRE 2000)*, Schaumburg, IL: IEEE Computer Society Press, 19-23 June 2000.
- [21] Hardgrave, B.C. and N.P. Dalal (1995): Comparing Object Oriented and Extended Entity Relationship Models, *Journal Of Database Management*, **6** (3).
- [22] Harris, M.J. and R. Rosenthal (1985): Mediation of Interpersonal Expectancy Effects: 31 Meta-Analyses, *Psychological Bulletin*, **97**.
- [23] Hu, P.J. and P.Y.K. Chau (1999): Examining the Technology Acceptance Model Using Physician Acceptance of Telemedicine Technology, *Journal of Management Information Systems*, **16** (2), Fall: p. 91-113.
- [24] Ivari, J. (1986): Dimensions Of Information Systems Design: A Framework For A Long Range Research Program, *Information Systems Journal*, (June).
- [25] Kaplan, B. and D. Duchon (1988): Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study, *MIS Quarterly*, December.
- [26] Keen, P.G.W. (1991): Relevance and Rigour in Information Systems Research: Improving Quality, Confidence, Cohesion and Impact, In *Information Systems Research: Contemporary Approaches and Emergent Traditions*, H.-E. Nissen, H.K. Klein, and R. Hirschheim (Eds.), Elsevier Science Publishers (North Holland).
- [27] Kesh, S. (1995): Evaluating The Quality Of Entity Relationship Models, *Information And Software Technology*, **37** (12).
- [28] Krogstie, J., O.I. Lindland, and G. Sindre (1995): Towards a Deeper Understanding of Quality in Requirements Engineering, In *Proceedings of the 7th International Conference on Advanced Information Systems Engineering (CAISE)*, Jyväskylä, Finland, June.
- [29] Krogstie, J. (1998): Integrating the Understanding of Quality in Requirements Specification and Conceptual Modelling, *Software Engineering Notes*, **28** (1).
- [30] Lauesen, S. and O. Vinter (2000): Preventing Requirement Defects, In *Proceedings of the Sixth International Workshop on Requirements Engineering: Foundation for Software Quality (REFSQ'2000)*, Stockholm, Sweden, June 5-6 2000.

- [31] Lee, H. and B.G. Choi (1998): A Comparative Study of Conceptual Data Modelling Techniques, *Journal Of Database Management*, **9** (2), Spring: p. 26-35.
- [32] Levitin, A. and T. Redman (1994): Quality Dimensions of a Conceptual View, *Information Processing and Management*, **31** (1).
- [33] Lindland, O.I., G. Sindre, and A. Sølvsberg (1994): Understanding Quality In Conceptual Modelling, *IEEE Software*, **11** (2), March: p. 42-49.
- [34] Martin, J. (1989): *Information Engineering*, Englewood Cliffs, N.J.: Prentice Hall, 3 v.
- [35] Moody, D.L. and G.G. Shanks (1994): What Makes A Good Data Model? Evaluating the Quality of Entity Relationship Models, In P. Loucopoulos (Ed.) *Proceedings of the Thirteenth International Conference on the Entity Relationship Approach*, Manchester, England, December 14-17.
- [36] Moody, D.L. and G.G. Shanks (1998): Evaluating and Improving the Quality of Entity Relationship Models: An Action Research Programme, *Australian Computer Journal*, November.
- [37] Moody, D.L. and G.G. Shanks (1998): What Makes A Good Data Model? A Framework For Evaluating And Improving The Quality Of Entity Relationship Models, *Australian Computer Journal*, August.
- [38] Moody, D.L. (2000): Building Links Between IS Research and Professional Practice: Improving the Relevance and Impact of IS Research, In R.A. Weber and B. Glasson (Eds.), *International Conference on Information Systems (ICIS'00)*, Brisbane, Australia, December 11-13.
- [39] Moody, D.L. (2001): *Dealing with Complexity: A Practical Method for Representing Large Entity Relationship Models (PhD Thesis)*, Melbourne, Australia: Department Of Information Systems, University of Melbourne.
- [40] Moody, D.L. and G.G. Shanks (2003): Improving the Quality of Data Models: Empirical Validation of a Quality Management Framework, *International Journal of Information Systems*, **forthcoming**.
- [41] Morris, C.W. (1970): *Foundations of the Theory of Signs*, Chicago: Chicago University Press.
- [42] Nordbotten, J.C. and M.E. Crosby (1999): The Effect of Graphic Style on Data Model Interpretation, *Information Systems Journal*, **9**.
- [43] Nunally, J. (1978): *Psychometric Theory (2nd edition)*, New York: McGraw Hill.
- [44] Olle, T.W., H.G. Sol, and A.A. Verrijn-Stuart, eds. *Information Systems Design Methodologies: A Comparative Review*. 1982, North-Holland: Amsterdam.
- [45] Olle, T.W., H.G. Sol, and C.J. Tully, eds. *Information Systems Design Methodologies: A Feature Analysis*. 1983, North-Holland: Amsterdam.
- [46] Olle, T.W., H.G. Sol, and A.A. Verrijn-Stuart, eds. *Information Systems Design Methodologies: Improving the Practice*. 1986, North-Holland: Amsterdam.
- [47] Pedhazur, E.J. and L.P. Schmelkin (1991): *Measurement, Design and Analysis: An Integrated Approach*, Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- [48] Rescher, N. (1977): *Methodological Pragmatism: Systems-Theoretic Approach to the Theory of Knowledge*, Oxford: Basil Blackwell.
- [49] Rosenthal, R. and D.B. Rubin (1979): A Note on Percent Variance Explained as a Measure of Importance of Effects, *Journal of Educational Psychology*, **74**.
- [50] Rosenthal, R. (1983): Assessing the Statistical and Social Importance of the Effects of Psychotherapy, *Journal of Consulting and Clinical Psychology*, **51**.
- [51] Shoval, P. and M. Even-Chaime (1987): Database Schema Design: An Experimental Comparison Between Normalisation and Information Analysis, *Database*, **18** (3), Spring: p. 30-39.
- [52] Shoval, P. and S. Shiran (1997): Entity-Relationship and Object-Oriented Data Modeling: An Experimental Comparison of Design Quality, *Data & Knowledge Engineering*, **21**: p. 297-315.
- [53] Sølvsberg, A. (1999): Data and what they refer to, In *Conceptual Modeling, Current Issues and Future Directions (Selected Papers from the Symposium on Conceptual Modeling, Los Angeles, CA, held before ER'97)*. P.P. Chen, et al. (Eds.), Springer Verlag: Berlin, p. 211-226.
- [54] Standish Group (1995): *The CHAOS Report into Project Failure*, The Standish Group International Inc. Available on-line at <http://www.standishgroup.com/visitor/chaos.htm>.
- [55] Standish Group (1996): *Unfinished Voyages*, The Standish Group International Inc. available on-line at <http://www.standishgroup.com/visitor/voyages.htm>.
- [56] van Vliet, H. (2000): *Software Engineering: Principles and Practice (2nd edition)*: John Wiley & Sons.
- [57] von Halle, B. (1991): Data: Asset or Liability?, *Database Programming and Design*, **4** (7), July.
- [58] Walrad, C. and E. Moss (1993): Measurement: The Key to Application Development Quality, *IBM Systems Journal*, **32** (3): p. 445-460.
- [59] Weber, R.A. (1996): Are Attributes Entities? A Study Of Database Designers' Memory Structures, *Information Systems Research*, June.
- [60] Weber, R.A. (1997): *Ontological Foundations Of Information Systems*, Melbourne, Australia: Coopers And Lybrand Accounting Research Methodology Monograph No. 4, Coopers And Lybrand.
- [61] Westrup, C. (1993): Information Systems Methodologies in Use, *Journal of Information Technology*, **8**.
- [62] Witt, G.C. and G.C. Simsion (2000): *Data Modeling Essentials: Analysis, Design, and Innovation*, The Coriolis Group.
- [63] Wynekoop, J.L. and N.L. Russo (1997): Studying Systems Development Methodologies: An Examination Of Research Methods, *Information Systems Journal*, **7** (1), January.
- [64] Zultner, R.E. (1992): The Deming Way: Total Quality Management for Software, In *Proceedings of Total Quality Management for Software Conference*, Washington, DC, April.