An Empirical Evaluation of the Requirements Engineering Tool for Socio-Technical Systems

Msury Mahunnah
Department of Software Science
Tallinn University of Technology
Tallinn, Estonia
msury.mahunnah@ttu.ee

Kuldar Taveter
Department of Software Science
Tallinn University of Technology
Tallinn, Estonia
kuldar.taveter@ttu.ee

Raimundas Matulevičius
Institute of Computer Science
University of Tartu
Tartu, Estonia
raimundas.matulevicius@ut.ee

***Abstract*—One of the major problems of requirements engineering is the lack of sufficient empirical evidence that evaluates the benefits of modelling tools for Model-Driven Engineering (MDE). In this paper, we report on the results of empirical study that compares the modelling effort and effectiveness of the novel software tool for modelling requirements of sociotechnical systems against modelling on paper. We have asked 8 persons who received 2 different treatments – modelling on software against modelling on paper to create 2 requirements models – goal and domain models – for 2 different case studies. The study finds that modelling effort with a software tool nearly equals to modelling effort on paper while modelling effectiveness with a tool is higher than modelling effectiveness on paper. The major limitation of this study is the use of students as participants and the use of small sample size. In the future work, we will conduct another empirical study with a large sample size of professionals that aims to increase the confidence in the results obtained from this empirical study.**

***Index Terms*—Requirements, sociotechnical system, model-driven engineering, tools.**

I. Introduction

A sociotechnical system is a software intensive system that has defined operational processes followed by human operators and which operates within an organization [1]. For example, if a computer game does not feel fun, we will not play it; if an ecommerce website does not feel trustworthy (irrespective of the actual security) we will not purchase from it; and if a social networking application does not feel engaging we will not use it [2]. Therefore, a social aspect of the system plays an important role to complement a technical aspect of system and form a sociotechnical system [1]. These aspects can be elaborated into different levels and perspectives. In [3], Whitworth suggest four levels that describe sociotechnical systems: physical, information, personal, and group level. In [4], da Conceição, *et al*. distinguish seven abstraction levels of sociotechnical systems in the maritime domain: natural environment, and reactive, automated reactive, proactive, planning scheduling, planning strategic, and political-economic levels. In addition to the abstraction levels, [5] proposes six perspectives of sociotechnical systems, which are orthogonal to the abstraction levels: goals, people, technologies, physical infrastructure, cultural assumptions, and processes, and working practices.

To cope with engineering sociotechnical systems considering the multitude of diverse abstraction levels and perspectives, Baxter and Sommerville [6] emphasize that appropriate models and abstractions should be used for representing sociotechnical considerations. Consequently, in [7], the Agent-Oriented Modelling (AOM) methodology is proposed and elaborated in [8]–[10]. It utilizes goal, role, organization and domain models during requirements engineering phase which is supported by the novel Agent-Oriented Modelling for Sociotechnical System (AOM4STS)[1] software tool. The AOM4STS software tool aims to reduce the effort and increase the effectiveness of the current practice of applying the AOM methodology by modelling on paper. However, the gap exists in the empirical evidence that compares the effort and effectiveness between modelling with the AOM4STS tool against modelling on paper.

In this paper, we fill the identified gap by answering the following research question: To what extent does the novel AOM4STS software tool improve the process of requirements modelling on paper? To establish complexity-reducing separation of concerns, we deduce the following sub-questions: (i) To what extent is the modelling effort with the AOM4STS software tool different from modelling on paper? (ii) To what extent is the modelling effectiveness with the AOM4STS software tool different from modelling on paper?

To find answers to the research questions, we carried out modelling experiments following the guidelines of experimentation in software engineering [11]. Our goal is to empirically compare the effort and effectiveness of modelling process using the AOM4STS tool against modelling on paper.

The rest of this paper is structured as follows. Section II presents related work. Section III provides an overview of the AOM methodology and briefly describes models for requirements engineering of sociotechnical systems. Section IV presents key features of the AOM4STS software tool. Section V describes an experiment conducted for empirical evaluation of the tool. Section VI summarizes the results of the experiment. Finally, conclusions are drawn in Section VII.

II. Related Work

The AOM methodology [7] stems from the Model-Driven Engineering (MDE) [12] paradigm that focuses on the systematic use of models as primary engineering artefacts throughout the system engineering lifecycle. Among the key benefits of

[1] goo.gl/Wr6q57

MDE paradigm are effective expression of domain concepts [13], decreasing system development time (effort), and improving system quality [14].

Despite the benefits of the MDE paradigm, various studies show that a domain-specific MDE language is not enough for industry-wide adoption and a tool supporting such language increases the complexity of the development process instead of diminishing it [15]. Elsewhere, Whittle et al. [16] interviewed 39 practitioners on tool-related issues affecting the adoption of MDE. The results of this study indicate that the complexity of the modelling tools is among the major issues hindering practical application of MDE. Moreover, the study [16] suggests the need for developing new software modelling tools that focus on early design stages, support creativity in modelling, and match the way people think rather than the other way round. Another study involved 15 MDE experts in a thought experiment to identify the biggest problems with current MDE technologies [17]. The results of this study found that steep learning curves and arduous user interfaces are among significant usability challenges to industry-wide adoption of MDE tools.

Considering the benefits of MDE languages and the challenges of using MDE tools, Gorschek et al. [18] conducted a survey with 3785 developers to find out the extent to which design models are used before actual coding. The results of this study found that design models are not used very extensively in industry. Moreover, in such companies where they are used, the notations are often *not* UML notations, the use of design models is informal and without tool support. Instead of relying on tools, the models are usually drawn on whiteboard or paper.

The findings from this review of related work point to the need of conducting research studies on MDE software tools to empirically compare claimed benefits of a modelling tool against modelling on a whiteboard or paper.

III. Agent-Oriented Methodology for Engineering Sociotechnical Systems

The AOM methodology proposed in [7] and elaborated in [8]–[10] is centered on the notions of agent, goal, role, and domain entity. A sociotechnical system (STS) is defined in AOM as a system consisting of diverse active components – both human and man-made (software and robots) – that collaborate in designing and sustaining the sociotechnical system. We term such active components as agents, which form a distributed system. AOM is an approach for engineering complex sociotechnical systems where a problem domain is conceptualized in terms of the goals to be achieved by the system, the roles required for achieving them, and the domain entities embodying the required knowledge.

Agent-oriented models for problem domain analysis, which are used for representing the requirements, help to improve communication between information technology (IT) and non-IT experts during the requirements elicitation phase in the development. These models provide a high-level description of the system and use graphical notations to enable project stakeholders to obtain a common understanding about the system requirements. Table I outlines the objective of each agent-oriented model for problem domain analysis.

Table I. Overview of models for problem domain analysis

ID	Model Name	Objective
1	Goal Modelling	To represent functional and non-functional requirements of the system as goals and quality goals, respectively, roles required for achieving the goals, and relationships among all of them.
2	Role Modelling	To list responsibilities and constraints of each role in the system.
3	Organization Modelling	To show the types of relationships that exist between the roles of the system.
4	Domain Modelling	To represent the knowledge represented within the system by capturing the types of domain entities (knowledge items) and the relationships between the roles and domain entities.

The scope of this paper includes goal and domain modelling. Role and organization modelling are not considered during the empirical study reported in Section V.

IV. AOM4STS Tool Support

The AOM4STS tool [19] supports the AOM methodology which involves incremental refinement of models in an iterative manner. Therefore, consistency checking becomes a necessary feature of the AOM4STS tool to ensure that the modelling artefacts represented in Table I remain consistent with each other. The following subsections A to B briefly describe the two key features of the AOM4STS tool. Due to space limitation, other features of AOM4STS tool are not describes in this paper.

A. Information propagation

According to Table I, models in the AOM methodology are divided horizontally along three abstraction layers and vertically along three viewpoint perspectives. Considering this, during the modelling process, the AOM4STS tool propagates information vertically across abstraction layers and horizontally across viewpoint perspectives.

In the vertical information propagation, models for problem domain analysis act as input for platform-independent design models while platform-independent design models act as input for platform-specific design models and prototypes.

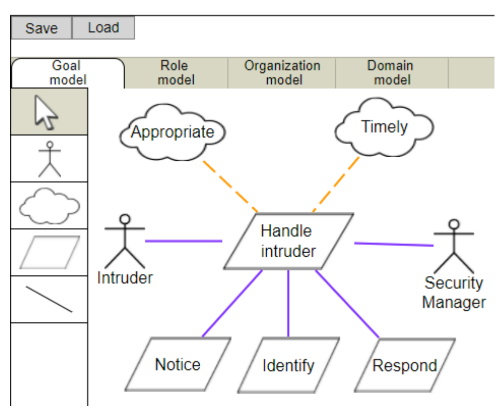

Fig. 1. Screenshot of the AOM4STS software tool

In the horizontal information propagation, AOM4STS propagates information across models of different viewpoint perspectives but within the same abstraction layer. The problem domain analysis layer contains four different models as outlined in Table I. The information in these models is propagated horizontally across the three viewpoint perspectives. For example, all the roles identified during goal modelling are horizontally propagated to role models, organization model, and domain model. Fig. 1 depicts a screenshot of the AOM4STS tool that describes the goal model of the Intruder Detection System [20].

B. Consistency checking

The AOM4STS tool continuously performs consistency checking to prevent certain errors from being made in the first place. The errors checked against are definition errors, simple typing errors, and violations of scope. The principle of detecting definition errors is that it is only possible to create a reference to an entity after the entity has been defined. For example, it is only possible to create a reference to a role in a domain model after that role has been defined in a goal model. Moreover, when the user deletes an entity, the tool deletes all references to the deleted entity. The effect is different when modelling on whiteboard or paper.

The principle of detecting simple errors allows users to create only syntactically correct connections between component types. The tool prevents all syntactically wrong connections and generates the corresponding error messages in the bottom frame of the tool containing user activity logs. For example, according to the AOM methodology, it is syntactically wrong to create a connection between a role and quality goal in the goal model.

Lastly, in preventing violations of scope constraints, the tool allows an analyst to neither increase nor decrease the scope of the project identified during the earlier modelling stages. In other words, once the goal modeller has defined the scope of the project, the role modeller, organization modeller, and domain modeller can only refine the requirements but not increase or decrease the scope of the project. This makes the AOM4STS tool suitable for an iterative (agile) modelling process that supports the AOM methodology [21].

V. Empirical Evaluation

The AOM4STS tool, which has been presented in Section IV, is an online diagramming software tool that supports the methodology of requirements engineering for sociotechnical systems described in Section III. In this section, we present an empirical study for evaluating requirements modelling for a sociotechnical system with the help of the AOM4STS tool in comparison with modelling the requirements for the same sociotechnical system using pen and paper. The design of the experiments follows the guidelines by Wohlin et al. [11] on how to set up and document empirical studies in software engineering.

A. Experimental Design

This section describes the plan for the experiment that was followed during the empirical study.

1) Goal of the study

The goal of the empirical study was to compare software-based processes of modelling requirements for sociotechnical system against paper-based processes of modelling the same requirements to find out if the benefits expected from using the AOM4STS tool were present when used by novice users of the AOM methodology and the AOM4STS tool in a realistic environment. Hence, the independent variables of this experiment were the modelling approaches that we wanted to compare: Modelling on Paper (MoP) and Modelling on Software (MoS). The former allows subjects to use pen and paper to create the requirements models while the latter allows subjects to use the AOM4STS tool for the same purpose.

The evaluation of a modelling approach can be characterized by two dependent variables: (1) the effort during modelling; and (2) the effectiveness of the modelling process.

With the objective of evaluating possible benefits of using the AOM4STS tool for modelling requirements for sociotechnical systems, in comparison with the use of pen and paper, we defined the following two research questions.

RQ1: To what extent is the modelling effort with the AOM4STS software tool different from modelling on paper?

RQ2: To what extent is the modelling effectiveness with the AOM4STS software tool different from modelling on paper?

2) Experimental Design

The experiment was run in a lecture room, as a blocked subject-object experiment [11] whereby a set of objects were assigned to a set of subjects in a random way. The two objects – requirements specifications of two sociotechnical systems – were assigned to each participant (subject).

3) Subjects

The participants of the experiment were 8 post-graduate students (MSc and PhD) taking the requirements engineering course. Among various sub-topics of this course are goal-oriented approaches and agent-oriented methodologies for requirements engineering. These participants of this study were not students taught by experimenters. Furthermore, they were using only paper-based requirements modelling in their requirements engineering course.

4) Objects

The objects of this study were two small sociotechnical systems – a Meeting Scheduler System (MES)[2] [22] and a Personalized Emergency System (PES)[3] [23]. The former is a computer-based service that supports setting up meetings while the latter is a system that supports a person, generally an older person, to remain living at home longer.

5) Data collection

The experiment was conducted for 3 hours in two consecutive days – 90 minutes on day 1 and another 90 minutes on day 2. In this experimental design, each subject performed the experiment tasks with both objects and with both treatments. This means that on day 1, half of the subjects were given the PES object and the remaining half of the subjects were given the MSS object. Moreover, half of those who received the PES object conducted the modelling on paper and the other half with the AOM4STS tool. Similarly, half of those subjects who received the MSS object conducted the modelling on paper and the other half with the software tool. On day 2, each subject

[2] https://goo.gl/AVrDzx

[3] https://goo.gl/wAMxnE

changed the object and treatment. This experimental design mitigates learning effects between the two objects and between the two treatments [11].

6) Data Analysis

For the comparison of the MoP and MoS treatments, we (1) collected data through questionnaires, (2) applied measures of central tendency – mean, median and mode [24] – to compare the impact of the collected data on the treatments, and (3) grouped the results based on the collected data to answer the two research questions RQ1 and RQ2.

To evaluate the two treatments, the questions from q4 to q10 from Table II were repeated for each treatment used during the experiment. Similarly, the answers to the questions from postq2 to postq5 from Table II, which mostly focused on the evaluation of the AOM4STS tool, were compared with respect to the value 3, which is the neutral answer according to the Likert scale used in the study. The answers to the questions from q1 to q3 from Table III, which captured the relative time spent on reading the tutorial and understanding the case study, and the actual time spent on modelling requirements for the system in %, were multiplied with the overall time used by the subject in the corresponding experiment to obtain time measurements that could be compared between the two treatments.

B. Execution

As for the task of the experiment, the subjects had to create two requirements models – goal model and domain model – for two case studies, one of which had to be modelled on paper and another one with the AOM4STS tool. Each case study had to be modelled with as much of detail as possible, with the given treatment, and by following the step-by-step description of the requirements for each case study. Before the beginning of the modelling task, each subject had to fill in a pre-questionnaire. After completion of the modelling task for each case study, a subject had to fill in a questionnaire for the corresponding case study and treatment used. Finally, each subject had to fill in a post-questionnaire. A collection of questions for each type of questionnaire is provided in Table II.

The pre-questionnaire aimed to assess the knowledge of the subjects with respect to computing studies, requirements engineering, and agent-oriented requirements modelling.

The questions from preq4 to preq6 as presented in Table II aimed to assess the knowledge of the AOM methodology acquired after completion of the tutorial, and therefore measured the adequateness of the tutorial given before the modelling experiment. The questionnaire associated with each treatment the questions from q4 to q10 as is described in Table II, which evaluated the adequateness of the objects of the case study and collected perceptions by the subjects based on the specific treatment applied. Finally, the questions in the post-questionnaire from postq2 to postq5, as listed in the bottom of Table II, collected data about the effectiveness of the requirements modelling with the AOM4STS tool as compared with modelling the requirements on paper.

TABLE II. A SET OF THE QUESTIONS IN THE QUESTIONNAIRES

1 – Strongly disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly agree	
preq4	Basic principles of the AOM modelling methodology are clear
preq5	The visual notations of the AOM methodology are clear
preq6	Basic knowledge of using the AOM4STS tool has been acquired
q4	The description of the case study was clear to me
q5	I had no difficulties in modelling the goal model
q6	I had no difficulties in modelling the domain model
q7	I had enough time for accomplishing the modelling task
q8	Goal decomposition was very useful in this task
q9	The concepts of the AOM methodology were detailed enough to model the requirements of the system
q10	The effort of modelling seems too high for an efficient use of the methodology in practice
postq2	The propagation of roles created in the goal model into the domain model is helpful for the modeller
postq3	The propagation of changes made to the roles in the goal model into the domain model helps to reduce the modelling effort
postq4	The modelling software supports creation of syntactically correct models by preventing and reporting syntactically wrong connections
postq5	The use of coloured connections in the creation of the models by the modelling software helps to improve the readability of the resulting models

Furthermore, the overall time needed for completing the experimental task was recorded before filling in the corresponding questionnaire. The participants were also asked to keep track of the time in fractions (in %) spent on various activities. An indicative period of 1 hour was given to the subjects as a suggestion for performing the experimental modelling task on each day of the experiment, but subjects were free to take the time they required for completing the experimental task. The questions that were asked on the time spent on activities in each experiment are presented in Table III.

TABLE III. QUESTIONS ON THE TIME SPENT ON THE ACTIVITIES IN EACH EXPERIMENT

Question label	Question description
duration	Time used for the task, in minutes
q1	Reading the description of the AOM methodology in %
q2	Reading and understanding the description of the case study in %
q3	Modelling the case study in %

VI. RESULTS ANALYSIS AND INTERPRETATION

This section presents the results of the empirical study by considering the measures of central tendency of the data collected from the subjects through questionnaires and provides the interpretation of the results that yields answers to the research questions RQ1 and RQ2.

A. Adequateness of the experimental settings

Before analysing the main factors of the empirical study, we analysed if the settings for the experiment were adequate. The pre-questionnaire asked about the subject's experience in the fields of computing, requirements analysis, and agent-oriented requirements modelling to measure the influence of these co-factors on the study.

1) Adequateness of the subjects

Although all the subjects were postgraduate students in the requirements engineering course, they had different experiences in computing. The results of the collected data[4] show that half of the subjects had little knowledge in computing, whereby 38% had experience in computing obtained through research projects, and 12% had experience in computing obtained through working as a computing professional in IT companies.

All the subjects were registered for the requirements engineering course to either acquire new knowledge or improve their knowledge in requirements engineering. The results of the collected data show that 75% of the subjects had little experience in requirements analysis while the remaining 25% had research experience in requirements analysis.

Before the subjects started to participate in the experiment, we gave a short tutorial about agent-oriented requirements modelling. After the tutorial, we did a short demonstration on agent-oriented goal modelling and domain knowledge modelling by using the AOM4STS tool. To be able to measure the effectiveness of the tutorial and demonstration for the subjects, we measured the prior experience of the subjects in agent-oriented requirements modelling. The results of the collected data shows that 75% of the subjects did not have any experience in agent-oriented requirements modelling while 25% had little experience in agent-oriented requirements modelling.

The results in Table IV provide a summary of the adequateness of the settings for the experiment after completing the tutorial and demonstration of the AOM4STS tool.

TABLE IV. RESULTS OF THE MEDIANS OF THE ADEQUATENESS OF THE SUBJECTS

Ref.	Question	Median
preq4	Basic principles of the AOM modelling methodology are clear	5
preq5	The visual notations of the AOM methodology are clear	5
preq6	Basic knowledge of using the AOM4STS tool has been acquired	5

The results presented in Table IV show that the subjects acquired adequate understanding of the AOM methodology and the AOM4STS modelling tool for participating in the modelling experiment to give undistorted feedback.

2) Adequateness of the case studies

Adequateness of the objects used in the experiment was evaluated by the questions q4 and q7, which were answered by the subjects after completion of the modelling task independently of the treatment used. The questions and results are presented in Table V.

On the one hand, for the question q4, the median value for the PES case study was 5, while the median value for the MES case study was 4. This means that the subjects considered that the descriptions of both case studies were nearly equally clear.

[4] https://goo.gl/qatU6s

TABLE V. RESULTS OF THE MEDIANS FOR THE ADEQUATENESS OF THE OBJECTS

Ref.	Question	Median (PES)	Median (MES)
q4	The description of the case study was clear to me	5	4
q7	I had enough time for accomplishing the modelling task	4	4

Although the description of the PES case study was considered clearer compared to the description of the MES case study, we believe the objects were adequate to provide unbiased results because both results were above the median value 3. On the other hand, for the question q7, the median value for both the PES and MES case studies was 4. This means that the subjects agreed that they had enough time for accomplishing the modelling task. The value 4 for each case study reduces the possibility of having biased results with respect to the time allocated for the experiment. Moreover, since the result for the question q4 was above the neutral value, which is 3 in the 1…5 Likert scale, the subjects did not experience time pressure when performing the modelling tasks. Consequently, the time allocated for the experiment did not have any influence on the results of the experiment. Therefore, we can claim that in overall the settings for the experiment were adequate.

B. Main factor: results and interpretation

In this section, we provide the results for the main factor of the experiment – the approach used – and compare the two treatments.

1) Evaluation of modelling effort

In this sub-section, we provide an answer to the research question RQ1 addressing the modelling effort, which was stated in Section V.A.1, based on the mean values represented in Fig. 2 and variance values shown in Fig. 3.

The question q0 records the overall time used by a subject for modelling a case study. The mean for modelling on paper (30) was nearly the same as the mean for modelling with the tool (29.6). However, the variance for modelling on paper (5.8) is noticeably higher than that for modelling with the tool (3.5).

The question q10 records modelling effort perceived by the subjects. The mean value of the modelling effort perceived by the subjects for modelling on paper and modelling with the tool were both close to 3 and their variances close to 0.7. Therefore, the subjects perceived the modelling effort on paper to be the same as the modelling effort with the tool.

The question q1 records the time used by the subjects for reading and understanding the modelling methodology. The mean time used by the subjects for reading and understanding the modelling methodology was slightly higher for subjects who conducted modelling with the tool (5.5) compared to those who modelled on paper (4.1). The variance of the time used by the subjects for reading and understanding the methodology was noticeably higher for subjects who conducted modelling with the tool (2.6) compared to those who modelled on paper (1.3).

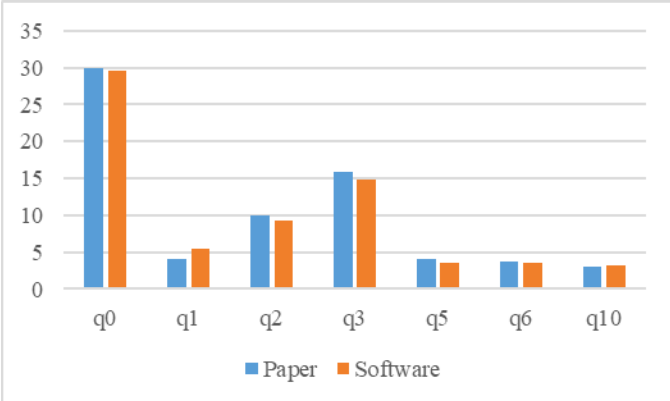

Fig. 2. Mean values for comparing modelling effort of the two treatments

Moreover, the question q2 records the time used by the subjects for reading and understanding the description of the case study. The mean time used by the subjects for reading and understanding the case study was slightly lower for subjects who conducted modelling with the tool (9.3) compared to those who modelled on paper (10.1). The variance of the time used by the subjects for reading and understanding the case study was noticeably lower for subjects who conducted modelling with the tool (3.1) compared to those who modelled on paper. Furthermore, the question q3 records the time consumed by the subjects for conducting the actual modelling using the two treatments. The mean time used by the subjects for conducting the actual modelling with the tool (14.8) was slightly lower than that for those who conducted the actual modelling on paper (15.8). The variance of the time used by the subjects for conducting the actual modelling with the tool was noticeably lower (3.6) than that for conducting the actual modelling on paper (6.7).

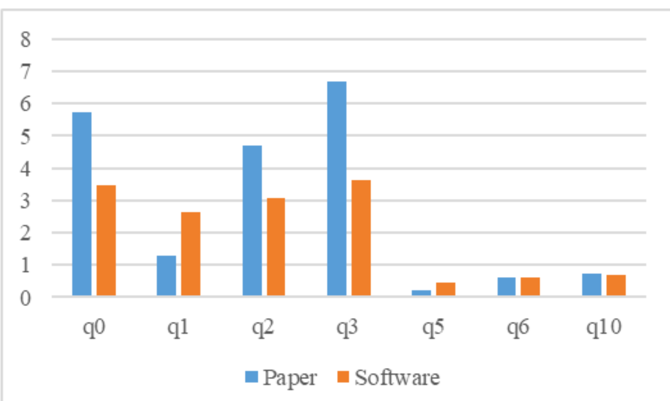

Fig. 3. Variance values for comparing modelling effort of the two treatments

The question q5 records the difficulty perceived by the subjects during goal modelling while q6 records the difficulty perceived by the subjects during domain knowledge modelling. For q5, the mean value of the difficulty perceived during goal modelling when modelling with the tool (3.6) is slightly lower than that for modelling on paper (4.1). However, the variance of the difficulty perceived by subjects during goal modelling when modelling with the tool (0.5) is noticeably higher than that for modelling on paper (0.2). For q6, the mean value of the difficulty perceived during domain modelling when modelling with the tool is slightly lower (3.5) than that for modelling on paper (3.8) but their variances are the same (0.6).

Considering all the collected data for q0 to q6 and q10, we must answer the research question RQ1 as follows: the modelling effort on paper is nearly the same as the modelling effort with the AOM4STS software tool. However, the variance values for comparing the modelling efforts of the two treatments are considerably different. In the reported study the higher variance values for the modelling effort on paper dominate as compared with the modelling effort with the tool. An explanation for this is that the tool imposes more constraints on the requirements modelling activities. The higher variances of the time for the questions q1 and q5 when using the tool require further research.

2) Evaluation of modelling effectiveness

In this sub-section, we provide an answer to the research question RQ1 addressing the modelling effectiveness, which was stated in Section V.A.1, based on the mean values represented in Fig. 4 (a) and variance values shown in Fig. 4 (b).

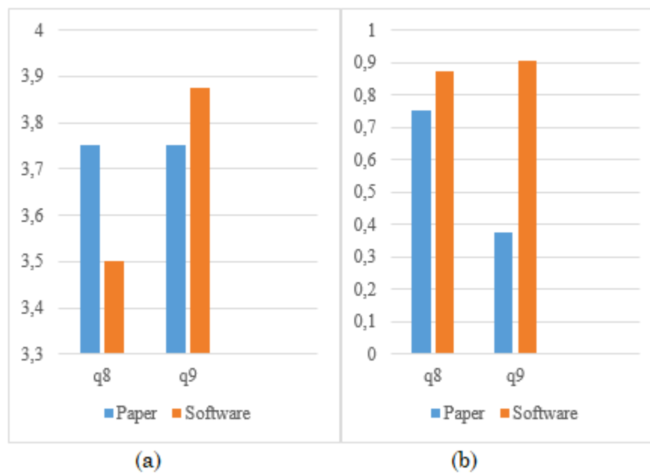

Fig. 4. (a) Mean and (b) Variance for comparing modelling effectiveness of the two treatments

The question q8 records the usefulness of goal decomposition during goal modelling. The mean value for the usefulness of goal decomposition on paper (3.75) as perceived by the subjects was slightly higher than that of modelling with the tool (3.5) with the variance of 0.75 when modelling on paper and 0.88 when modelling with the tool.

The question q9 records the utility of the concepts of goal modelling and domain knowledge modelling perceived by the subjects for requirements modelling. The mean value of the subjects who conducted modelling with the tool was 3.88 while the same value for those for those who conducted modelling on paper was 3.75. The variance of the subjects who conducted modelling with the tool was 0.91 while the same value for those who conducted modelling on paper was 0.38.

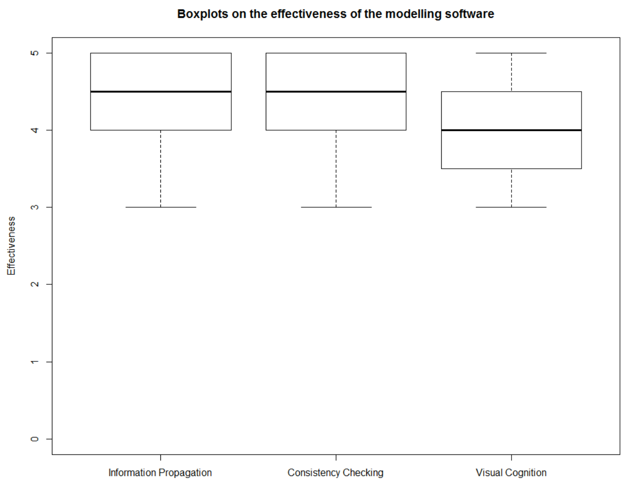

Fig. 5. Boxplot on the effectiveness of the modelling tool with respect to information propagation, consistency checking, and visual cognition

Furthermore, the subjects agreed on the effectiveness of the key features of the AOM4STS modelling tool that are not present in paper-based modelling with the median values 4.5 for information propagation, 4.5 – for consistency checking, and 4 – for visual cognition. The distribution of these results is depicted by the boxplot presented in Fig. 5. Moreover, the results in the boxplot clearly show that none of the subjects disagrees or strongly disagrees with the effectiveness of the AOM4STS modelling tool with respect to information propagation, consistency checking, and visual cognition.

Considering all the collected data from q8 and q9 and the postquestionnaire (information propagation, consistency checking, and visual cognition), we must answer the research question RQ2 as follows: the effectiveness of modelling requirements with the modelling tool was higher than the effectiveness of modelling requirements on paper except for goal decomposition which was slightly more effective when modelled on paper compared to modelling with the tool.

C. Validity Evaluation

In this experiment, there is one major threat to the internal validity [11]. This empirical study was not conducted by professionals in the industrial environment. According to [25], empirical evaluation by professionals in the real environment embraces all of the complexities of human practice in real organisations, gives stronger internal validity, and assures a more rigorous assessment of the effectiveness of the artefact. However, the research results by [26], [27] show that professionals and students perform similarly in empirical evaluations of software engineering artefacts, especially when they apply a new approach for the first time.

Concerning the external validity [11], it is highly probable that similar results will be obtained when running this experiment in a similar way with other subjects because the subjects of this experiment decided to register for the course of agent-oriented modelling of sociotechnical systems based on their interest in advanced software engineering and convenience. However, all the resources used in this experiment are publicly available in the experiment package[5] to encourage repetition of the study.

The threat to conclusion validity [11] relates to the sample size during the empirical study which involved modelling of 8 real world sociotechnical systems. According to [28], a large sample size helps to statistically observe nearly any legitimate differences between experimental conditions. Moreover, a large sample size improves the quality of research contributions. However, the systematic review of 1,700 software engineering papers published from 2001 to 2011 [29] on controlled experiments of software engineering tools with human participants reports on a large range of participants from 1 to 2,600 (the latter was a field deployment) with a median of 10 participants. Therefore, the sample size during this empirical study is very close to the median sample size of similar empirical studies.

The construct validity [11] includes two major threats. The first threat to the construct validity is that the used metrics may not be appropriate ones for evaluating the effectiveness of the modelling guidelines. For example, is "the comparison between the number of entities produced by the AOM approach and the number of the resulting entities of CPN in CPN Tools" an appropriate metric for evaluating the effectiveness of the modelling guidelines? The second threat to the construct validity is that the experiment was conducted as a part of the course, where the students were graded. This implies that the students may bias their data, as they believe that it will give them better grades. However, in the beginning of the course it was emphasised that the grade did not depend on the actual data. The grade was instead based on the completeness of the requirements, proper delivery, and the understanding of the topics expressed in the reports that were handed in by students at the end of the course.

VII. Conclusions

We conducted an empirical study with the objective of evaluating the effort and effectiveness of modelling requirements by goal models and domain models using pen and paper in comparison with the use of the modelling tool developed for engineering requirements for sociotechnical systems. The study involved experimental tasks of modelling requirements for sociotechnical systems and was conducted by postgraduate students registered for the requirements engineering course at the University of Tartu. The assessment results of experimental settings show that a short tutorial about goal modelling and domain knowledge modelling and a brief demonstration of the newly developed modelling tool were adequate. That is, these measures provided subjects with sufficient knowledge to perform adequately the modelling tasks.

The answer to the first research question lets us to conclude that the modelling effort on paper is nearly the same as the modelling effort with the AOM4STS software tool. However, the higher variance values for the modelling effort on paper dominate as compared with the modelling effort with the tool. An explanation for this is that the tool imposes more con-

[5] https://goo.gl/eVMe2B

straints on the requirements modelling activities. As the answer to the second research question, we can also conclude based on the results of the empirical study that the effectiveness of modelling requirements with the modelling tool was higher than the effectiveness of modelling requirements on paper by considering information propagation, consistency checking, and visual cognition. However, goal decomposition activity was slightly more effective when modelled on paper compared to modelling with the tool.

The answers to the research questions and particularly the answer to the second research question allow us to conclude that the support by modelling tools is essential for engineering requirements for sociotechnical systems because for such systems requirements should be modelled at different abstraction levels and from different perspectives that should be consistent with each other. In the future work, we will conduct another empirical study with a large sample size of professionals that aims to increase the confidence in the results obtained from this empirical study.

Acknowledgment. This research is partly supported by the Estonian Research Council (grant IUT20-55).

REFERENCES

[1] R. Lock and I. Sommerville, "Modelling and analysis of socio-technical system of systems," in *Engineering of Complex Computer Systems (ICECCS), 2010 15th IEEE International Conference on*, 2010, pp. 224–232.

[2] J. Marshall, "Agent-based modelling of emotional goals in digital media design projects," in *Innovative Methods, User-Friendly Tools, Coding, and Design Approaches in People-Oriented Programming*, IGI Global, 2018, pp. 262–284.

[3] B. Whitworth, "A brief introduction to sociotechnical systems," in *Encyclopedia of Information Science and Technology, Second Edition*, IGI Global, 2009, pp. 394–400.

[4] V. P. da Conceição, J. Dahlman, and A. Navarro, "What is maritime navigation? Unfolding the complexity of a Sociotechnical System," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2017, vol. 61, pp. 267–271.

[5] M. C. Davis, R. Challenger, D. N. Jayewardene, and C. W. Clegg, "Advancing socio-technical systems thinking: A call for bravery," *Appl. Ergon.*, vol. 45, no. 2, pp. 171–180, 2014.

[6] G. Baxter and I. Sommerville, "Socio-technical systems: From design methods to systems engineering," *Interact. Comput.*, vol. 23, no. 1, pp. 4–17, 2011.

[7] L. S. Sterling and K. Taveter, *The art of agent-oriented modeling*. MIT Press, 2009.

[8] T. Miller, B. Lu, L. Sterling, G. Beydoun, and K. Taveter, "Requirements elicitation and specification using the agent paradigm: the case study of an aircraft turnaround simulator," *IEEE Trans. Softw. Eng.*, vol. 40, no. 10, pp. 1007–1024, 2014.

[9] A. Norta, M. Mahunnah, T. Tenso, K. Taveter, and N. C. Narendra, "An agent-oriented method for designing large socio-technical service-ecosystems," in *Services (SERVICES), 2014 IEEE World Congress on*, 2014, pp. 242–249.

[10] K. Taveter and A. Norta, "Agile Software Engineering Methodology for Information Systems' Integration Projects," in *International Conference on Future Data and Security Engineering*, 2017, pp. 215–230.

[11] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[12] M. Brambilla, J. Cabot, and M. Wimmer, "Model-driven software engineering in practice," *Synth. Lect. Softw. Eng.*, vol. 1, no. 1, pp. 1–182, 2012.

[13] J. Whittle, J. Hutchinson, and M. Rouncefield, "The state of practice in model-driven engineering," *IEEE Softw.*, vol. 31, no. 3, pp. 79–85, 2014.

[14] P. Mohagheghi, W. Gilani, A. Stefanescu, and M. A. Fernandez, "An empirical study of the state of the practice and acceptance of model-driven engineering in four industrial cases," *Empir. Softw. Eng.*, vol. 18, no. 1, pp. 89–116, 2013.

[15] F. D. Giraldo, S. España, and Ó. Pastor, "Evidences of the mismatch between industry and academy on modelling language quality evaluation," *ArXiv Prepr. ArXiv160602025*, 2016.

[16] J. Whittle, J. Hutchinson, M. Rouncefield, H. akan Burden, and R. Heldal, "A taxonomy of tool-related issues affecting the adoption of model-driven engineering," *Softw. Syst. Model.*, vol. 16, no. 2, pp. 313–331, 2017.

[17] G. Mussbacher *et al.*, "The relevance of model-driven engineering thirty years from now," in *International Conference on Model Driven Engineering Languages and Systems*, 2014, pp. 183–200.

[18] T. Gorschek, E. Tempero, and L. Angelis, "On the use of software design models in software development practice: An empirical investigation," *J. Syst. Softw.*, vol. 95, pp. 176–193, 2014.

[19] A. Sapožnikov, "Design and Development of a Graphical Tool for Agent-Oriented Modelling. MSc thesis (in Estonian)." Tallinn University of Technology, Faculty of Information Technology, Department of Informatics, 2016.

[20] L. Sterling and K. Taveter, "Building agent-based appliances with complementary methodologies," in *Proceedings of the 2006 conference on Knowledge-Based Software Engineering: Proceedings of the Seventh Joint Conference on Knowledge-Based Software Engineering*, 2006, pp. 223–232.

[21] T. Tenso, A. H. Norta, H. Rootsi, K. Taveter, and I. Vorontsova, "Enhancing requirements engineering in agile methodologies by agent-oriented goal models: Two empirical case studies," in *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, 2017, pp. 268–275.

[22] E. S. Yu, "Towards modelling and reasoning support for early-phase requirements engineering," in *Requirements Engineering, 1997., Proceedings of the Third IEEE International Symposium on*, 1997, pp. 226–235.

[23] T. Miller, S. Pedell, A. A. Lopez-Lorca, A. Mendoza, L. Sterling, and A. Keirnan, "Emotion-led modelling for people-oriented requirements engineering: The case study of emergency systems," *J. Syst. Softw.*, vol. 105, pp. 54–71, 2015.

[24] E. R. Babbie, *The basics of social research*. Cengage Learning, 2013.

[25] J. Venable, J. Pries-Heje, and R. Baskerville, "FEDS: a framework for evaluation in design science research," *Eur. J. Inf. Syst.*, vol. 25, no. 1, pp. 77–89, 2016.

[26] D. Falessi *et al.*, "Empirical software engineering experts on the use of students and professionals in experiments," *Empir. Softw. Eng.*, pp. 1–38, 2017.

[27] I. Salman, A. T. Misirli, and N. Juristo, "Are students representatives of professionals in software engineering experiments?," in *Proceedings of the 37th International Conference on Software Engineering-Volume 1*, 2015, pp. 666–676.

[28] M. Kaptein and J. Robertson, "Rethinking statistical analysis methods for CHI," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 1105–1114.

[29] A. J. Ko, T. D. Latoza, and M. M. Burnett, "A practical guide to controlled experiments of software engineering tools with human participants," *Empir. Softw. Eng.*, vol. 20, no. 1, pp. 110–141, 2015.