Applying AHP for Collaborative Modeling Evaluation:

Experiences from a Modeling Experiment

Denis Ssebuggwawo, Department of Computer Science, Faculty of Science, Kyambogo University, Kampala, Uganda

Stijn Hoppenbrouwers, Model Based Information Systems Group, Faculty of Engineering, HAN University of Applied Sciences, Arnhem, The Netherlands

Henderik A. Proper, Chair of Information Systems, Radboud University Nijmegen, The Netherlands & Public Research Center-Henri Tudor, Luxembourg

ABSTRACT

Collaborative modeling is one of the approaches used to enhance productivity in many enterprise modeling and system development projects. Determining the success of such a collaborative effort needs an evaluation of a number of factors which affect the quality of not only the end-products – the models, but also that of other modeling artifacts: the modeling language, the modeling procedure and the support tool. Although a number of quality frameworks have been developed, few of these frameworks have received practical validation and many offer little guidance about how the evaluation is operationalized. The Collaborative Modeling Evaluation (COME) framework presented in this paper offers a holistic approach to the evaluation of the four modeling artifacts. It employs the Analytic Hierarchy Process (AHP), a well-established method from Operations Research, to score the artifacts' quality dimensions and to aggregate the modelers' priorities and preferences. Results from a modeling experiment demonstrate both the theoretical and practical significance of the framework.

Keywords: *Analytic Hierarchy Process (AHP), Collaborative Modeling, Collaborative Modeling Evaluation (COME), Modeling Artifacts, Multi-criteria Decision Analysis*

INTRODUCTION

Collaborative Modeling (Renger et al., 2008; Rittgen, 2007), which is closely related to Group Model Building (Vennix, 1996), is a process that can enhance productivity in Information Systems Design and Business Process Re-engineering. During the collaborative effort of system development, stakeholders "move through a process in which they combine their expertise, their insights and their resources to bring them to bear for the task at hand" (de Vreede & Briggs, 2005, p. 1). The importance of involving different hierarchical level representatives in a (re-) engineering process is recognized by Dean et al. (1994). However, the emphasis in the bulk of the literature is on tools and techniques used by the stakeholders in order to achieve the desired model quality (*completeness and correctness*). Yet, it has been

DOI: 10.4018/jismd.2013010101

argued that model quality alone, especially its clarity and completeness, which are often emphasized, is no longer enough (Mendling & Recker, 2007). Following that observation, it is our contention that if we are concerned with the quality of the final model, we also need to evaluate other modeling artifacts that are used in, and produced during, the modeling session.

Rather than taking the end-products (models) to the so-called "modeling expert (s)," we advocate the evaluation of such models and other modeling artifacts - which include the modeling language, the modeling procedure and the support tool (Ssebuggwawo et al., 2010) - to be done by the collaborative modelers themselves. This essentially guarantees stakeholders' satisfaction if the evaluation of the models and the process is integrated within the modeling session. After all, it is their model and it is their process. This, however, is compounded by the fact that modelers posses different knowledge, skills, expertise and often lack the required competencies (Frederiks et al., 2005) which may not only affect the process of modeling, but also the evaluation of the modeling artifacts. Moreover, they often have different priorities and preferences about the modeling artifacts to be evaluated and their associated quality dimensions. One way of overcoming the limitations encountered during the modeling process and evaluation is to position the modeling process and evaluations within the communicative process (Hoppenbrouwers et al., 2005). This fits in well since the modeling process is collaborative in nature and exchanges between and among the modelers are expected and assumed to eventually lead to agreement and consensus about the final quality of the modeling artifacts.

Communication plays a vital and important role in system development, and in conceptual modeling (Veldhuijzen et al., 2004). The communicative process should render consensus and agreement transparent to the modelers. This is, however, not always the case since many stakeholders (with varying skills, expertise, knowledge, priorities, and preferences) are involved in system development. The heterogeneity of the group makes it hard for them to agree on each and every issue. Yet, agreement and consensus are key pillars in such an interactive and collaborative environment. For this to be achieved, participants need to engage in various types of conversation during the creation of *agreed models*. Such conversations involve *negotiation*, which results in *accepts, rejects, modifications*, etc. (Rittgen, 2007). This communicative, argumentative and negotiation process is vital for reaching agreement and consensus about the quality of the different modeling artifacts. Due to the differences in their knowledge stored in their mental models, skills, competencies and expertise, priorities and preferences, there is always some bias and subjectivity – a fact that makes the overall decision-making process subjective (Saaty, 2008b) which eventually overflows into the evaluations. This begs the question whether there exists (an) evaluation framework(s) that can help us evaluate the four modeling artifacts yet at the same time reduce the subjectivity and aggregate the modelers' priorities and preferences. We describe, in this paper, a framework that can help us achieve this. The major contribution of this paper is the COME framework that can be used by participants in the modeling effort to collaboratively evaluate the different modeling artifacts without guidance of a facilitator.

RELATED WORK

A number of frameworks have been developed to evaluate some of the artifacts that are used in, and produced during, a modeling session. Prominent among these is the Semiotic Quality (SEQUAL) framework (Krogstie et al., 2006). The SEQUAL framework is a versatile framework that is strongly rooted in the existing theory – the *semiotic theory* – and can thus "claim theoretical validity" (Rittgen, 2010, p. 2). It can be used to evaluate not only the end-products of the modeling effort – models, but also the modeling language. It may help determine quality with respect to new knowledge acquisition, new knowledge transfer, learning and level of agreement (Krogstie et al., 2006). It does not, however, bring out firmly and explicitly the evaluation of the modeling procedure and the support tool which, we believe, have an impact on the overall quality of the modeling process. Moreover, as rightly observed by Rittgen (2010), it does not state how the quality dimensions can be measured despite giving and suggesting a number of dimensions. The Quality of Modeling (QoMo) framework of Bommel et al. (2007) extends the SEQUAL framework by incorporating the rules and goals, the knowledge of the modelers, the activities modeled, modeling language, the domain modeled and the agreement between and among the modelers into the evaluation framework. It has, unfortunately, not been validated in practice. The COME framework we present in this paper is part of the ongoing effort to extend, apply and validate some of the concepts in the QoMo framework. In addition, the COME incorporates a Multi-criteria Decision Analysis (MCDA) technique (Guitouni & Martel, 1998) – notably the Analytic Hierarchy Process (AHP) of Saaty (1980) not only to evaluate, score, rate and rank the quality dimensions of the modeling artifacts, but also aggregate the priorities and preferences of the modelers.

The Guidelines of modeling (GoM) (Schuette et al., 1998) is a model-based quality framework that goes beyond the syntactical rules in model evaluation. The goal of GoM is to improve the quality of process models (*product quality*) and that of information modeling (*process quality*). To achieve this, it identifies six quality factors: *correctness, relevance, economic efficiency, clarity, comparability,* and *systematic design* which are further classified into basic guidelines (correctness, relevancy, economic efficiency) – meaning they are essential to determining the quality of the products and the process, and optional guidelines (clarity, comparability, systematic design) – meaning they are just additional or desirable features. Despite these guidelines, the GoM still lacks a sound theoretical methodology and provides little empirical proof (Recker, 2006). The Moody-Shanks framework (Moody & Shanks, 1994), which looks only at evaluation of one modeling artifact – the model, is one of the few frameworks that develop quality metrics and shows how they can be applied for model evaluation (Moody, 1998). Unfortunately, it fails to include, or offer guidelines and metrics about other modeling artifacts.

Although it has been stated that the subjectivity and bias that is carried along by the modelers in the evaluation of the modeling artifacts is based on their prior experiences – knowledge that is stored in their mental models – there are still other factors that influence their judgments in the evaluation of the modeling artifacts. Most of these are psychological while others are behavioural. Modelers' satisfaction is one of the affective factors that can best be used to capture most of these psychological and behavioural factors. Satisfaction connotes psychological factors such as attitudes/perceptions, beliefs, and intentions which can be explained and measured using the Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1975) and the Theory of Planned Behaviour (TPB) (Ajzen, 1991). There have been some attempts to evaluate (collaborative) modeling processes using some of these psychological factors – through satisfaction (Duivenvoorde et al., 2009) and a few frameworks include measurement of the participants' satisfaction with respect to the modeling process and outcomes or end-products (Briggs et al., 2003). Also, a few frameworks have been developed recently, especially in the area of Collaboration Engineering (CE), to study outcome factors that describe the effects of a collaborative modeling effort (Renger et al., 2008). Still, these do not wholly satisfy our requirements of integrating all the four artifacts in a collaborative modeling evaluation, collaboratively scoring the quality dimensions, aggregating the scores and synthesizing them to obtain the final quality score for any modeling artifact and/or modeling approach used.

Driven by the conviction that increased understanding of the communicative process – including the argumentative, negotiation and decision-making process will help develop better *support* for evaluating the modeling artifacts used in, and resulting from, the collaborative modeling effort, we feel the need for an approach that enables us to evaluate these modeling artifacts and makes visible how/why the evaluation renders the results good or bad. Building on the strengths of existing frameworks while avoiding the identified pitfalls, we present in the next sections a methodology that can be used to evaluate the modeling artifacts taking into account the subjectivity and bias of the modelers and we show how the evaluation framework can be used to score the different dimensions of the modeling artifacts and eventually select the modeling approach that satisfies the modelers' quality goals.

THE COME FRAMEWORK

The Collaborative Modeling Evaluation (COME) framework we present follows and extends the approach suggested by Pleiffer and Niehaves (2005) to evaluate the different artifacts used in, and produced during, the modeling process. Their approach follows a design science approach (Hevner, 2004) to identifying the different IS research artifacts and evaluating them. Because their framework employs the philosophical notions of structuralism, it still focuses mainly on the inner structure of the models and the evaluation of their quality. Although our approach extends their framework by evaluating a wider range of modeling artifacts involved in the modeling process, it also fundamentally differs from theirs in the way it scores the quality dimensions of the artifacts and the method used to evaluate the artifacts. In our case, we apply principles and concepts from the Analytic Hierarchy Process (AHP) (Saaty, 1980) to measure and evaluate the modeling process artifacts. The AHP is, essentially, a method for making complex decisions on the basis of subjective opinions by multiple stakeholders, a case that is prominently reflected in collaborative modeling. In our case the process renders the score for an individual modeling session which can then be compared with a similarly calculated score for another session. Given that variables between the sessions are sufficiently controlled, this enables well-founded judgment about which method works best for the situation at hand. The advantages of our evaluation framework and the AHP approach lie in advanced management of subjectivity, aggregation of individual priorities, and aggregation of preferences into group priorities and preferences. Also, the AHP helps the stakeholders reach consensus about their preferences and priorities.

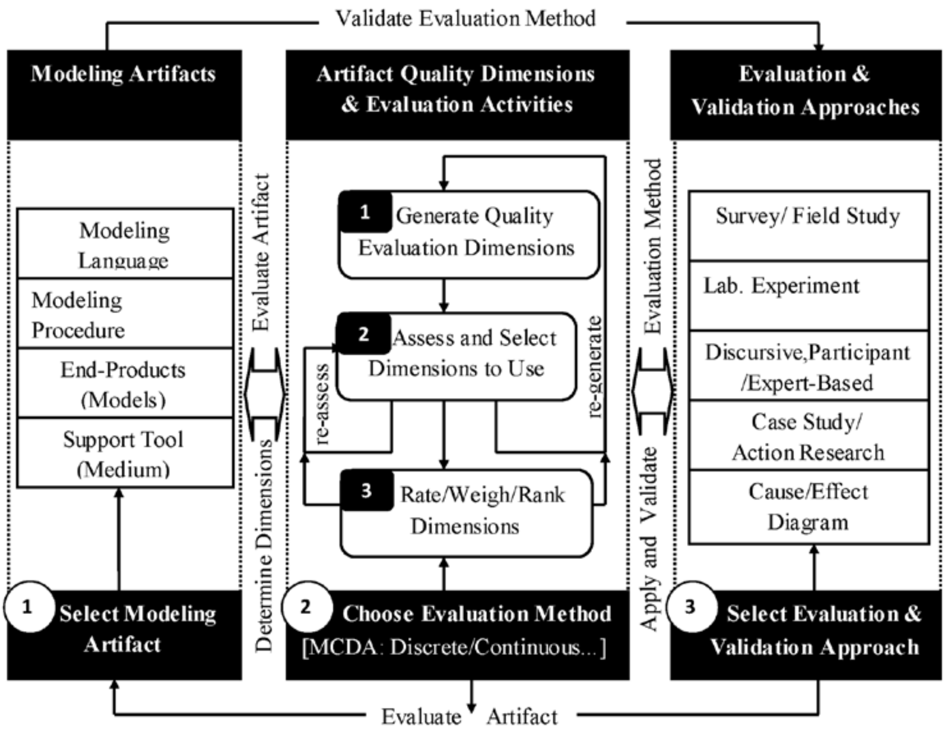
The COME framework can be used to evaluate the different modeling artifacts used in, and produced during, a collaborative modeling process. It is based on a conceptual model which describes three steps followed during the evaluation. This conceptual model is given in Figure 1. The model shows how quality dimensions (criteria) or factors associated with each modeling artifact are generated, selected and scored (rated, ranked and/or weighted) by the modelers within the evaluation process. A Multi-criteria Decision Analysis (MCDA) approach can be used to facilitate the decision making process in which the modelers reconcile their subjectivity, biases, priorities and preferences. Though the COME conceptual model in principle offers flexibility and freedom in selecting the MCDA approach to use, we single out one MCDA approach – the Analytic Hierarchy Process (AHP) – that is used to score the modeling artifacts dimensions, aggregate the modelers' priorities and preferences. Within the COME conceptual model we have three main steps:

1. *Selecting the modeling artifact(s) to evaluate.*
2. *Choosing the evaluation method to apply in the evaluation of the modeling artifact(s).*
3. *Choosing an evaluation and validation approach to evaluate and validate the evaluation methods and modeling artifacts.*

SELECTING THE MODELING ARTIFACTS

This step involves determining and selecting the modeling artifacts whose quality is to be measured and/or assessed during a collab-

Figure 1. COME conceptual model



orative modeling session. The four artifacts, whose dimensions are briefly described in the subsequent sections, include the following: 1) modeling language (ML) which refers to the syntactic or domain meta language that provides concepts (constructs) in which modelers define the problem, express and communicate the solution; 2) modeling procedure (MP) which details the processes, techniques, strategies or approaches (methods) of how the problem is defined and how the solution is reached; 3) end-products (EP) which are the final outcomes (models) of the communicative process that may have used the modeling language concepts and a modeling procedure to represent real world entities in the problem domain and solution domain; and 4) support-tool (ST) - (medium) which refers to either an electronic or non-electronic group support system that aids the communicative process, generation of the outcomes and/or evaluation of the outcomes. It should be noted that these modeling artifacts and their quality dimensions may be generated from the existing literature, e.g., from SEQUAL, GoM, Moody-Shanks frameworks, etc. The forward direction (right arrow) of the double-headed arrow between step 1 and step 2 means "determine the quality dimensions in step 2 of the identified modeling artifacts in step 1." Likewise, the backward arrow (left arrow) of the same double-headed arrow means "apply the identified quality dimensions in step 2 to evaluate the modeling artifacts in step 1."

CHOOSING THE EVALUATION METHOD

In step 2 we choose the method to evaluate the modeling artifacts. Within this step three activities take place. These activities are indicated by the following sub-steps: (1) generating quality dimensions (criteria, factors) which are the characteristics or features of the modeling artifacts upon which quality assessment will be done. These may come and/or are generated from those existing in the literature (e.g., from SEQUAL, GoM, Moody-Shanks frameworks, etc.), (2) assessing and selecting the dimensions to use (may involve narrowing the scope and grouping the dimensions), and, (2) rating, weighting and/or ranking the dimensions using an evaluation method. A Multi-criteria Decision Analysis (MCDA) method can be used for sub-step (2) and sub-step (3) as well as determining a measurable and quantifiable quality of the dimensions. This is explained in the AHP approach described later in the paper. Our choice for the AHP is prompted by its ability to reduce the subjectivity or bias associated with the individual judgments when computing and aggregating the individual and group priorities. Sub-steps 2 and 3, as indicated by the internal (uni-directional) arrows in step 2, involve either a re-assessment or a re-generation of the quality dimensions. The rated, weighted and/or ranked dimensions are used to evaluate the modeling artifact(s) as shown by the double-headed arrow between steps 1 and 2. The forward direction (right arrow) of the double-headed arrow between step 2 and step 3 means "apply the evaluation method in step 2 in the evaluation and validation approach in step 3." Likewise, the backward arrow (left arrow) of the same double-headed arrow means "validate using step 3 the evaluation method in step 2."

SELECTING THE EVALUATION AND VALIDATION APPROACH

This step involves selecting an evaluation and validation approach for the evaluation method and the modeling artifacts. This means that the evaluation method in step 2 is also evaluated to determine its appropriateness for evaluating the dimensions and the modeling artifacts and it is then validated with an appropriate approach. The work of Siau and Rossi (1998) is an excellent survey on the literature about evaluation approaches for modeling methods. These approaches, although given for the evaluation of IS methods, can easily be tailored for the evaluation of the evaluation method used and we are concerned here with the application of only empirical approaches as opposed to non-empirical approaches. Since we are looking at the evaluation process within the whole communicative process, we follow the discursive, participant/IT-based approach. In this approach, different persons with their subjective experiences are brought together with a goal of engaging in a dialogue to reach a more objective view of and valuation of the some facts (Wolff & Frank, 2005). It should be noted that the evaluation of the modeling artifacts through the quality dimensions and the validation of the evaluation method using any selected evaluation and validation approach is cyclic. This is indicated by the outer arrows around the three boxes of the three steps.

THE EVALUATION METHOD: THE AHP APPROACH

This section looks at an MCDA method that we use in sub-steps (1), (2), and (3) of step 2 –selecting the evaluation method – of the COME framework, in Figure 1, to assign quality scores to the modeling artifacts, rank and weigh the quality dimensions of the artifacts and finally evaluate and determine the modeling artifacts that meet the modelers' quality goals. This evaluation method can be used during the group decision-making and/or negotiation process to reconcile the modelers' subjective opinions, views, priorities and judgments (Ssebuggwawo et al., 2009). The MCDA that is selected is the Analytic Hierarchy Process (AHP) developed by Saaty (1980). The AHP is a complex multi-criteria approach for aiding decision making. According to Saaty (2008a, 2008b), the AHP is a flexible tool for dealing with both qualitative and quantitative multi-criteria decision. It integrates different evaluative measures into an overall score for ranking, evaluating and selecting alternatives. The main feature of the AHP is that it is based on pair-wise comparisons and has a rich mathematical foundation. It should be pointed out from the outset that the alternatives, in this research, could refer to quality dimensions when selection is to be performed, modeling artifacts when evaluation and selection is to be done with respect to the quality dimensions, or collaborative modeling approaches when more than one modeling approach is used during a modeling session. Evaluation and selection in this case is with respect to the modeling artifacts. The Analytic Hierarchy Process consists of mainly three main steps:

1. *Structural decomposition,*
2. *Comparative judgment,*
3. *Synthesizing.*

These steps are divided into a number of steps which are summarized in Figure 2 (Nga et al., 2005).

STRUCTURAL DECOMPOSITION

Problem Identification

This step involves identifying the unstructured problem to solve. It could be an evaluation, selection, or a location/allocation problem. Problem identification means also identifying the characteristics or features of the problem which can be used in decision making. These could be criteria, sub-criteria, attributes and alternatives. By weighting the different quality attributes, sub-criteria and criteria for each modeling artifact, modelers are able to assign and determine their priorities and preferences.

Hierarchy Construction

This step involves decomposing the problem into a hierarchical structural with distinctive levels. The structure can be obtained using decision-tree like diagrams. The topmost level, in the hierarchy, is the goal level followed by the criteria level, which is also followed by the sub-criteria, sub-sub-criteria and attributes levels (if any) up to the lowest level which consists of alternatives. Figure 3 is an example of a hierarchical structure in which a problem is decomposed into a goal, criteria, sub criteria up to alternatives, each on a different level of the hierarchy.

Comparison Scale

In the comparison step, each of the elements is assigned and ranked using a nine (1 - 9) point scale (Saaty, 1980), in a questionnaire-like instrument, see Figure 4, in order to determine their relative importance to each other.

Figure 2. The analytic hierarchy process (AHP) steps (adapted from Ngai et al., 2005)

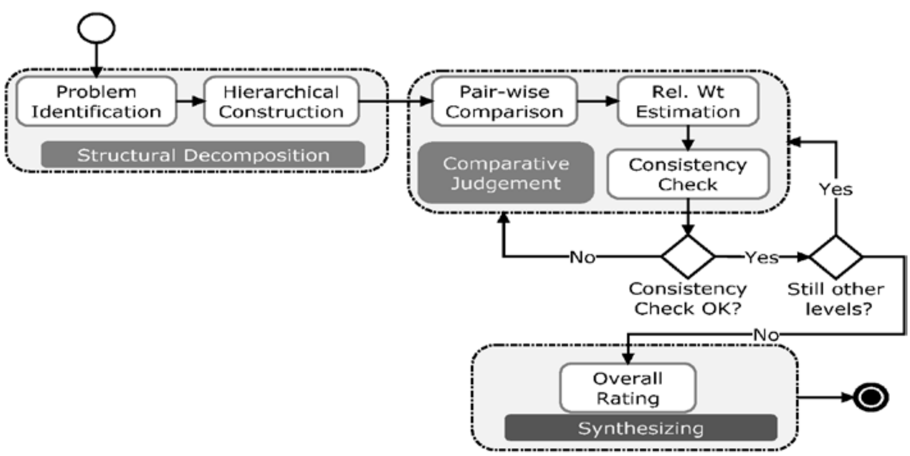
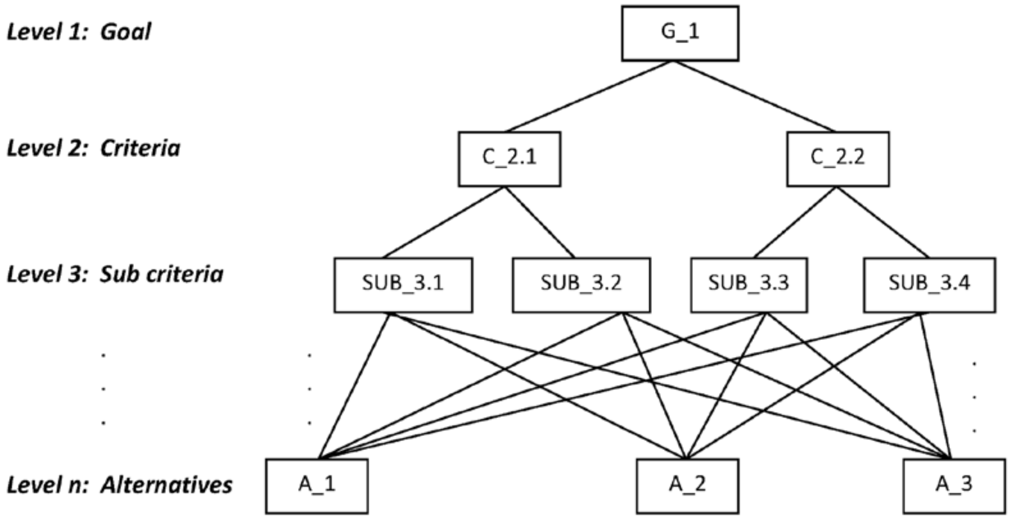


Figure 3. An example of an AHP hierarchical structure



Forming a Comparative Matrix

The outcome of the comparative judgment step is a comparative matrix the entries of which are the comparison values between the *ith* row and the *jth* column indicating the relative importance (from the fundamental scale of) one criterion over another. This comparison value gives the importance of the row's criterion relative to the column's criterion.

Let $\boldsymbol{A} = \left(a_{ij}\right)$ be an $n x n$ comparative (judgment) matrix and let a_{ij} be its entry. Then

$$\left(a_{ij}\right) = \begin{cases} 1/a_{ji}, & i \neq j \\ 1, & i = j \end{cases} \tag{1}$$

This means that the elements, a_{ii}, for all *i*, on the principal diagonal are all equal to 1. The purpose of the pair-wise comparison, according to Saaty (1980) is to determine the (priority) vector, $\boldsymbol{w}$, with weights $w_1, w_2 \ldots, w_n$ which represent the expert's relative opinion/judgment for the criteria, sub-criteria or attributes (Ssebuggwawo et al., 2009). This priority vector is shown in Equation 2:

$$\boldsymbol{w} = (w_1, w_2 \ldots, w_n)^T, \quad where\, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \tag{2}$$

The relation of the weights w_i to matrix **A** is given in Equation 3:

$$a_{ij} = w_i / w_j, \quad 1 \leq i, j \leq n \tag{3}$$

Saaty (1980, 2008a) observes that matrix $\boldsymbol{A} = (a_{ij}),\ i, j \in \{1, \ldots, n\}$ where the entries, a_{ij}, are given by Equation 2 are all positive and is called the *reciprocal matrix*, since it satisfies the property given in Equation 4:

$$a_{ji} = 1/a_{ij} = w_j/w_i, j > i \quad or\ a_{ij} = 1/a_{ji} = w_i/w_j, i < j \tag{4}$$

Figure 4. AHP questionnaire evaluation instrument

Element	← *Element A's Scale* →								← *Element B's Scale* →									*Element*
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
A_1	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	B_1
A_2	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	B_2
. . .																		. . .
A_n	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	B_n

Relative Weights Estimation: Eigenvector Method

There are a number of methods for computing the (priority) vector of the relative weights and aggregating individual and group judgments or priorities. The most popular aggregation methods are Aggregation of Individual Judgments (AIJ) and Aggregation of Individual Priorities (AIP) (Escobar & Moreno-Jimenez, 2007). For prioritization, the right Eigenvector Method (EGVM) and the Row Geometric Mean Method (RGMM) are the most popular. We prefer to use EGVM to show how the relative weights are computed because of its simplicity and transparency. The relative weights of all the attributes are computed from the eigenvalue problem, which is a system of homogeneous equations, of the form shown in Equation 5:

$$\boldsymbol{A}'\boldsymbol{w} = \lambda_{max}\boldsymbol{w} \; or \; \left(\lambda_{max}\boldsymbol{I} - \boldsymbol{A}'\right)\boldsymbol{w} = 0 \quad (5)$$

where λ_{max} is the largest eigenvalue of $\boldsymbol{A}$, called the principal eigenvalue of $\boldsymbol{A}'$ and **I** is the usual identity or unit matrix (Saaty, 1980). The importance of λ_{max} is to reduce the inconsistencies in the modelers' judgments and is thus the measure we use to control the modelers' subjectivity in their evaluations of the quality dimensions of the modeling artifacts (Ssebuggwawo et al., 2009).

Consistency Check

To check whether matrix judgments (decisions) are consistent, we need to check the consistency of the comparative matrices at each level of the hierarchy. This is done via the Consistency Index (C.I) and the Consistency Ratio (C.R) (Saaty, 1980, 2008a, 2008b) calculated, respectively, using Equation 6:

$$C.I = \left(\ddot{e}_{max} - n\right) / \left(n - 1\right), \; C.R = C.I \; / \; R.I \quad (6)$$

where $R.I$ is the random index (the average consistency index) calculated as the average of a randomly generated pair-wise matrix of the same order as $\boldsymbol{A}$, often computed from a table of random indices (Escobar & Moreno-Jimenez, 2007). Saaty (2008a) gives some threshold values for the consistency ratio (C.R) above which matrix $\boldsymbol{A}$, is of insufficient consistency and thus the evaluations or judgments made by the evaluators (collaborative modelers) are assumed to be unacceptable. These threshold values are given in Equation 7:

$$C.R \leq \begin{cases} 0.05, & n = 3 \\ 0.08, & n = 4 \\ 0.10, & n > 4 \end{cases} \quad (7)$$

SYNTHESIZING: OVERALL RATING AND RANKING

This step consists of determining overall rating and ranking of alternatives whose priorities may be given as normalized or idealized priorities. It determines the overall priority (preference) rating of the alternatives by aggregating the relative weights of the criteria. The Synthesizing step has the following sub steps (Saaty, 2008a)

1. *Synthesis of the weight of each criterion with respect to the goal.*
2. *Synthesis of the comparisons to get the (local) priority of alternatives with respect to each criterion.*
3. *Multiplication of the local priorities of each alternative by the local priorities of each criterion and summing up the local priority products to obtain the overall (global) priority for each alternative.*

To execute step (3), suppose we have got *m* alternatives. Let w'_{ik} be the local priority for the *kth* alternative, A_k, for $k \in \{1,2,\ldots,m\}$, with respect to the *ith* criterion, C_i. Let w'_i be the local priority of C_i with respect to the goal, G. Then the alternative global priority w'_{A_k}, of alternative A_k with respect to all local priorities of the criteria is given by Equation 8**:**

$$w'_{A_k} = \sum_{i=1}^{n} w'_{ik} w'_i, \quad w'_{A_k} > 0, \quad \sum_{k=0}^{m} w'_{A_k} = 1 \quad (8)$$

The computations in Equation 6 of the alternative global priorities (w'_{A_k}) are aided by arranging the alternatives (A_k), criteria (C_i .), criteria local priorities with respect to the goal (w'_i), alternatives local priorities with respect to the criteria (w'_{ik}) as shown in Figure 5.

Idealized Priorities

An alternative way of expressing overall (global) priorities for alternatives, according to Saaty (2008b), is to use an idealized form. Priorities for the ideal mode are obtained by dividing each priority by the largest one. Let w''_{A_k} be the idealized overall priority for alternative *k*, $k \in \{1,2,\ldots,m\}$. Then the idealized priority is computed from Equation 7:

$$w''_{A_k} = w'_{A_k} / \max\{w'_{A_k}\}, \quad k \in \{1,2,\ldots,m\} \quad (9)$$

APPLICATION OF THE COME FRAMEWORK

In this section we look at how the COME framework can be used. We apply it to a modeling session experiment which was carried out to validate the concepts of the framework. This modeling experiment involved IT experts in a big Telecommunication Company in Uganda. The goal of the modeling session experiment was two-fold: (i) we wanted to validate the COME framework which, originally, had been tried in numerous exploratory experiments that involved mainly university students both in the Netherlands and Uganda and (ii) we wanted to validate two research instruments (one AHP-based, the other SEQUAL/QoMo, TRA/TPB-based) that were designed and had been tested mainly with students in exploratory modeling session experiments. It should be noted that in this paper we report on only the AHP-based research instrument due to our selection of the evaluation method which is the AHP. The SEQUAL/QoMo, TRA/TPB-based research instrument will be reported on elsewhere.

Figure 5. Aid to computation of global priorities

Criteria local priorities w.r.t. goal

Altern.\Criteria	C_1	C_2	C_i	.....	C_n	Alternative Global Priority
	w'_1	w'_2	w'_i	.....	w'_n	
A_1	w'_{11}	w'_{21}	w'_{i1}	.....	w'_{n1}	w'_{A_1}
A_2	w'_{12}	w'_{22}	w'_{i2}	.....	w'_{n2}	w'_{A_2}
. A_k .	. w'_{1k} .	. w'_{2k} .	. w'_{ik} .	. . .	. w'_{nk} .	. w'_{A_k} .
A_m	w'_{1m}	w'_{2m}	w'_{im}	.....	w'_{nm}	w'_{A_m}

Alternative local priorities w.r.t. criteria

MODELING SESSION EXPERIMENTAL SET UP AND THE TASK

Selection of Subjects

The participants that took part in the modeling experiment came mainly from the IT Department of the organization. We preferred persons with some background in computing, although not necessarily with modeling skills to take part in the modeling experiment. These participants had varied background computing e.g., in databases, web-design and hosting, networking, programming and modeling in entity-relationships (E-R), Object-role Modeling (ORM) and Unified Modeling Language (UML).

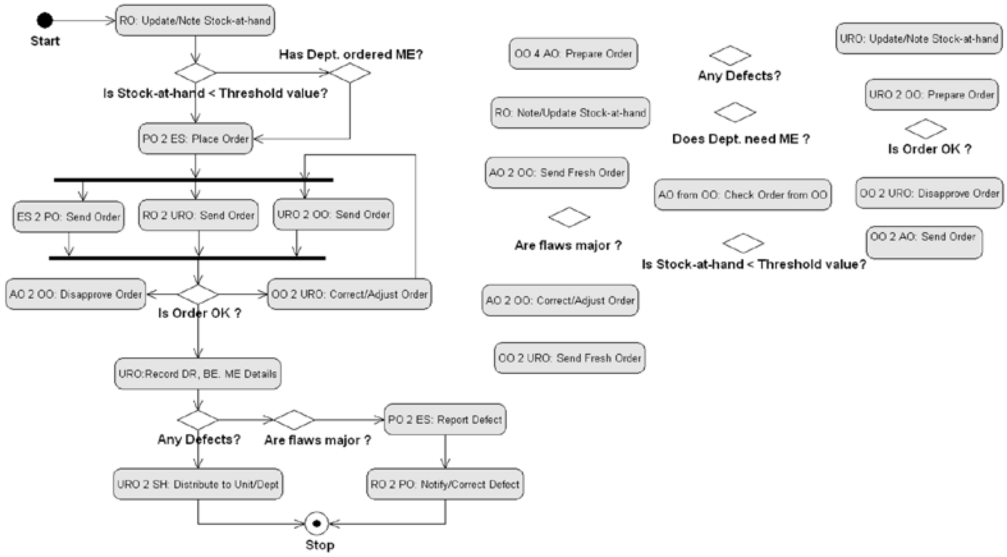
Modeling Task

The main task that was given to the participants was about developing a model for the University Teaching Hospital's Pharmacy and Medical Equipment Department showing the procurement process of medical drugs and equipment and distributing these to the different wards and departments of the University Teaching Hospital. This task was chosen on the basis that since procurement and distribution of drugs and equipment is not different from that of IT products, it would be found interesting by the participants and thus it would be easier for them to brainstorm and generate ideas about the problem being addressed. Participants were asked to first generate, individually, as many ideas in the idea-generation task (Delbecq & Van de Ven, 1975) about the problem. Figure 6 is a snapshot of part of the models that were generated.

Modeling Session Experiment

The modeling session experiment had two phases which in total lasted for three hours. The first phase required modelers to generate a model of a case that was given to them using a Unified Modeling Language (UML)-based environment embedded within the Collaborative Modeling Architecture (COMA) tool of Rittgen (2008). Prior to the actual modeling session, modelers were introduced to the inner workings of the modeling tool and working definitions of what is meant by the modeling language, the modeling procedure, the end-product (model) and support tool were introduced to them. This role was played by the modeling session facilitator at the beginning of the modeling session. Our choice for the COMA tool was based on a number of factors prominent among which are: its simplicity, its integration of two of the modeling artifacts: the modeling language (UML) and the support tool, which is COMA itself. The third modeling artifact – the end-product (model) was to be developed using the

Figure 6. Snapshot of the group model and ideas generated during the intellective and idea generation phases

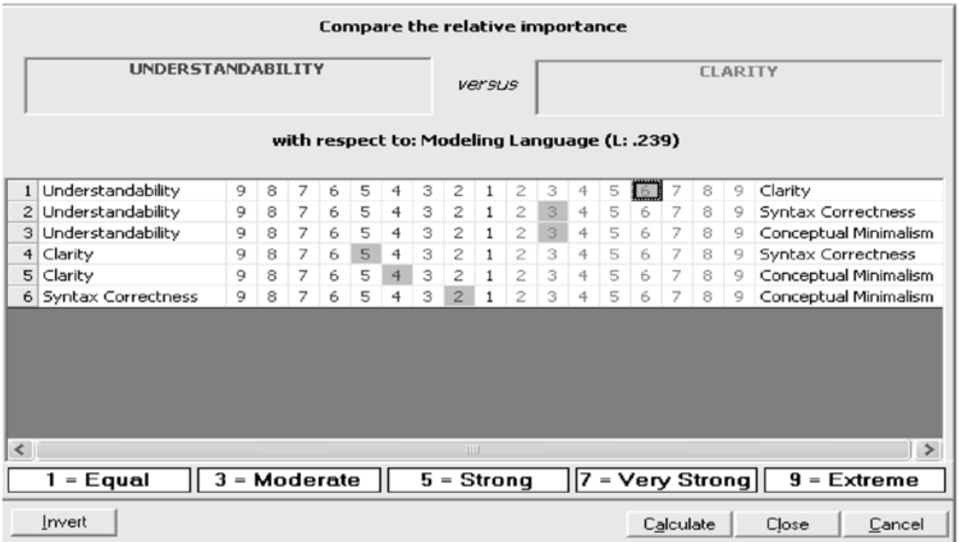


COMA environment while the fourth modeling artifact – the modeling procedure – was left to the participants to determine and follow. At the end of the first phase, modelers were given a paper-based pre-survey instrument to evaluate the modeling artifacts that had been introduced to them. This questionnaire is given in the Appendix. The goal of this questionnaire was to get the individual scores to the modeling artifact quality dimensions. This questionnaire used the AHP fundamental scale of Saaty (1980). The second phase of the modeling session experiment required participants to collectively use a post-survey research instrument, which is a computer-based evaluation tool that employs the AHP methodology implemented in Expert-Choice Software (Expert-Choice, 2011) to evaluate the modeling artifacts used in, and produced during, the modeling session in phase one. It should be noted that we used two types of questionnaires a paper-based pre-survey instrument and a post-survey (electronic) questionnaire. The pre-survey questionnaire is what is presented in the Appendix while the post-survey questionnaire is given in Figure 7. The paper-based questionnaire for all the four modeling artifacts is 6 pages long. Only part of this is presented in the Appendix for one of the modeling artifacts — the modeling language — since the others can similarly be fitted within this structure using Figure 4 and the quality dimensions in Figure 8 for each of the modeling artifacts.

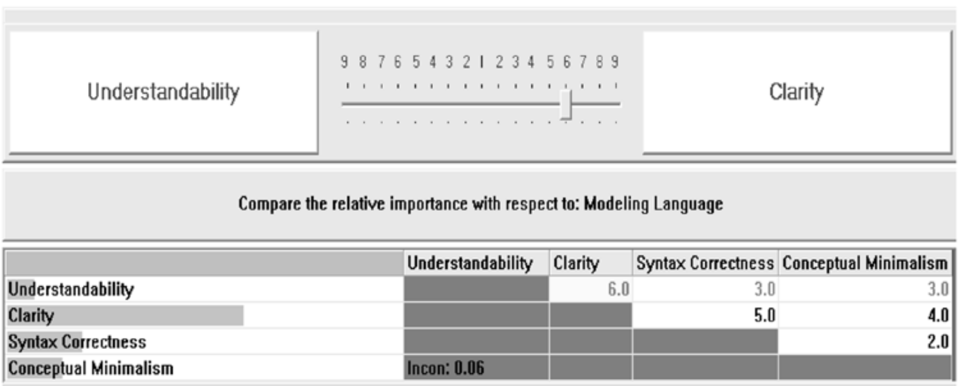
MODELING ARTIFACT SELECTION

Step 1 of the COME framework conceptual model in Figure 1 requires selection of the modeling artifacts to be used in the evaluation process. Although, the COME framework is generic in that it allows selection of one or more modeling artifacts to evaluate and allows selection of any other evaluation method other an MCDA-based method and also gives freedom in the evaluation approach selection, we already argued our case that in collaborative modeling all four artifacts have an impact on the effectiveness and efficiency of the modeling session, see also (Ssebuggwawo et al., 2009, 2010). We, therefore, selected all the

Figure 7. Expert Choice AHP-based questionnaire (a) and comparative matrix (b)



(a) Expert-choice AHP questionnaire



(b) Comparative matrix

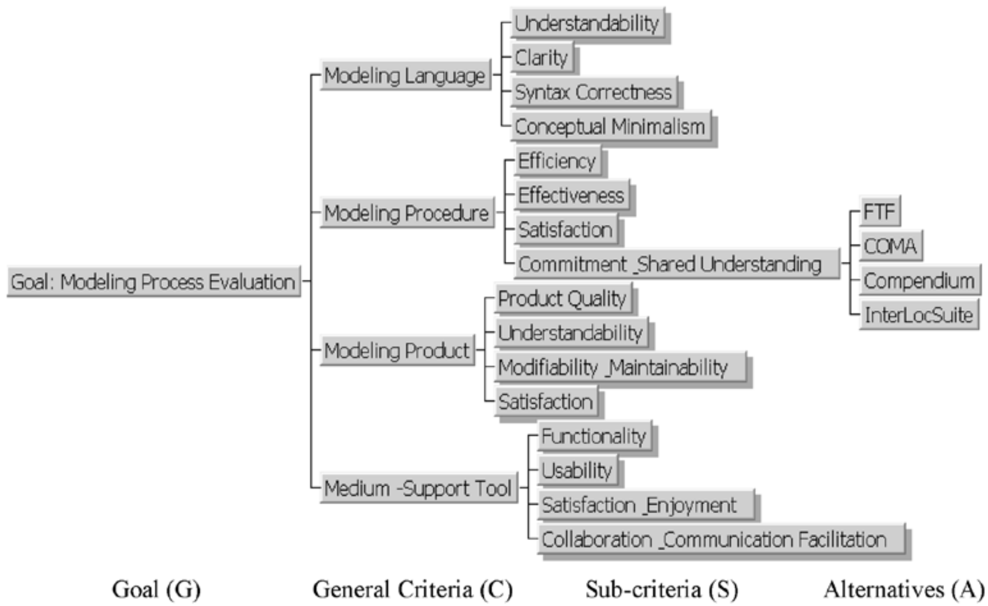
four modeling artifacts: modeling language, modeling procedure, end-product (model) and support tool.

CHOOSING THE EVALUATION METHOD

Step 2 of the COME framework conceptual model requires choosing an evaluation method to use in the evaluation of the modeling artifacts. We selected an MCDA method – the AHP method – due to our goal of trying to find an appropriate technique to help the modelers score (rate/rank/weigh) the different modeling artifact quality dimensions (criteria or factors) and a method that can help us aggregate the individual and group scores (and, thus priorities and preferences) given to the different quality dimensions of the modeling artifacts.

Within Step 2 of the COME framework, sub-step (1) requires generation of the dimen-

Figure 8. Structural decomposition of modeling process evaluation



sions of the modeling artifact(s) to be used in the evaluation of that particular modeling artifact. This is normally done during either a brain-storming session where participants, during a brainstorming or an idea-generation task (Delbecq & Van de Ven, 1975), generate the dimensions in a freewheeling fashion or a literature survey is done to identify relevant quality dimensions for each artifact (Pfeiffer & Niehaves, 2005).

Each participant was thus, initially, given a piece of paper on which to write down the dimensions he/she felt were relevant for evaluating each of the identified modeling artifacts discouraging discussion and criticism at this stage. Sub-step (2) requires assessing and selecting the dimensions to finally use. Through the modeling session facilitator, each participant presented their quality dimensions in a round-robin fashion and other members were allowed to discuss them, thus allowing group interaction – mainly negotiation - and through Delbecq and Van de Ven's (1975) Nominal Group Technique (NGT) these were "subjectively" ranked, voted on and agreed upon – thus resulting into group-accepted quality dimensions. It should be noted that this procedure, though done democratically through a voting process, does not eliminate or reduce the subjectivity or bias still inherent in the quality dimensions that are generated, since many times many people may tend to just follow, simply give-in to or go-by, what the majority has proposed. Through guidance from the modeling session facilitator, the generated and subjectively ranked quality dimensions were categorized and grouped into some of the quality categories that exist in the literature. These categories for the four modeling artifacts are given in Table 1, and are defined and explained in Ssebuggwawo et al. (2009, 2010).

Due to the subjective nature of the evaluation and ranking of the generated quality dimensions, the participants had to evaluate the groupings of these quality dimensions as a group using sub-step 3 of step 2 in the COME framework in Figure 1. This involved using the AHP-based evaluations using pair-wise comparisons of the quality dimensions.

Table 1. Groupings of modeling artifact quality dimensions

Modeling Artifact	Dimension Groupings
Modeling Language (ML)	Understandability, Clarity, Syntax Correctness, Conceptual Minimalism
Modeling Procedure (MP)	Efficiency, Effectiveness, Satisfaction, Commitment & Shared Understanding
End-product (EP)	Product Quality, Understandability, Modifiability & Maintainability, Satisfaction
Support Tool/Medium (ST)	Functionality, Usability, Satisfaction & Enjoyment, Collaboration & Communication Facilitation

SELECTION OF THE EVALUATION/VALIDATION APPROACH

The COME framework shows that an evaluation and validation approach has to be selected in step 3 in Figure 1. We selected the discursive, participant, expert-based approach for the evaluation and validation. In this approach, different persons with their subjective experiences are brought together with a goal of engaging in a dialogue to reach a "more objective view and valuation of some facts" (Wolff & Frank, 2005).

RESULTS AND DISCUSSION

This section gives and discusses some of the sample results from the modeling session experiment. We use these results to explain the main concepts discussed, especially, from the AHP evaluation method, since this is the method that is at the center of the COME evaluation framework presented. The structural decomposition step in Figure 2 with its two sub-steps: problem identification and hierarchical construction in Figure 3 is shown in Figure 8.

It should be noted that the alternatives are shown only for one of the secondary criterion (commitment and shared understanding), but they do exist for all the other secondary criteria as well. This is done to avoid cluttering the diagram. We also emphasize that the problem being looked at in this paper is an evaluation problem of the collaborative modeling process based on the four modeling artifacts. Our goal is to develop an evaluation framework or framework that can help in the scoring of the different attributes. Therefore, two collaborative modeling approaches (CMA) which constitute the alternatives were paid much attention to. These were the face-to-face (FTF) and the collaborative modeling architecture (COMA). Face-to-Face was used during the communicative process were modelers negotiated and agreed on the final score as a group using the AHP Expert-choice software while the COMA tool was mainly used to develop the models. Compedium and InterLoc Suite were used as dummy alternatives that could facilitate idea generation/issue building and for synchronous conversational dialog exchanges.

Table 2 shows results of the comparative judgment step with its three sub-steps: pair-wise comparison, relative weight estimation and consistency checking.

There are a few observations about the comparative matrix shown in Table 2. First, it should be noted that these scores are obtained from the comparative matrix in Figure 7 (b) and Equation 4 is used to complete the (reciprocal) pair-wise comparative matrix. Quality dimensions along the matrix (in the first column of Table 3) are given numbers (in brackets) which are also repeated on top of the matrix. This is done for convenience, otherwise the quality dimensions themselves would be repeated on top as shown in Figure 7 (b). The group scores (6, 3, 3, 5, 4, 2) that were given by participants in the collaborative modeling

Table 2. Pair-wise comparative matrix and priority vector of the modeling language (ML)

Modeling Language (ML)	(1)	(2)	(3)	(4)	Priorities Vector(w)
Understandability (1)	1	**1/6**	**1/3**	**1/3**	0.067
Clarity (2)	6	1	5	4	0.603
Syntax Correctness (3)	3	1/5	1	2	0.190
Conceptual Minimalism (4)	3	1/4	½	1	0.141
λ_{max} = 4.168 C.I = 0.056 C.R = 0.063					

session are given in the upper (grey-coloured) part of the comparative matrix (above the main diagonal of 1's) with the reciprocals of 6, 3 and 3 being entered since these scored appeared in the right-half of the Expert-choice AHP questionnaire in Figure 7(a). Note that elements on the left are those along the pair-wise comparative matrix while those on the right are those on top of the pair-wise comparative matrix in Table 2. The priority vector w in Equation 2 which is computed from the system of homogeneous equations given in Equation 5 is shown in the last column of Table 2. These priority values indicate that in trying to measure and evaluate the quality of the modeling language, modelers attach more priority to clarity of the modeling language followed by its syntax correctness, conceptual minimalism and understandability is of less priority.

To check that these values are not subjective and biased, i.e., the scores given by modelers shown in the reciprocal comparative matrix are consistent, we use Equations 5 and 6 with R.I = 0.89 for a matrix of order (n =) 4 (Saaty, 2008a, 2008b), to compute the largest eigenvalue ($\ddot{e}_{max}$), the consistency index (C.I) and consistency ratio (C.R). These values are shown at the bottom of Table 2. As can be seen from these values, the consistency ratio (C.R) is less than 0.08, which is the threshold value given in Equation 7 which the matrix and the hence the evaluations would be inconsistent and thus biased or subjective. This means that the evaluations are consistent and they are, therefore, acceptable. Tables 3(a) through 3(d) give the results for the other modeling artifacts, interpreted similarly.

Table 4 gives the synthesized final global priorities for the alternatives (collaborative modeling approaches-CMAs) which are computed using Equation 8 and the Figure 5. Both normalized and idealized priorities are given in the last column of Table 4. It should be noted that our research looked at only CMA1 (FTF) and CMA2 (COMA) while CMA 3 and CMA4 were used as dummy approaches. Interpretation of these results shows that the face-to-face (FTF) approach is judged a better approach than the COMA approach, i.e., CMA2 is 56.5% as good as the CMA1 approach.

Use of the Scores Calculated and the Results of the Evaluation

A number of observations can be made about the calculated scores shown in Table 2, Tables 3(a) through 3(d), and Table 4. The last column in each of these tables is the most interesting since it gives calculated priorities and/or preferences using the assigned scores to the quality dimensions of the modeling artifacts. These calculated priorities are used to determine the level of satisfaction about the quality by the modelers and which of the quality dimensions and/or modeling artifacts meets their quality goals. The higher the value of the calculated priority, the higher the satisfaction with the quality of the dimensions and/or modeling artifacts. This means that a dimension and/or modeling quality with a higher priority value is preferred and satisfies the quality goals of the modelers. Table

Table 3. Pair-wise comparative matrices and priority vectors of the modeling artifacts

(a) Modeling Procedure (MP)	(1)	(2)	(3)	(4)	Priorities vector(w)
Efficiency (1)	1	1	1	1	0.241
Effectiveness (2)	1	1	1	1	0.241
Satisfaction (3)	1	1	1	3	0.331
Commit. & Shared Understanding (4)	1	1	1/3	1	0.188
λ_{max} = 4.154 C.I = 0.051 C.R = 0.057					
(b) End-Product (EP)	**(1)**	**(2)**	**(3)**	**(4)**	**Priorities vector(w)**
Product Quality (1)	1	1/2	1/2	1/2	0.136
Understandability (2)	2	1	1/2	1	0.237
Modifiability & Maintainability (3)	2	2	1	1/2	0.287
Satisfaction (4)	2	1	2	1	0.340
λ_{max} = 4.186 C.I = 0.062 C.R = 0.070					
(c) Support Tool /Medium (ST)	**(1)**	**(2)**	**(3)**	**(4)**	**Priorities vector(w)**
Functionality (1)	1	1	2	1	0.288
Usability (2)	1	1	2	2	0.330
Satisfaction & Enjoyment (3)	1/2	1/2	1	2	0.207
Collaboration & Commun. Facilit. (4)	1	1/2	1/2	1	0.175
λ_{max} = 4.186 C.I = 0.062 C.R = 0.070					
(d) Modeling Process Evaluation -Goal	**(1)**	**(2)**	**(3)**	**(4)**	**Priorities vector(w)**
Modeling Language (ML) (1)	1	1	1	1	0.288
Modeling Procedure (MP) (2)	1	1	1/2	1	0.330
End-Product (EP) (3)	1	2	1	4	0.207
ST (4)	1	1	1/4	1	0.175
λ_{max} = 4.186 C.I = 0.062 C.R = 0.070					

3(d) and Table 4 give, respectively, the results of the evaluation for the modeling artifacts and the modeling approaches. They can be used to determine which of the modeling artifacts is of better quality. For example, in Table 3(d) the modeling procedure (MP) is of better quality than the modeling language (ML), which is also of better quality than the end-product (EP). The support-tool (ST) is of least quality. From Table 4, as already argued, the face-to-face (FTF) approach is judged a better approach than the

Table 4. Final global priorities with respect modeling process evaluation goal

Criteria Alt.	*Modeling Language (0.239)*	*Modeling Procedure (0.191)*	*End Products (0.404)*	*Support Tool (0.167)*	*Alt. Global Priorities (Normalized) (Idealized)*	
CMA1	0.207	0.437	0.318	0.314	0.310	1.000
CMA2	0.160	0.178	0.162	0.219	0.175	0.565
CMA3	0.367	0.157	0.323	0.235	0.290	0.935
CMA4	0.266	0.228	0.197	0.231	0.225	0.726

Key: CMA1: FTF, CMA2: COMA, CMA3: Compendium, CMA4: InterLoc Suite

COMA approach, i.e., CMA2 is 56.5% as good as the CMA1 approach.

Experiences, Benefits, and Consequences of Using the COME Framework

Although we could not apply the Theory of Reasoned Action/Theory of Planned Behaviour (TRA/TPB) models in the modeling session reported about to capture the perceptions of the modelers and their intention to use the evaluation approach in future, these were assessed through a post-survey interview where modelers were asked about their experiences with using the evaluation approach presented to them. Most of the participants found the method easy to understand and use. They enjoyed using both the modeling tool – COMA and the AHP's Expert-choice tool. However, selection of the modeling artifacts and generation of the quality dimensions were the difficult parts and most of them observed that it would not be possible to select the artifacts or generate the required dimensions without guidance from the modeling session facilitator. This may be not a surprising observation since most of them, although had background in computing as observed in the "Selection of Subjects" section, had never been involved in decision analysis and evaluations.

From the experiences of using the COME framework in the modeling experiment, and from the results obtained, we can draw the following benefits and/or consequences: 1) the COME framework integrates all the four modeling artifacts in the evaluation process which we feel have an impact on the overall quality of the modeling process and its success. It should be noted, however, that it is still possible to evaluate any of these modeling artifacts at any time, if one so wishes. 2) the COME framework provides a mechanism for developing and generating quality dimensions for the modeling artifacts and metrics for scoring, weighting and/or ranking the modeling artifacts and their quality dimensions, 3) it is possible to aggregate both the individual and group scores to obtain the final score, 4) evaluation of the modeling artifacts can be done collaboratively by the modelers themselves through the COME framework and their subjectivity or bias (inconsistency judgment) is then reduced/minimized or eliminated through the Multi-criteria Decision Analysis (MCDA) techniques such as the AHP approach, and 5) the COME framework can be used to determine the most effective modeling approach for collaborative modeling by synthesizing the priorities as shown in Table 4.

CONCLUSION AND FURTHER RESEARCH

We have presented and illustrated a research approach aimed at evaluating not only the end-products (models) developed from a collaborative modeling session, but also capable of bringing on board other modeling artifacts (the modeling language, modeling procedure, and support tool) that are used in, and produced from, a collaborative modeling session. We presented a conceptual framework – the Collaborative Modeling Evaluation (COME) – with three main steps: 1) Select modeling artifact to evaluate, 2) choose evaluation method with three sub-steps, and 3) select evaluation and validation method. A methodological approach, embedded within the COME framework which is based on the Analytic Hierarchy Process (AHP) that can help modelers not only score and determine the quality dimensions, but can also be used to aggregate their priorities and preferences was also presented. The AHP approach has three major steps: 1) structural decomposition, 2) comparative analysis, and, 3) synthesizing. We have shown that use of the COME framework based on a multi-criteria methodology like the AHP, can reduce the subjectivity that is inherent in the modelers' evaluations. We analyzed an actual collaborative modeling session. Findings were also presented, answering our research questions within scope of the modeling experiment. We also used our findings to perform a partial validation of the COME framework, and thus demonstrated its applicability for evaluation purposes.

One major conclusion about the COME framework and the evaluations for collaborative modeling is its use by the modelers themselves to score the different modeling artifacts used in, and produced from, the modeling effort. Through the COME framework modelers can score, weigh, rate and/or rank the artifacts and determine which one satisfies their quality goals. The major contribution of this paper is the COME framework that can be used by participants in the modeling effort to collaboratively evaluate the different modeling artifacts without guidance of a facilitator. The COME framework can be used to evaluate not only the models – which are the solutions of the modeling task, but also the artifacts, such as the modeling language, the modeling procedure and support too that are used to generate or develop these models. Any task or problem that can be solved jointly using a modeling language to generate models or using a well-structured modeling procedure with the use support of tool is a candidate for the COME framework. Such problems include, but are not limited to, business process modeling or re-engineering problems, enterprise modeling, enterprise engineering or enterprise architecture problems, etc.

We do not claim that our approach is definitive and static. There clearly is ample room for elaboration and improvement. Similar analyses of different (in particular, more restrictive) modeling contexts should be performed in the future, which will no doubt require refinement of the method. Additional future work involves exploring use of the COME framework in aggregating the scores given by different groups of modelers in different environments (face-to-face (FTF) or computer-mediated communication (CMC) rather than the individual scores which were aggregated into group scores to determine quality of the modeling artifacts. In the near future, we also plan to carry on in this line of research and work towards realization of a tool that facilitates synchronous and asynchronous communication and evaluation of the modeling artifacts. Our main aim is to lay a foundation for the evaluation and design of advanced, modeler-oriented support tools for collaborative modeling. We hope to have shown that the sort of evaluation presented can be fruitful, in particular in view of (empirical), HCI-style research into stakeholder-oriented, collaborative creation of models and collaborative evaluation of the modeling artifacts.

ACKNOWLEDGMENT

This work has been partially sponsored by the Fonds National de la Recherche Luxembourg (www.fnr.lu), via the PEARL programme, and NUFFIC – The Netherlands Organization for International Organization for Higher Education under the second NUFFIC Uganda project (NPT-UGA-238).

REFERENCES

Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. doi:10.1016/0749-5978(91)90020-T

Briggs, R. O., de Vreede, G. J., & Reinig, B. A. (2003). A theory and measurement of meeting satisfaction. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences* (p. 25.3). Washington, DC: IEEE Computer Society.

de Vreede, G. J., & Briggs, R. O. (2005). Collaboration engineering: Designing repeatable processes for high-value collaborative task. In *Proceedings of the 38th Annual Hawaii International Conference on Systems Sciences* (pp. 17-17c). Washington, DC: IEEE Computer Society.

Dean, D., Orwig, R., Lee, J., & Vogel, D. (1994). Modeling with a group modeling tool: Group support, model quality and validation. In *Proceedings of the 27th Annual Hawaii International Conference on System Sciences: System Sciences: Collaboration Technology Organizational Systems and Technology* (Vol. 4, pp. 214-223). Washington, DC: IEEE Computer Society.

Delbecq, A. L., & Van de Ven, A. H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Glenview, IL: Scott, Foresman.

Duivenvoorde, G. P. J., Kolfschoten, G. L., Briggs, R. O., & de Vreede, G. J. (2009). Towards an instrument to measure the successfulness of collaborative effort from the participant perspective. In *Proceedings of the 42nd Annual Hawaii International Conference on System Sciences* (pp. 1-9). Washington, DC: IEEE Computer Society.

Escobar, M. T., & Moreno-Jimenez, J. M. (2007). Aggregation of individual preference structures in AHP-group decision making. *Group Decision and Negotiation*, *16*, 287–301. doi:10.1007/s10726-006-9050-x

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behaviour. An introduction to theory and research*. Reading, MA: Addison-Wesley.

Frederiks, P. J. M., & van der Weide, T. P. (2005). Information modeling: The process and the required competencies of its participants. *Data & Knowledge Engineering*, *58*(1), 4–20. doi:10.1016/j.datak.2005.05.007

Guitouni, A., & Martel, J. M. (1998). Tentative guidelines to help choosing the appropriate MCDA method. *European Journal of Operational Research*, *109*, 501–521. doi:10.1016/S0377-2217(98)00073-3

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *Management Information Systems Quarterly*, *28*(1), 75–105.

Hoppenbrouwers, S. J. B. A., Proper, H. A., & van Reijswoud, V. E. (2005). Navigating the methodology jungle - The communicative role of modeling techniques in information system development. *Computing Letters*, *1*(3), 1–69. doi:10.1163/1574040054861276

Krogstie, J., Sindre, G., & Jorgensen, H. (2006). Process models representing knowledge action: A revised quality framework. *European Journal of Information Systems*, *15*, 91–102. doi:10.1057/palgrave.ejis.3000598

Mendling, J., & Recker, J. (2007). Extending the discussion of model quality: Why clarity and completeness may not always be enough. In *Proceedings of the 19th Conference on Advanced Information Systems Engineering*. Trondheim, Norway: Tapir Academic Press.

Moody, D. L. (1998). Metrics for evaluating the quality of entity relationship models. In T. W. Ling, S. S. Ram, & M. L. Lee (Eds.), *Proceedings of the 18th International Conference on Conceptual Modeling* (LNCS 1507, pp. 211-225).

Moody, D. L., & Shanks, G. (1994). What makes a good data model? A framework for evaluating and improving the quality of entity relationship models. *Australian Computer Journal*, *13*, 97–100.

Ngai, E. W. T., & Chan, E. W. C. (2005). Evaluation of knowledge management tools using AHP. *Expert Systems with Applications*, *29*, 889–899. doi:10.1016/j.eswa.2005.06.025

Pfeiffer, D., & Niehaves, B. (2005). Evaluation of conceptual models: A structuralist approach. In *Proceedings of the 13th European Conference on Information Systems: Information Systems in a Rapidly Changing Economy* (pp. 459-470). Regensburg, Germany: Association of Information Systems.

Recker, J. (2006). Towards understanding of process quality: Methodological considerations. In *Proceedings of the 14th Conference on Information Systems*, Goteborg, Sweden (pp. 434-445). Regensburg, Germany: Association of Information Systems.

Renger, M., Kolfschoten, G. L., & de Vreede, G. J. (2008). Challenges in collaborative modeling: A literature review and research agenda. *International Journal of Simulation and Process Modeling*, *4*(3-4), 248–263. doi:10.1504/IJSPM.2008.023686

Rittgen, P. (2007). Negotiating models. In J. Krogstie, A. L. Opdahl, & G. Sindre (Eds.), *Proceedings of the 19th Conference on Advanced Information Systems Engineering* (LNCS 4495, pp. 561-573).

Rittgen, P. (2008). *Collaborative modeling architecture (COMA)*. Retrieved November 2, 2011, from http://www.coma.nu/COMA_Tool.pdf

Rittgen, P. (2010). Quality and perceived usefulness of business models. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, Sierre, Switzerland (pp. 65-72). New York, NY: Association of Computing Machinery.

Saaty, T. L. (1980). *The analytic hierarchy process*. New York, NY: McGraw-Hill.

Saaty, T. L. (2008a). The analytic hierarchy and analytic network measurement process: Applications of decisions under risk. Honorary invited paper. *European Journal of Applied Mathematics*, *1*, 122–196.

Saaty, T. L. (2008b). Decision making with the analytic hierarchy process. *International Journal of Service Sciences*, *1*(1), 83–98. doi:10.1504/IJSSCI.2008.017590

Schuette, R., & Rotthowe, T. (1998). The guidelines of modeling: An approach to enhance the quality of information models. In T. W. Ling, S. Ram, & M. L. Lee (Eds.), *Proceedings of the 17th International Conference on Conceptual Modeling* (LNCS 1507, pp. 240-254).

Siau, K., & Rossi, M. (1998). Evaluation of information modeling methods: A review. In *Proceedings of the 31st Annual Hawaii International Conference on System Sciences*. Washington, DC: IEEE Computer Society.

Ssebuggwawo, D., Hoppenbrouwers, S. J. B. A., & Proper, H. A. (2009). Evaluating modeling sessions using the analytic hierarchy process. In A. Persson & J. Stirna (Eds.), *Proceedings of the 2nd IFIP WG 8.1 Working Conference on Practice of Enterprise Modeling Conference* (LNBIP 39, pp. 69-83).

Ssebuggwawo, D., Hoppenbrouwers, S. J. B. A., & Proper, H. A. (2010). Assessing collaborative modeling quality based on modeling artifacts. In P. van Bommel, S. Hoppenbrouwers, S. Overbeek, E. Proper, & J. Barjis (Eds.), *Proceedings of the 3rd IFIP WG 8.1 Working Conference on Practice of Enterprise Modeling Conference* (LNBIP 68, pp. 76-90).

van Bommel, P., Hoppenbrouwers, S. J. B. A., & Proper, H. A. (2007). QoMo: A modeling process quality framework based on SEQUAL. In *Proceedings of the Workshop on Exploring Modeling Methods for Systems Analysis and Design, held in conjunction with the 19th Conference on Advanced Information Systems Engineering* (pp. 118-127). Trondheim, Norway: Tapir Academic Press.

Veldhuijzen van Zanten, G., Hoppenbrouwers, S. J. B. A., & Proper, H. A. (2004). System development as a rational communicative process. *Journal of Systemics. Cybernetics and Informatics*, *2*(4), 47–51.

Vennix, J. A. M. (1996). *Group model building: Facilitating team learning using system dynamics*. Chichester, UK: John Wiley & Sons.

Wolff, F., & Frank, U. (2005). A multi-perspective framework for evaluating conceptual models in organizational change. In *Proceedings of the 13th European Conference on Information Systems, Information Systems in a Rapidly Changing Economy* (pp. 1283-1294). Regensburg, Germany: Association of Information Systems.

Denis Ssebuggwawo is a lecturer in the Department of Computer Science at Kyambogo University, Kampala, Uganda. He obtained his PhD in Computer Science (specializing in Information Systems) from the Institute of Computing and Information Sciences at Radboud University Nijmegen in 2012. Between 2001 – 2002, he undertook a Master Class in Scientific Computing at Wiskunde Onderzoekschool (WONDER), the Dutch Research School in Mathematics, formerly the Mathematical Research Institute (MRI), at Radboud University Nijmegen. He also holds a master degree in mathematics from Makerere University in Uganda. Denis's research interests border along information systems, particularly, the analysis and evaluation of "modeling artifacts" used in, and produced during, the collaborative modeling process, the application of communication, argumentation theory and dialogue games in collaborative modeling, and evaluation and selection of collaborative modeling approaches using multi-criteria decision analysis (MCDA) techniques.

Stijn Hoppenbrouwers holds Master's degrees in English Language & Literature (Utrecht) and Linguistics (Bangor, Wales). Since 1996 he has been involved as a researcher and designer in various applied science projects concerning the creation of languages and tools for modelling in information system development and enterprise engineering. He acquired his PhD in Computer Science at Radboud University Nijmegen in 2003 and was appointed assistant professor there in 2005. In 2012 he became professor at HAN University of Applied Sciences at Arnhem. Stijn's current research increasingly involves joint work with industry and mostly concerns collaborative modelling, emphasizing communication and conceptualisation with and for stakeholders untrained in systems modelling.

Henderik A. Proper is a senior research manager at the Public Research Centre - Henri Tudor in Luxembourg. He also holds a chair in Information Systems at the Radboud University Nijmegen. He leads the Enterprise Engineering research team involving researchers from these two institutions, who jointly work on a joint research program on model-enabled enterprises. Erik has a mixed background, covering a variety of roles in both academia and industry. His professional passion is the further development of the field of enterprise engineering and enterprise architecture. His long experience in teaching and coaching a wide variety of people enables him to involve and engage others in this development. He has co-authored several journal papers, conference publications and books. His main research interests include enterprise architecture, enterprise engineering, enterprise modelling, systems theory, business/IT alignment and conceptual modelling.

APPENDIX

Evaluation Instrument for the Modeling Artifacts - Using the AHP Fundamental Scale

As part of an on-going research to understand and evaluate the quality of modeling process, we would kindly request you to spend about 10 - 25 minutes of your time and fill-out this questionnaire instrument.

Please use the following evaluation instruments to determine the importance of each of the criteria of the modeling artifacts with respect to each other. Tick () or use a cross in the white-colored/non-shaded circles.

NOTE

1. If the criterion (i) on the left is considered to be "***more important*** than ***that on the left***" (j), use the "***LEFT HALF.***"
2. If the criterion (j) on the right is "***more important than that on the left***" (i), use the "***RIGHT HALF.***"
3. If an element/criterion (i) is compared to ***itself*** (equal importance) we give it rank 1 (see, shaded (dark) circles in the questionnaire).
4. Use the following AHP Fundamental scale, presented in Table 5 and Table 6, to rank the elements given in the questionnaire.

Table 5. AHP fundamental scale

Intensity of importance on absolute scale (rank)	Definition	Explanation
1	Equal importance	***criterion i*** is equally as important as ***criterion j***
2	Weak or slight importance	" is weaker or of slight importance than "
3	Moderate importance	" is of moderate importance than " (Experience or judgment moderately favours criterion i to criterion j)
4	Moderate plus	" is moderately & essentially more important than "
5	Essential or strong importance	" is essentially or strongly more important than " (Experience or judgment strongly favours criterion i to criterion j)
6	Strong plus	" is essentially and strongly more important than "
7	Very strong or demonstrated importance	" is very strongly more important or is of more demonstrated importance than " (Experience or judgment very strongly favours criterion i to criterion j)
8	Very, very strong	" is very, very strongly more important than "
9	Extreme importance	" is extremely more important than " (Experience or judgment extremely favours criterion i to criterion j)

Table 6. Evaluation instrument for the modeling language – using the fundamental scale

Criterion	← Left half scale →								← Right half scale →									Criterion
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
Understandability	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	Understandability
Understandability	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	Clarity
Understandability	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	Syntax Correctness
Understandability	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	Conceptual Minimalism
Clarity	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	Clarity
Clarity	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	Conceptual Minimalism
Clarity	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	Syntax correctness
Syntax Correctness	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	Syntax Correctness
Syntax Correctness	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	Conceptual Minimalism
Conceptual Minimalism	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	Conceptual Minimalism