IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4975

The Impact of UML on Systems Analysis: End-User Communication and Modeling Tools

Michael Eierman
MIS Team, College Business Administration, Universityof Wisconsin Oshkosh, 800 Algoma Blvd., Oshkosh, WI 54901,
eierman@uwosh.edu

Bruce C. Hungerford
MIS Team, College Business Administration, Universityof Wisconsin Oshkosh, 800 Algoma Blvd., Oshkosh, WI 54901,
hungerfo@uwosh.edu

INTRODUCTION

The Unified Modeling Language (UML) has received significant attention as the tool of the future for modeling information systems. According to Rumbaugh, Jacobson, and Booch (1999), models serve several purposes: (1) to capture and state requirements and knowledge so that all stakeholders understand and agree on them; (2) to facilitate thinking about the design of a system; (3) to capture design decisions separate from the requirements; (4) to generate usable work products; (5) to manage information about large systems; (6) to explore multiple solutions economically; and (7) to master complex systems. The UML is purported to facilitate the development of models to help achieve these goals. However, IS modeling was done prior to the development of the UML with tools such as Data Flow Diagrams (DFDs) and Entity-Relationship Diagrams (ERDs). These tools have been developed over many years of information systems development and are (and have been) taught to thousands of information systems professionals. Many organizations still use these tools, and many other organizations use these tools but are considering switching to the UML. While the UML is being promoted as the future of IS modeling, there is little empirical evidence that suggests it is better at fulfilling the purposes of a model identified above. This research seeks to inform the decision whether to adopt the UML over traditional modeling languages by comparing outcomes of the two types of languages.

BACKGROUND

The UML was developed as a unification of three object-oriented analysis and design techniques. The object-oriented approach is considered a paradigm shift in systems development because the focus of analysis, design, and programming is on objects, which have both properties and methods. This contrasts with the traditional systems development approach, which focuses on data requirements and programs to operate on the data.

The benefits of the object-oriented approach are expected to include (1) more effective systems analysis, (2) more effective communication between users and developers, (3) more effective systems design, (4) easier translation between systems design and implementation, and (4) more productive programming.

THEORY

There has been very little empirical research comparing the UML to traditional modeling languages. The limited work that has been done has focused on object-oriented development methods versus traditional analysis and design approaches. This work does not specifically include the UML. However, based on this work, there is limited evidence that the type of modeling language may have an impact. For example, Krovi and Chandra (1998) find evidence that an object-oriented model is easier to understand than a process model because it more closely resembles the cognitive representations used by individuals.

Fowler and Scott (2000) suggest that the fundamental reason to use the UML is communication. An examination of the purposes of a model presented by Rumbaugh, Jacobson, and Booch (1999) finds that a significant number of those purposes involve communication of some sort. Therefore, in the absence of prior work, we refer to research on message effects and graphics comprehension in communication theory to develop a theoretical basis for this study.

Sager (1994) states that communication is a purposeful human activity concerned with affecting the knowledge structure of individuals. The message is the vehicle through which communication occurs (Bowers, 1989). A message must be processed cognitively for the message to have an effect (Kellermann and Lim 1989). Therefore, if a model of an information system is used primarily to communicate with users of the information system, the model may be conceptualized as a message that must be cognitively processed by users to develop their understanding of the system.

An information system model using traditional techniques such as DFDs and ERDs or the UML is a graphical representation of the system. Larkin and Simon (1987) suggest diagrams may be superior to verbal descriptions because:

- Diagrams can group together all information that is used together, thus avoiding large amounts of search activity for the elements needed to make an inference.
- Diagrams typically use location to group information about a single element, avoiding the need to match symbolic labels.
- Diagrams automatically support a large number of perceptual inferences, which are extremely easy for humans.

Winn (1994) presents an overview of how the symbol system of graphics interacts with the viewers' perceptual and cognitive processes. In his description, the graphical symbol system consists of two elements: (1) Symbols that bear an unambiguous one-to-one relationship to objects in the domain of interest; and (2) The spatial relations of the symbols to each other. Thus, how symbols are configured spatially will affect the way viewers understand how the associated objects are related and interact. Zhang (1997), in an experiment conducted using a Tic-Tac-Toe board and its logical isomorphs, shows that external representations of information are more than just memory aids. Her research suggests that the form of representation determines the information that can be perceived in a diagram.

These studies suggest that different methods of graphically representing an information system may impact the ability of the user to comprehend the features and functioning of the system being modeled. The current research seeks to understand if there is a substantive difference between the traditional modeling methods and the UML in communicating with users. To process a diagram, the individual must use three types of processes (Larkin and Simon 1987): (1) Search, in which the user of the diagram seeks to locate specific sets of elements; (2) Recognition, in which the user of the diagram matches elements located in the search with data required; and (3) Inference, in which the user adds new information to his or her understanding. In the context of this research, the first process, Search, may indicate how easily an individual locates specific information in either of the two modeling approaches. However, this research does not investigate differences in search between the approaches for two reasons. First, the ability to find required information effectively will have a direct impact on differences in the other two processes. If individuals cannot easily find the information they are looking for, they will not be as effective at recognizing required information, nor will they be as effective at making inferences about that information. Second, effective investigation of search activity requires sophisticated technology and research techniques that we are not prepared to perform at this time. This research also does not investigate the third process, Inference. The ability to make inferences about the information acquired from a model requires individuals who have sufficient knowledge about a situation to allow them to apply the new information. Our subject population for this study does not have that knowledge. Accordingly, the current research examines if there is a difference in the ability of individuals using different types of information systems models to recognize required data. The basic hypothesis and two related hypotheses investigated in this study are:

H1: There is no difference in the ability to recognize required information between individuals using traditional information system modeling tools (e.g., DFD and ERDs) and those using the UML.

$H1_1$: There is no difference in the ability to recognize required information about the data used in the information system between individuals using traditional information system modeling tools (e.g., DFD and ERDs) and those using the UML.

$H1_2$: There is no difference in the ability to recognize required information about the process used in the information system between individuals using traditional information system modeling tools (e.g., DFD and ERDs) and those using the UML.

METHOD

The hypotheses were investigated with a field experiment using undergraduate students enrolled in Introduction to Marketing sections. Subjects were randomly assigned to one of two treatments: (1) system description with traditional modeling diagrams or (2) system description with the UML. The experiment lasted approximately one hour. In the first 20 minutes of the experimental session, the subjects read a training document that included an explanation of the symbols used in the modeling language treatment to which they were assigned, a written description of an information system, and a model of an information system that was annotated to explain how the model represented the information system. In the final 40 minutes of the session, the subjects were presented a model of a different information system and a set of 20 questions about the data and operation of that information system. Subjects were allowed to use the training document, if desired, during this portion of the experiment. Upon completion of the survey questions, subjects were asked to answer 10 questions concerning their perception of the task and the diagrams provided in the model. Finally, subjects were asked to provide basic demographic data. All students received extra credit in their marketing class for participation in the experiment. Students who scored above 70% on the survey were put in a pool from which four individuals were chosen at random to receive a $20.00 prize.

The training document was developed to explain the symbols used in each of the modeling languages and to provide an example of their use. For the traditional modeling language, this included Data Flow Diagrams, Entity-Relationship Diagrams, Decision Tree Diagrams and a Data Dictionary. For the UML, this included Use Case Diagrams, Class Diagrams, Sequence Diagrams, and State Diagrams. The documents were developed to provide the same level of explanation for each of the modeling languages. These were reviewed for accuracy and completeness by two MIS professors not involved in the project. The training information system was a Student Registration System. This type of system was selected because it was believed that students would have some knowledge of how such a system worked and would therefore find it easier to understand the example. The treatment information system described by the models was an on-line grocery store. Each model was reviewed to ensure that it provided enough information to answer all the questions on the survey. The instruments used in the experiment were tested in three different pilots. In the first pilot, the training documents were provided to the subjects, but were explained by one of the authors. It was determined from this experience that we could not guarantee the same level of training between experimental sessions because of student questions and level of detail covered. The decision was therefore made to provide self-paced training by allowing subjects to examine the training documents on their own. The other two pilots focused on ensuring that both model types provided the same information and that the questions on the survey matched the information provided.

The experiment was administered in four sessions. Both modeling languages were tested in the same session by randomly assigning the subjects to use one or the other treatment. Of the 68 students participated in the experiment, nine were eliminated during analysis because they indicated that they had been exposed to the modeling language prior to the experiment.

RESULTS

Data analysis was performed on 59 questionnaires: 30 subjects received the traditional modeling language treatment, and 29 subjects received the UML treatment. The subjects were similar demographically in all aspects except gender. A disproportionate number of females received the UML treatment. Gender differences were tested, and no significant difference in performance between male and female subjects was found.

Data was collected about subjects' perception of the task and the diagrams used in the models. The first set of these questions (Table 1) is examined to determine if one modeling language was perceived to be harder to use than the other. There were no significant differences between the traditional modeling language and the UML in subjects' perception of the difficulty of the task. In both languages subjects found it somewhat difficult to relate the diagrams to develop an overall understanding of the system and somewhat difficult to find specific information to answer the questions. Subjects in both languages also found the overall task to be somewhat difficult. The questions and answers for both UML and traditional languages received an almost identical evaluation of being somewhat easy to understand. The questions and answers were exactly the same for both languages, which suggests that subjects understood what they were supposed to do and implies that other differences detected in the study are attributable to differences in the modeling languages.

The study also was interested in determining if some diagrams were more useful than others in understanding the system. Subjects were asked if they found each of the diagrams presented useful for answering the questions (Table 2). The subjects that received the UML treatment showed very little difference, finding all of the diagrams to be somewhat useful. However, the subjects that received the traditional treatment showed a much wider variation. While they found all of the diagrams to

Table 1: Perception Questions

	Traditional	UML
I found it easy to relate the various documents to understand overall system.	2.70	2.57
It was difficult to find the specific information necessary to answer the questions.	3.60	3.36
Some documents were much more useful than others.	4.07	3.71
Multiple-choice questions and answers were easy to understand.	3.70	3.64
The overall task was difficult.	3.63	3.54
Scale: 1 – Strongly Disagree, 2 – Disagree, 3 – Neutral, 4 – Agree, 5 – Strongly Agree		

Table 2: Diagram Usefulness/Used Significantly More Than the Others

Traditional Diagram *(UML Diagram)*	**Traditional**		**UML**	
	Score*	Use	Score*	Use
I found the Entity-Relationship *(Class)* Diagram useful in answering the questions	3.17	16.7%	3.29	41.4%
I found the Data-Flow *(Sequence)* Diagram useful in answering the questions	3.03	23.3%	3.39	24.1%
I found the Decision Tree *(State)* Diagram useful in answering the questions	3.30	0.0%	3.36	0.0%
I found the Data Dictionary *(Use Case View)* useful in answering the questions	3.9	43.3%	3.25	10.3%
All were used approximately equally.		16.7%		34.5%

Note: Each row shows traditional and UML "equivalents" – comparisons should be done primarily by column rather than by row.
*Scale: 1 – Strongly Disagree, 2 – Disagree, 3 – Neutral, 4 – Agree, 5 – Strongly Agree

Table 3: UML vs. Traditional Score Statistics

					Equality of Means		
Type	**Treatment**	**Mean**	**Std. Deviation**	**Std. Error Mean**	**t**	**df**	**Sig. (2-tailed)**
Overall	UML	.5987	.10898	.02024	0.867	57	.390
	Traditional	.5685	.15390	.02810			
Data	UML	.6334	.13490	.02505	0.193	57	.848
	Traditional	.6256	.17566	.03207			
Process	UML	.5628	.15076	.02799	1.093	52.515	.279
	Traditional	.5107	.21121	.03856			

N: UML, 29; Traditional, 30
Levene's Equality of Variances Test: Sig. – Overall, .066; Data, .080; Process, .009

be somewhat useful, the traditional treatment subjects rated the Data Dictionary as much more useful than the other software artifacts. This result is borne out by an additional question that asked the students to identify which diagram they used significantly more than the others (Table 2). Subjects that received the traditional treatment chose the Data Dictionary as being used significantly more often than other diagrams by an almost 2 to 1 margin. Subjects that received the UML treatment chose the Class diagram as being used significantly more often than other diagrams, but a large percentage of them indicated that all the diagrams were used equally.

The hypotheses were tested using a single factor, two treatment level design. The dependent variables are based on the scores of the subjects. They are ratio in nature – a subject with a score of 70% did twice as well as a subject with a score of 35%. Statistical tests were performed on the subjects' overall scores, scores on questions concerning the data represented in the model (a total of 10), and scores on questions concerning the process represented in the model (10). The normality of the output for overall, data, and process scores was checked, and all met the skewness and kurtosis requirements. Variance assumptions were checked for each hypothesis using Levene's Test for Equality of Variances; equality could not be rejected at the .05 level for overall and data scores, but was rejected for the process scores. Each hypothesis was then examined using the appropriate t-test (Morgan and Griego 1998).

The average score for subjects on all questions was 59.87% for the UML treatment and 56.85% for the Traditional treatment (Table 3). Test statistics demonstrate that there is not a significant difference between the mean scores. Therefore, the first hypothesis cannot be rejected: There is no difference in the ability to recognize required information between individuals using traditional information system modeling tools (e.g., DFDs and ERDs) and those using the UML.

The average score for subjects on data questions was 63.34% for the UML treatment and 62.56% for the Traditional treatment (Table 3). Test statistics demonstrate that there is not a significant difference between the mean scores. Therefore, the second hypothesis cannot be rejected: There is no difference in the ability to recognize required information about the data used in the information system between individuals using traditional information system modeling tools (e.g., DFDs and ERDs) and those using the UML.

The average score for subjects on process questions was 56.28% for the UML treatment and 51.07% for the Traditional treatment (Table 3). Test statistics demonstrate that there is not a significant difference between the mean scores. Therefore, the third hypothesis cannot be rejected: There is no difference in the ability to recognize required information about the process used in the information system between individuals using traditional information system modeling tools (e.g., DFD and ERDs) and those using the UML.

CONCLUSION

The results of this study suggest that the Unified Modeling Language does not hold a significant advantage over traditional modeling languages for communication with end-users. Subjects using either the UML or traditional language models on average were able to recognize required information just over half the time. Neither modeling language had an advantage communicating data or process information to end-users. This suggests that the decision to use either of the languages has little impact when trying to communicate system design information to users.

The study has obvious limitations. First, the subjects were undergraduate students with no significant prior knowledge of the information system presented to them. This may not be a realistic representation of users involved in the design of a system. Second, the subjects were not able to ask questions or discuss the design with an individual knowledgeable about the design. This is also probably not representative of the "real world." Finally, the system used in the research, while not trivial, was not of the size and complexity found in systems typically developed by business.

However, despite the limitations, this study does imply that the move to UML as a modeling language for information systems may not confer a tremendous advantage over use of traditional modeling languages. More studies are needed that explore different aspects of model use in information system design.

REFERENCES

Bowers, J.W. "Introduction," in *Message Effects in Communication Science*, J.J. Bradac (ed.), Sage Publications, Newbury Park, CA, 1989, pp. 10-23.

Fowler, M., and Scott, K. *UML Distilled: A Brief Guide to the Standard Object Modeling Language*, Addison-Wesley, New York, 2000.

Kellermann K., and Lim, T. "Inference-Generating Knowledge Structures in Message Processing," in *Message Effects in Communication Science*, J.J. Bradac (ed.), Sage Publications, Newbury Park, CA, 1989, pp. 102-128.

Krovi, R., and Chandra, A. "User cognitive representations: The case for an object oriented model," The *Journal of Systems and Software* (43), 1998, pp. 165-176.

Larkin, J.H., and Simon, H.A. "Why a Diagram is (Sometimes) Worth Ten Thousand Words," *Cognitive Science* (11), 1987, pp. 65-99.

Morgan, G.A., and Griego, O.V. *Easy Use and Interpretation of SPSS for Windows: Answering Research Questions with Statistics*, Lawrence Erlbaum Associates, Mahwah, NJ, 1998.

Rumbaugh, J., Jacobson, I., and Booch, G. *The Unified Modeling Language Reference Manual*, Addison-Wesley, New York, 1999.

Sager, J. *Language Engineering and Translation: Consequences of Automation*, John Benjamins, Amsterdam, 1994.

Winn, W. "Contributions of Perceptual and Cognitive Processes to the Comprehension of Graphics," in *Comprehension of Graphics*, W. Schnotz and R.W. Kulhavy (eds.), North-Holland, Amsterdam, 1994.

Zhang, J. "The Nature of External Representations in Problem Solving," *Cognitive Science* (21:2), 1997, pp. 179-217.

Related Content