Systematic derivation of conceptual models from requirements models: a controlled experiment

Sergio España, Marcela Ruiz
PROS Research Centre, Universitat Politècnica de València
{sergio.espana, lruiz}@pros.upv.es

Arturo González
DSIC, Universitat Politècnica de València, Spain
agdelrio@dsic.upv.es

Abstract—There is an open challenge in the area of model-driven requirements engineering. Model transformations that allow deriving (platform-independent) conceptual models from (computation-independent) requirements models are being proposed. However, rigorous assessments of the quality of the resulting conceptual models are needed. This paper reports a controlled experiment that compares the performance of subjects applying two different techniques for deriving object-oriented, UML-compliant conceptual models. We compare the quality of the OO-Method conceptual models obtained by applying a text-based derivation technique (which mimics what OO-Method practitioners actually do in real projects) with the quality obtained by applying a novel communication-based derivation technique (which takes as input Communication Analysis requirements models). The results show that there is an interaction between the derivation technique and the OO-Method modelling competence of the subject: the derivation technique has a significant impact on model completeness within the high-competence group. No impact has been observed on model validity. We also discuss new challenges raised by the evaluation.

Keywords- Information systems; requirements model; business process model; conceptual model; model transformation; empirical validation; controlled experiment; OO-Method; Communication Analysis; model-driven requirements engineering

I. INTRODUCTION

Enterprise information systems can be developed following the model-driven development (MDD) paradigm. This way, models that represent the information system are used to generate the software source code automatically. Several methods and tools are currently available (e.g. Accelero, Integranova, MetaEdit+, Rational Rhapsody). Given the increasing maturity of the field, MDD methods are starting to cover the requirements engineering stage by providing transformations to shift from (computation-independent) requirements models to (platform-independent and/or platform-specific) conceptual models. However, a systematic review of the literature has highlighted the lack of empirical validations of model-driven requirements engineering proposals [1]. The benefit of using model transformations in early stages of software development is yet to be investigated.

This paper takes up the open challenge of assessing the quality of model-driven requirements engineering proposals. It focuses on a specific approach that integrates Communication Analysis, a communication-oriented business process modelling and requirements engineering method [2], with the OO-Method [3], an object-oriented model-driven development framework with automatic code generation capabilities.

Previous works have presented a derivation technique that allows to systematically derive OO-Method conceptual models from Communication Analysis requirements models [4] [5] [6]. The researchers have applied the technique in four lab demos (one is presented in full detail in [7]), demonstrating the feasibility of the approach and allowing for an incremental improvement of the technique [6]. However, no rigorous assessment of the performance of analysts applying the technique and of the quality of the derived conceptual models has been performed yet.

This paper presents a controlled experiment in which we have compared the performance of subjects applying a text-based conceptual model derivation (a practice that intends to mimic what OO-Method practitioners actually do in real development projects) with the performance of subjects applying the communication-based derivation technique.

The main contributions of this paper are the following:

- An experimental design that allows comparing model-driven requirements engineering methods is provided; it operationalises the conceptual model quality framework by Lindland et al [8].
- Results are reported and discussed; we have found a significant interaction between the conceptual modelling competence and the derivation technique.

The paper is structured as follows. Section 2 reviews related works. Section 3 describes the experimental planning (experimental context, design, variables, parameters, hypotheses and procedure). Section 4 summarises the results, analysis and interpretation. Section 5 discusses issues with the potential to threaten the validity of the experiment. Section 6 further discusses the results and some lessons learned. Section 7 presents the conclusions and future work.

II. RELATED WORKS

A. Experiments in model-driven requirements engineering

According to a systematic review in model-driven requirements engineering [1], there is a lack of experimental validations in the area. Although the review indicates that 2 out of the 72 primary studies report experimental validations, we have only found evidence of one actually addressing a truly controlled experiment: Shin et al. [9] experimentally assess the effectiveness of a scenario advisor tool. However, our interest lies in investigating the quality of models produced by applying transformations to requirements models.

1 Throughout the paper, we use the term requirements model and conceptual model to distinguish these two abstraction levels.
2 CARE Technologies - Integranova Model Execution System
3 Besides a set of rules intended to be manually applied by an analyst, a model transformation has been implemented in ATL, automating the derivation as much as possible. This paper focuses on the manual derivation technique.
Fortuna et al. [10], have indeed carried out an experiment where the focus is put in model transformations. They compare the uniformity of the domain models (i.e. class diagrams) derived from Use Case models with respect to those derived from InfoCases models (an extension of the former). The experiment involved six participants. The results reveal that InfoCases can reduce the heterogeneity among several domain models derived from the same use-case model.

In contrast, our experiment adopts the conceptual model quality framework proposed by Lindland et al. [8], which has been widely used for experimental purposes, including a previous validation of Communication Analysis [11]. It distinguishes syntactic, semantic, and pragmatic quality. In this paper we focus on semantic quality, which can be further decomposed into completeness and validity. To the best of our knowledge, this is the first time that the framework is applied to assess model-driven requirements engineering methods.

In 2009 we designed and conducted an ambitious pilot experiment in University of Twente, addressing model-driven systems development [12]. We could not draw any statistically significant results but we learned many lessons that have influenced the design of the experiment reported in this paper.

B. Integration of Communication Analysis and OO-Method

Communication Analysis offers a requirements structure and several elicitation, analysis, and modelling techniques for business process and requirements engineering [2]. The Communicative Event Diagram is intended to describe business processes from a communicational perspective. A communicative event is the organisational action that is triggered as a result of a given change in the world, intended to account for that change by gathering information about it. For each communicative event, an Event Specification Template is defined; among other requirements, it contains a description of the new, meaningful information that is conveyed to the information system in the event. This is specified by means of Message Structures, a technique based on structured text.

Recently, Communication Analysis has been integrated into the OO-Method, an object-oriented MDD framework. The OO-Method conceptual model consists of four complementary views: the Object Model is an extended UML Class Diagram that specifies the information system memory; the Dynamic Model is a collection of UML State Machine Diagrams that represents the valid lifecycles of objects and interactions among them; the Functional Model declaratively specifies the reaction of the information system; the Presentation Model supports both modelling and automatic code generation (a.k.a. model compilation).

A model transformation strategy has been defined to derive OO-Method conceptual models from Communication Analysis requirements models [4] [5]. It allows deriving initial versions of the Object Model, the Dynamic Model, and the Functional Model. The conceptual model can later be modified, but the initial version already has the capability of being compiled. This way, the resulting MDD method covers the software life-cycle from requirements engineering to code generation. The derivation technique is fully described in [6].

III. EXPERIMENTAL PLANNING

A. Experimental context

The Master in Software Engineering, Formal Methods and Information Systems of the Universitat Politècnica de València (Spain) offers a last-year course named “Software Technology for Web Environments” (TAW), in which students learn to create object-oriented conceptual models for information system development. In the 2010 edition, its 31 students were trained in the OO-Method and the Integranova MDD CASE tool and, thus, they were appropriate subjects for an experiment regarding object-oriented conceptual modelling. The OO-Method complies with the Model-Driven Architecture and the Integranova model compiler can generate code not only for web applications but for desktop applications as well.

B. Experimental design

Two treatments are compared, each being a different technique for deriving conceptual models from requirements models:

- Text-based derivation. This technique takes as input a natural language requirements model and, by means of a careful reading and a linguistic analysis, facilitates the identification of actors (agents), business objects (classes of objects), operations (class services), changes in the state of objects (states and transitions), etc. This way, a conceptual model is created. This is how the OO-Method has been taught for many years and how Integranova practitioners work under conditions of practice.

- Communication-based derivation. This technique takes as input a Communication Analysis requirements model and, by applying a set of transformation rules, a conceptual model is derived. The derivation technique is described in [4] [5] [6].

In order make the comparison, we planned a multi-test within object study [13] each subject (in our case a team of students acting as surrogates for analysts) uses exactly one of the techniques to create a conceptual model and then we assess the quality of this model; we used a single case description as input for all the teams. This design implies only one round.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Treatment</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group TS</td>
<td>Text-based derivation</td>
<td>Photography</td>
</tr>
<tr>
<td>Group CA</td>
<td>Communication-based derivation</td>
<td>Agency case</td>
</tr>
</tbody>
</table>

The experimental object is the description of a photography agency. This description specifies the needs for an information system that supports the work practice of a photography agency (it also includes some organisational forms). This enterprise acts as an intermediary between photographers that provide illustrated reports and publishing houses that request and buy those reports. However, given the different nature of the two derivation techniques being compared, two versions of the object were prepared (both containing the very same information); namely, a textual requirements model and a Communication Analysis requirements model. An experienced

---

5The plan was to form two-student teams, but we had to accommodate contingencies and four teams had only one student. We acknowledge that it has not been assessed whether two-student teams had a significant advantage. In any case, by random chance, one-student teams were evenly distributed between the two treatments.

Integranova analyst created a reference solution for the case; the conceptual model has a size of 537 IFPUG function points (we applied a functional size measurement procedure using the corresponding tool of the Integranova suite).

C. Variables

We identified four types of variables [14]:

Response variables

The response variables (a.k.a. dependent variables) are related to actual effectiveness; more specifically, to the semantic quality of the conceptual models. We decompose it based on the framework by Lindland et al. [8]

- **Semantic completeness.** The metric used to measure this variable is the degree of completeness of the conceptual model (degCompl), which is the result of a careful model inspection by an expert reviewer. To do the measurement the reviewer checks the conceptual model against a template that lists each and every statement about the domain that should be included in a complete model.

- **Semantic validity.** The metric used to measure this variable is the number of validity errors in the conceptual model (errValid), as a result of the expert reviewer inspection. To do the measurement the reviewer notes down each invalid statement found in the conceptual model, while reviewing it to assess the semantic completeness.

Factors

We want to assess the impact of the following factor (a.k.a. independent variable) on the response variables.

- **Derivation technique.** Two different sets of guidelines for the derivation of conceptual models from requirements specifications are used: communication-based derivation guidelines and text-based derivation guidelines. Each set of guidelines assumes a specific requirements modelling language: textual specification and Communication Analysis model, respectively. Thus, these two requirements modelling languages have been applied to create the requirements models that are input for the derivation task.

Blocking variables

In order to take into account undesired variations in the subjects we define the following blocking variable.

- **The OO-Method modelling competence level of the team.** We expected all subjects to start the training without any prior competence in the OO-Method modelling language and the Integranova tool (in any case, this is assessed by a questionnaire). It always turns out that teams acquire different levels of modelling competence during the training. Thus, we assessed their competence to block them in two groups: high-competence and average-competence teams. The metric used to measure this variable is an indirect metric that aggregates three metrics: (i) the average of several OO-Method training exercises (contributes up to 0.2 points to the competence level), (ii) the result of a 30-question OO-Method knowledge test (up to 0.3 points), (iii) and the completeness of a conceptual model made by the team (up to 0.5 points). The result (modCompet, stands for modelling competence) is a normalised value ranging from 0 to 1.

Parameters

The variables that we do not want to influence the experimental results have been fixed. However, we measure some of them in order to ascertain the homogeneity of the subjects and, in case of inevitable heterogeneity, whether this has a significant impact on the response variables.

- **Problem domain.** The general domain is information systems analysis; more specifically, a particular business is chosen: the Photography Agency case.

- **Conceptual modelling language used for the target conceptual model: the OO-Method.**

- **Conceptual modelling tool.** Integranova is used as a tool to support the OO-Method conceptual models.

- **Requirements engineering competence.** We expect that all the subjects will start the training with a similar level of expertise in requirements engineering (in any case, this is assessed by a questionnaire).

- **Domain competence.** All the subjects will start the training without any prior competence in the photography agency domain.

- **Amount of information included in the requirements specifications.** The same information is included in both requirements specifications. The only difference lies in how it is expressed (the modelling language and the structure of the specification).

D. Hypotheses

The hypotheses formulated from the research questions defined above are the following (suffix are added to the variable name to indicate the technique: CA stands for Communication Analysis, TS stands for textual specification, High refers to the high-competence group, Avg refers to the average-competence group). Null hypotheses (H₀) and alternative hypotheses (H₁) are provided.

- **Hypothesis 1** - For high-competence teams, the text-based and the communication-based derivation techniques allow obtaining conceptual models with same degree of semantic completeness.

  \[ H_{10}: \ degCompl_{CA\_High} = degCompl_{TS\_High} \]

  \[ H_{11}: \ degCompl_{CA\_High} ≠ degCompl_{TS\_High} \]

- **Hypothesis 2** - For average-competence teams, the text-based and the communication-based derivation techniques allow obtaining conceptual models with same degree of semantic completeness.

  \[ H_{20}: \ degCompl_{CA\_Avg} = degCompl_{TS\_Avg} \]

  \[ H_{21}: \ degCompl_{CA\_Avg} ≠ degCompl_{TS\_Avg} \]
• Hypothesis 3 - For high-competence teams, the text-based and the communication-based derivation techniques allow obtaining conceptual models with same number of validity errors.

\[ H_{3p}: err\text{Valid}_{CA}\text{High} = err\text{Valid}_{TS}\text{High} \]

\[ H_{3i}: err\text{Valid}_{CA}\text{High} \neq err\text{Valid}_{TS}\text{High} \]

• Hypothesis 4 - For average-competence teams, the text-based and the communication-based derivation techniques allow obtaining conceptual models with same number of validity errors.

\[ H_{4p}: err\text{Valid}_{CA}\text{Avg} = err\text{Valid}_{TS}\text{Avg} \]

\[ H_{4i}: err\text{Valid}_{CA}\text{Avg} \neq err\text{Valid}_{TS}\text{Avg} \]

Table II shows the mean value and standard deviation of the group knowledge of and experience in the specification techniques. We consider that a subject reports having good knowledge of the syntax of a specification technique when s/he rates it over 3 in the 5-point Likert scale. We make a similar consideration for the experience in applying the techniques. Results indicate that, in general, the subjects report not being very knowledgeable of the syntax of the techniques. The better-known techniques are the Class Diagram and the Relational Model; 15 subjects (57.7% of the total amount of subjects) have reported having a moderately high or a high knowledge level of their syntax.

With regards to their experience in applying the techniques, the Class Diagram has been used by 16 subjects (61.5%) to solve moderately complex cases or even real cases in a professional environment. Use Case Diagrams, Relational Model and Entity Relationship Diagrams have also been used in moderately complex or real cases by 13 subjects (50.0%), 12 subjects (46.2%) and 10 subjects (38.5%), respectively.

We offer three charts showing the distribution of knowledge and experience ratings. According to our experience, the chart that corresponds to the Data Flow Diagram is quite representative of what master students usually claim to know about methods and techniques they have been taught during their studies.

Both the knowledge and the experience follow a bell-shaped curve, the mode and the mean being around 3 (this varies among the techniques; see the means in Table II). This means that they consider that they learnt the syntax good enough to pass their exams, and that they have applied them to solve small exercises that lack the complexity of real cases professionally.

E. Experimental procedure

The experimental procedure is depicted in Figure 1 and explained next.

Demographic questionnaire

A demographic questionnaire, which was supported by SurveyGizmo (a website that offers support to online surveys), was applied with the purpose of identifying the background and experience using different specification techniques, such as Data Flow Diagram (DFD), Entity Relationship Diagram (ERD), Activity Diagram (AD), Class Diagram (CD), Use Case Diagram (UCD), State Transition Diagram (STD), Business Process Modeling Notation (BPMN), Relational Model for database design (RM) and Communication Analysis models (CA). The subjects rated in a 5-point Likert scale their knowledge level of the syntax of the different specification techniques (where 1 means “low” and 5 means “high”). They also rated their experience in applying the techniques (where 1 means “I have never use this technique” and 5 means “I have solved real cases professionally”).

![Figure 1. Experimental procedure](image1)

![Figure 2. Distribution of knowledge and experience in the Data Flow Diagram](image2)

### Table II. Descriptive Statistics of Demographic Data (N=26)

<table>
<thead>
<tr>
<th>Knowledge in</th>
<th>DFD</th>
<th>ERD</th>
<th>AD</th>
<th>CD</th>
<th>UCD</th>
<th>STD</th>
<th>BPMN</th>
<th>RM</th>
<th>CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.04</td>
<td>3.58</td>
<td>2.81</td>
<td>3.85</td>
<td>3.69</td>
<td>3.12</td>
<td>2.12</td>
<td>3.69</td>
<td>1.73</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>0.91</td>
<td>1.02</td>
<td>1.05</td>
<td>1.08</td>
<td>1.01</td>
<td>1.07</td>
<td>1.30</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>Rating &gt; 3 (num. and %)</td>
<td>8</td>
<td>11</td>
<td>4</td>
<td>15</td>
<td>14</td>
<td>7</td>
<td>6</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>Experience in</td>
<td>DFD</td>
<td>ERD</td>
<td>AD</td>
<td>CD</td>
<td>UCD</td>
<td>STD</td>
<td>BPMN</td>
<td>RM</td>
<td>CA</td>
</tr>
<tr>
<td>Mean</td>
<td>2.92</td>
<td>3.46</td>
<td>2.50</td>
<td>3.65</td>
<td>3.38</td>
<td>2.77</td>
<td>1.92</td>
<td>3.46</td>
<td>1.54</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>0.84</td>
<td>0.94</td>
<td>1.02</td>
<td>0.79</td>
<td>0.94</td>
<td>0.90</td>
<td>1.32</td>
<td>0.94</td>
<td>0.81</td>
</tr>
<tr>
<td>Rating &gt; 3 (num. and %)</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>16</td>
<td>13</td>
<td>5</td>
<td>3</td>
<td>12</td>
<td>1</td>
</tr>
</tbody>
</table>
The distribution for the Class Diagram indicates a higher knowledge level and practical experience; this is convenient for our experiment since we expect them to create OO-Method conceptual models and we will place special interest in the Object Model (an extended Class Diagram). The Use Case Diagram and, to a lesser extent, the Entity Relationship Diagram and the Relational Model have a similar distribution.

Figure 3. Distribution of knowledge and experience in the Class Diagram

The knowledge level of the subjects in the Communication Analysis modelling techniques is, as expected, very low. The reason is the fact that it is hardly known; unless the students have attended the course in which it is taught, the chances they have ever heard of the method are very low. Surprisingly, the Business Process Modelling Notation has a similar distribution.

![Distribution of knowledge and experience in Communication Analysis](image1)

Figure 4. Distribution of knowledge and experience in Communication Analysis

It must be noticed that, among the 26 respondents, 6 subjects have been taught Communication Analysis in the past. Although only one of them reports having a good knowledge level of the syntax of Communication Analysis modelling techniques and having applied them in moderately complex cases, we want to take advantage of their knowledge in our experiment, as explained later on.

**OO-Method training**

Then the students were trained in the OO-Method, covering the concepts and modelling primitives of the Object Model, the Dynamic Model and the Functional Model. The students were also trained in using Integranova to create conceptual models. The training lasted 21 hours over 12 sessions (6 weeks), including theory and exercises with and without the tool.

**OO-Method modelling competence assessment**

After the OO-Method training, the students were evaluated in order to assess their knowledge level and their modelling competence. To assess the knowledge level of the subjects in the OO-Method (actually, knowledge and comprehension according to Bloom’s taxonomy [15]) a 30-question test was used. To assess their modelling competence (covering the application, analysis and -to some extent- synthesis levels of Bloom’s taxonomy), they were handed out a small\(^6\) case description and they created its corresponding conceptual model. To create this conceptual model they freely teamed up in pairs, and the same teams were maintained for the rest of the experiment. Each team has an identifier team number (ranging from 1 to 17). Two researchers assessed the completeness of the models. As explained above, we aggregated in a variable the results of the training exercises, the knowledge test and the model quality assessment. This variable allowed us to classify teams into two levels of OO-Method modelling competence: high and average. We considered a team to have a high OO-Method modelling competence when the value of the competence variable is higher than or equal to 0.75, whereas the team has an average modelling competence then the value of the competence variable is lower than 0.75. Since the variable ranges from 0 to 1 but teams having a score lower than 0.5 would be discarded, 0.75 constitutes a reasonable threshold.

**Allocation of teams to treatment groups**

Then the teams were split into two treatment groups: group $T_S$ and group $C_A$ (whose names stand for textual specification and Communication Analysis, respectively). We applied blocking in order to be able to compare the effect of the OO-Method modelling competence on conceptual model quality, as well as the interaction of the effects of the OO-Method modelling competence and the derivation technique. For this comparison to be possible, we ensured balance between both treatment groups.

A purely random and balanced allocation can be done as follows. A list of the subjects is made (actually teams, in our case); the ordering criteria of the list (if any) are irrelevant. A random sequence of distinct numbers is generated (e.g. using the services of random.org, a website that generates random numbers using atmospheric noise). Each team in the list is assigned a number of the randomly generated sequence (the first team in the list is assigned the first random number in the sequence, and so on). Then, the list of teams is ordered according to the random number. Finally, each team of the list is alternatively allocated to a group (the first team to first group, the second team to the second group, the third team to the first group again, and so on). In order to decide which group is allocated a team in the first place, a coin is flipped.

If blocking is intended, then the procedure is similar, but the teams are first split into the different blocks according to the blocking parameter (we defined two levels of OO-Method modelling competence, in our case). The above mentioned allocation procedure is performed for each of the blocks separately.

These procedures are ideal; when working with human subjects, reality imposes some constraints that need to be dealt with flexibly, trying not to compromise or bias the allocation. In our case we needed to purposely allocate some teams to a specific treatment group (referred to as fixed allocation). Five teams in which at least one of the subjects knew about Communication Analysis were allocated in treatment group CA straight away. The rationale is that these subjects had already been trained in Communication Analysis, including some heuristic guidelines concerning the derivation of data models; thus, had they been allocated to the treatment group

---

\(^6\) The model used as a reference solution has 5 classes, 16 attributes, 36 services and 5 relationships.
TS, we would not have been able to prevent them from applying their knowledge of these guidelines instead of the textual-based derivation guidelines. Additionally, when we informed that the class would be split into two groups with different dates for the training sessions, one team reported not being able to assist one of the dates. In any case, we managed to ensure the balance between both treatment groups, also taking into account the blocking of competence levels (see [6] for further details and an account for every decision taken).

Treatment: training in a derivation technique

Then each treatment group was trained in a different conceptual model derivation technique:

- The teams in group TS were trained in text-based derivation of conceptual models.
- The teams in group CA were trained in communication-based derivation of conceptual models.

In both trainings, derivation guidelines and heuristics are offered. First the concepts and the guidelines are explained and illustrative examples are offered, then the subjects solve small exercises. Once the subjects know the derivation technique, they apply it to a mid-sized example; namely, the SuperStationery case. The solution to this case is discussed later on in class. The training for each group amounted 3.5 hours over two sessions (1 week).

Experimental task: derivation of a conceptual model

After the training sessions, both groups were merged once again and they were assigned a case to solve; namely, the Photography Agency case. The input for this task is different for each treatment group. Group CA receive a Communication Analysis requirements model of the case, whereas group TS receive a textual requirements model of the case. Both models contain the same information but expressed differently. The time allotted for deriving the conceptual model of the case and creating it with the Integranova Modeler was 7 hours over 4 sessions (2 weeks). After each session, the teams had to send a snapshot of their model to the researchers. The models sent after the fourth session are considered the actual result from this task and are the main artefact analysed in the experiment.

Post-task questionnaire and discussion

After the modelling task, the subjects answer a questionnaire along with a discussion to gather some feedback on the derivation technique. This post-task questionnaire was supported online by SurveyGizmo and asked for the three bigger advantages of applying the derivation technique (compared to not having any derivation technique at all), the three main drawbacks of the derivation technique (compared to having complete freedom and lack of rules for creating the conceptual model), and in which aspects the subjects thought the derivation technique could be improved (for instance, indicating what guidelines they missed in the derivation technique, or which modelling primitives of the conceptual model were the most difficult to derive from the requirements model). The aim was to elicit some feedback from the subjects. We later treated the qualitative data to draw some conclusions.

IV. Results analysis and interpretation

a) OO-Method modelling competence of the teams

The OO-Method modelling competence of the teams was calculated by means of metric numCompetNor, which normalises the numeric indicator between 0 and 1. Table III shows the descriptive statistics of the competence level, both disregarding the treatment (for the whole sample of teams) and taking it into account (splitting by treatment). Note that one team was not assessed (it seemed a dropdown at that moment).

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>Mean</th>
<th>Std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>modCompet</td>
<td>16</td>
<td>0.7196</td>
<td>0.09439</td>
</tr>
<tr>
<td>modCompet_CA</td>
<td>8</td>
<td>0.7341</td>
<td>0.05554</td>
</tr>
<tr>
<td>modCompet_TS</td>
<td>8</td>
<td>0.7050</td>
<td>0.12459</td>
</tr>
</tbody>
</table>

Figure 5 shows the boxplot of the modelling competence, depending on the treatment. The outlier in group TS corresponds to a team that did not perform well in the competence assessment.

![Figure 5](image-url)

We expect that the difference between the modelling competences of the teams belonging to the two treatment groups is not significant. Otherwise, it would be a threat to the validity of the treatment comparison. To compare the means an independent-samples T-test can be applied, as long as its three assumptions are fulfilled.

Assumption 1: The two samples are independent of one another. Given the experimental design, both treatment groups are independent samples (each team was trained in only one derivation technique).

Assumption 2: The data from the two samples are both normally distributed. We apply the Shapiro-Wilk test of normality to modCompet and it indicates that the distribution is normal both for group TS ($p=0.281>0.05$) and group CA ($p=0.925>0.05$). The Kolmogorov-Smirnov test of normality is usually applied to larger samples; in any case it also indicated that the distribution is normal ($p=0.200>0.05$).

Assumption 3: The two samples have approximately equal variance on the dependent variable. According to the Levene test for the equality of variances, both competence levels have
approximately equal variance on the response variable treatment ($p=0.134>0.05$).

Since the three assumptions are fulfilled, we apply the test and it indicates that the difference between the means is not significant ($t=0.603$ and $p=0.556>0.05$). In other words, the difference between the modelling competences of the teams belonging to the two treatment groups is not significant. Thus, we can be confident that the allocation did not introduce any bias with regards to the modelling competence.

b) Actual effectiveness of the derivation techniques

An expert reviewer used the reference model and a correction template to review the conceptual models. The conceptual models of the 17 teams where thoroughly reviewed; the mean review time were 1:03:30 (ca. 1 hour and 3 minutes). The measures obtained by the expert reviewer are analysed in the following. See Table IV.

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>Mean</th>
<th>Std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of completeness</td>
<td>13</td>
<td>0.6815</td>
<td>0.07616</td>
</tr>
<tr>
<td>Validity errors</td>
<td>13</td>
<td>0.6316</td>
<td>0.08585</td>
</tr>
</tbody>
</table>

The degree of conceptual model completeness ($deg\text{Compl}$) is 68.15% for the communication-based derivation technique and 64.36% for the text-based derivation technique. With regards to validity errors ($err\text{Valid}$), the conceptual models created by the teams applying communication-based derivation had an average of 15.56 errors, whereas those created by teams applying text-based derivation had an average of 20.50 errors. However, it is necessary to further analyse the significance of these differences.

However, as discussed above, we should take into account the blocking and analyse separately the high-competence teams and the average-competence teams.

**Completeness of the models (blocked comparison)**

We investigate now the difference in the means of the degree of completeness but now blocking by modelling competence level. We compare the performance of teams with high and average OO-Method modelling competence separately. Table V shows the descriptive statistics.

<table>
<thead>
<tr>
<th>Competence level</th>
<th>Measure</th>
<th>N</th>
<th>Mean</th>
<th>Std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>$deg\text{Compl}_{CA, High}$</td>
<td>4</td>
<td>0.7439</td>
<td>0.05256</td>
</tr>
<tr>
<td></td>
<td>$deg\text{Compl}_{TS, High}$</td>
<td>4</td>
<td>0.6517</td>
<td>0.05191</td>
</tr>
<tr>
<td>Average</td>
<td>$deg\text{Compl}_{CA, Avg}$</td>
<td>5</td>
<td>0.6316</td>
<td>0.05027</td>
</tr>
<tr>
<td></td>
<td>$deg\text{Compl}_{TS, Avg}$</td>
<td>4</td>
<td>0.6355</td>
<td>0.11969</td>
</tr>
</tbody>
</table>

It appears that the difference in the effect of the derivation technique over the degree of completeness is bigger for high-competence teams than for average-competence teams. Within the high competence level, the models created by the teams applying communication-based derivation (CA) have a mean degree of completeness of 74.38% whereas those created by the teams applying text-based derivation (TS) have a mean degree of completeness of 65.17%; there is a difference of 9.22% in favour of the communication-based derivation technique. Within the average competence level, the mean degree of completeness for group CA is 63.16% whereas for group TS is 63.55%; there is a difference of 0.38% in favour of the text-based derivation technique. Figure 6 shows the corresponding boxplots.

![Boxplot of the model completeness (deg\text{Compl}) for both treatment groups, blocking by competence level](image)

**High-competence teams**

To investigate whether the difference for the high competence level is significant we intend to apply the independent-samples T-test, which relies on three assumptions.

Assumption 1: The two samples are independent of one another. This is true, see above.

Assumption 2: The data from the two samples are both normally distributed. The sample is too small to outline a bell curve, so we omit analysing the histograms and perform a more formal test to assess normality. The Shapiro-Wilk test on normality indicates that the distribution is normal (for TS, $p=0.771>0.05$; for CA, $p=0.187>0.05$).

<table>
<thead>
<tr>
<th>Derivation technique</th>
<th>Kolmogorov-Smirnov*</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>TS</td>
<td>0.224</td>
<td>4</td>
</tr>
<tr>
<td>CA</td>
<td>0.282</td>
<td>4</td>
</tr>
</tbody>
</table>

Assumption 3: The two samples have approximately equal variance on the dependent variable. According to the Levene test for the equality of variances (see Table VII), both competence levels have approximately equal variance on the response variable treatment ($p=0.789>0.05$).

Since the three assumptions are fulfilled, we can apply an independent samples T-test to verify the null hypothesis $H_{10}$. As shown also in Table VII, the test results indicate that the difference is indeed significant ($p=0.047<0.05$). Thus the null
hypothesis $H_{10}$ is refuted with a 95% confidence and the alternative hypothesis $H_{11}$ is corroborated. This means that the communication-based derivation technique, when applied by analysts with a high competence in OO-Method modelling, allows obtaining conceptual models with greater degree of completeness than the text-based derivation technique.

**TABLE VII.** MEAN COMPARISON FOR THE DEGREE OF COMPLETENESS FOR THE HIGH COMPETENCE LEVEL (INDEPENDENT SAMPLES T-TEST FOR degCompl)

<table>
<thead>
<tr>
<th>Levene's test for equality of variances</th>
<th>T-test for equality of means</th>
<th>95% confidence interval of the difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>Sig.</td>
<td>$t$</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>$-$</td>
<td>0.079</td>
<td>0.789</td>
</tr>
<tr>
<td>$-$</td>
<td>2.495</td>
<td>5.99</td>
</tr>
</tbody>
</table>

= Equal variances assumed, # Equal variances not assumed

**Average-competence teams**

With regards to the average competence level, the T-test cannot be applied because one of the two samples does not have a normal distribution (see Table VIII); that is, the sample for group CA does not pass the Shapiro-Wilk test of normality (for TS, $p=0.304>0.05$; but for CA, $p=0.029<0.05$).

**TABLE VIII.** TESTS OF NORMALITY FOR THE DEGREE OF COMPLETENESS (degCompl) FOR THE AVERAGE COMPETENCE LEVEL

<table>
<thead>
<tr>
<th>Derivation technique</th>
<th>Kolmogorov-Smirnov*</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>$df$</td>
</tr>
<tr>
<td>TS</td>
<td>0.224</td>
<td>4</td>
</tr>
<tr>
<td>CA</td>
<td>0.381</td>
<td>5</td>
</tr>
</tbody>
</table>

* Lilliefors Significance Correction

Therefore, we apply a non-parametric test for unpaired samples that allows for non normal distributions (see Table IX). The Mann-Whitney U test indicates that the difference is not significant ($p>0.05$) so we cannot refute $H_{20}$ that is, we cannot conclude that the derivation technique has a significant effect on the degree of completeness of the models created by analysts of an average OO-Method modelling competence.

**TABLE IX.** RESULTS OF THE MANN-WHITNEY U TEST ON THE DEGREE OF COMPLETENESS (degCompl) FOR THE AVERAGE COMPETENCE LEVEL

<table>
<thead>
<tr>
<th>Ranks</th>
<th>Derivation technique</th>
<th>N</th>
<th>Mean rank</th>
<th>Sum of ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TS</td>
<td>4</td>
<td>5.25</td>
<td>21.00</td>
</tr>
<tr>
<td></td>
<td>CA</td>
<td>5</td>
<td>4.80</td>
<td>24.00</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Test Statistics$^b$

<table>
<thead>
<tr>
<th></th>
<th>Mann-Whitney U</th>
<th>Wilcoxon W</th>
<th>Z</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>Exact Sig. (2*tailed Sig.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9.000</td>
<td>24.000</td>
<td>-0.246</td>
<td>0.806</td>
<td>0.905$^a$</td>
</tr>
</tbody>
</table>

$^a$ Not corrected for ties. $^b$ Grouping variable: derivation technique

**Discussion**

Summing up, we can conclude that the communication-based derivation technique has a significant effect on the degree of completeness of the derived conceptual models, but only when the team of analysts has a high competence level on the conceptual modelling technique. This result can be explained by our observations during the experiment on how the subjects interiorised the derivation technique. Those subjects with a higher competence in the OO-Method seemed to grasp the concepts and guidelines of the communication-based derivation technique better that those subjects with lower competence. It may be due to the fact that highly the outstanding modellers are those analysts with better capacity for abstraction; they would be better suited to take advantage of the systematic derivation rules. On the contrary, the subjects in the average competence level could be those with more difficulty to apply abstract thinking; if that was so, they could sense the communication-based derivation technique as a burden, yet another complex task to solve. We acknowledge that this just an ad-hoc hypothesis that needs to be further investigated and, eventually, perhaps corroborated.

**VALIDITY OF THE MODELS (BLOCKED COMPARISON)**

We investigate now the difference in the means of the degree of completeness blocking by modelling competence level. We compare the performance of teams with high and average OO-Method modelling competence separately. Table X shows the descriptive statistics. The means show differences favourable to the communication-based derivation technique (see also Figure 7), but their significance needs to be assessed.

**TABLE X.** DESCRIPTIVE STATISTICS OF THE VALIDITY ERRORS (errValid), BLOCKING BY COMPETENCE LEVEL

<table>
<thead>
<tr>
<th>Competence level</th>
<th>Measure</th>
<th>N</th>
<th>Mean</th>
<th>Std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>errValid_CA_High</td>
<td>4</td>
<td>16.00</td>
<td>11.690</td>
</tr>
<tr>
<td></td>
<td>errValid_TS_High</td>
<td>4</td>
<td>21.75</td>
<td>20.172</td>
</tr>
<tr>
<td>Average</td>
<td>errValid_CA_Avg</td>
<td>5</td>
<td>15.20</td>
<td>8.758</td>
</tr>
<tr>
<td></td>
<td>errValid_TS_Avg</td>
<td>4</td>
<td>19.25</td>
<td>5.123</td>
</tr>
</tbody>
</table>

**Figure 7.** Boxplot of the validity errors (errValid) for both treatment groups, blocking by competence level
High-competence teams

We first investigate the teams with a high competence level. The distribution of the validity errors is not normal (see Table XI), according to the Shapiro-Wilk test of normality (it fails for TS, \( p = 0.003 > 0.05 \)).

Therefore, we apply a non-parametric test for unpaired samples that allows for non normal distributions (see Table XII). The Mann-Whitney U test indicates that the difference is not significant (\( p > 0.05 \)) so we cannot refute \( H_{3b} \) that is, we cannot conclude that the derivation technique has a significant effect on the validity errors of the models created by analysts of a high OO-Method modelling competence.

<table>
<thead>
<tr>
<th>Derivation technique</th>
<th>Kolmogorov-Smirnov*</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>TS</td>
<td>0.436</td>
<td>4</td>
</tr>
<tr>
<td>CA</td>
<td>0.318</td>
<td>4</td>
</tr>
</tbody>
</table>

a. Lilliefors Significance Correction

TABLE XII. RESULTS OF THE MANN-WHITNEY U TEST ON THE VALIDITY ERRORS (errValid) FOR THE AVERAGE COMPETENCE LEVEL

<table>
<thead>
<tr>
<th>Derivation technique</th>
<th>N</th>
<th>Mean rank</th>
<th>Sum of ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS</td>
<td>4</td>
<td>5.00</td>
<td>20.00</td>
</tr>
<tr>
<td>CA</td>
<td>4</td>
<td>4.00</td>
<td>16.00</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Test Statisticsb

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>6.000</td>
<td></td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>16.000</td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>-0.581</td>
<td></td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>0.561</td>
<td></td>
</tr>
<tr>
<td>Exact Sig. [2*(1-tailed Sig.)]</td>
<td>0.686c</td>
<td></td>
</tr>
</tbody>
</table>

a. Not corrected for ties
b. Grouping variable: derivation technique

Average-competence teams

We now investigate the teams with an average competence level. We intend to apply a parametric test to compare the means.

Assumption 1: The two samples are independent of one another. This is true, see above.

Assumption 2: The data from the two samples are both normally distributed. The Shapiro-Wilk test on normality indicates that the distribution is normal (see Table XIII; for TS, \( p = 0.594 > 0.05 \); for CA, \( p = 0.755 > 0.05 \)).

TABLE XIII. TESTS OF NORMALITY FOR THE VALIDITY ERRORS (errValid) FOR THE AVERAGE COMPETENCE LEVEL

<table>
<thead>
<tr>
<th>Derivation technique</th>
<th>Kolmogorov-Smirnov*</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>TS</td>
<td>0.237</td>
<td>4</td>
</tr>
<tr>
<td>CA</td>
<td>0.181</td>
<td>5</td>
</tr>
</tbody>
</table>

a. Lilliefors Significance Correction

* This is a lower bound of the true significance.

Assumption 3: The two samples have approximately equal variance on the dependent variable. According to the Levene test for the equality of variances (see Table XIV), both competence levels have approximately equal variance on the response variable treatment (\( p = 0.245 > 0.05 \)).

TABLE XIV. MEAN COMPARISON FOR THE VALIDITY ERRORS FOR THE HIGH COMPETENCE LEVEL (INDEPENDENT SAMPLES T-TEST FOR errValid)

<table>
<thead>
<tr>
<th>Levene’s test for equality of variances</th>
<th>T-test for equality of means</th>
<th>95% confidence interval of the difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>TS</td>
<td>1.607</td>
<td>0.245</td>
</tr>
<tr>
<td>CA</td>
<td>0.865</td>
<td>0.555</td>
</tr>
</tbody>
</table>

= Equal variances assumed; ≠ Equal variances not assumed

Since the three assumptions are fulfilled, we can apply an independent samples T-test to verify the null hypothesis \( H_0 \). As shown also in Table XIV, the test results indicate that the difference is not significant (\( p = 0.443 > 0.05 \)). Thus the null hypothesis cannot be refuted. This means that we cannot conclude that the derivation technique has a significant effect on the validity errors of the models created by analysts of an average OO-Method modelling competence.

Discussion

Our procedure for measuring semantic completeness was very thorough: we assessed each and every requirement that ought to be represented in the conceptual model. On the contrary, our procedure for measuring validity we less systematic: we only counted invalidities that we came across while measuring completeness; we expect to have found many of them but we cannot ensure that all invalidities were spotted. We can assume that we found approximately the same percentage of invalidities in each model; however, we can neither be sure of this. Therefore, there are threats to the validity of the results concerning model validity (if you will forgive the repetition).

Despite the non-significant statistical results, we still have two ad-hoc hypotheses that need to be further investigated and, eventually, perhaps corroborated.

A possible explanation is that the communication-based derivation technique is still not covering all the primitives of the OO-Method and, whenever the subjects intend to increase the completeness of their models and try to model some statements about the domain for which the derivation technique does not offer guidelines, they make mistakes that increase the invalidity. Improving and extending the communication-based derivation technique could lead to a significant difference in the validity of the models.

Or it could actually occur that, if a more thorough procedure for measuring validity is followed, then the results will become significantly different. The difficulty lies in the fact that covering all invalidities is too time-consuming when models as big as the Photography Agency are considered. Such an investigation should be made using a smaller case.
V. EVALUATION OF THE EXPERIMENT VALIDITY

This section discusses issues with the potential to threaten the validity of the experiment [13].

A. Conclusion validity

A random heterogeneity of subjects could give rise to greater variability in the measures, due to confounding factors. However, we verified that the subjects had an homogeneous background by means of a questionnaire, so there is no threat. As a trade-off, homogeneity limits external validity.

However, due to different OO-Method modelling competences of the subjects, we opted for applying blocking: we classified teams into two levels of modelling competence. This way we could assess the effect of the modelling competence in the quality of the resulting models, as well as investigate the interaction of the modelling competence with the treatment (the conceptual model derivation technique). This provides a more detailed analysis of the mechanisms in play.

The proposed measures related to actual efficacy have been widely used by many researchers, since they come from a well-known quality framework by Lindland et al. [8]. However, the instruments used to measure the quality are subject to threats (see below). Also, each model was only assessed by one reviewer; he has been applying the OO-Method since 2005 to develop software applications, to teach conceptual modelling and for research purposes, so he is competent enough for the purposes of the experiment. We acknowledge that it is convenient to have the models reviewed by more expert reviewers and to calculate the inter-reviewer agreement. This was not possible due to resource constraints. Note that 17 models where thoroughly reviewed; the mean review time was 1:03:30 (ca. 1 hour and 3 minutes). We plan, however, to research on different ways of reviewing the models so as to decrease the review time, while still obtaining a precise measure of the quality.

B. Internal validity

Instrumentation is the effect caused by the instruments used in the experiment. To avoid it, we used SurveyGizmo as a support for web-based forms. This allows enforcing the compulsoriness of certain answers and minimises transcription errors (since the results can be directly downloaded as a spreadsheet). There is a risk of the subjects not understanding textual parts of the instruments correctly, due to the fact that it is in English and the mother tongue of the subjects is Spanish. Our impression during the experiment operation was that some aspects of the case description were not fully grasped. Whenever we found a team having a misconception due to a misinterpretation of the text, we helped them understand the meaning of the text, but some misconceptions may have remained unnoticed. There is a risk related to the allocation of subjects to groups. We applied a random allocation procedure mixed with fixed allocations. Although complex, this procedure was designed with the honest intention of avoiding bias. Since in this experiment the treatment is a method, the only sensible decision with respect to those teams that already knew one of the methods was to allocate them to the group that received that treatment. Otherwise, their combined knowledge of both methods would have had an uncontrolled influence on the results. With respect to the only team that requested being allocated to a specific treatment, it should be noted that the team did not know at that moment, what the treatment was; their request was due to their unavailability to attend the training corresponding to the other treatment. Since we were aware of their commitment to the course, we did not sense in this action any threat to the validity. It is simply a contingency related to using students as experimental subjects.

C. Construct validity

Two of the researchers involved in the experiment are authors of Communication Analysis who have expectancies about this method performing better. In order to reduce the threat of bias, two experimenters without expectancies have been involved. The experiment includes a single information system description so it may under-represent the construct of all information systems.

D. External validity

With respect to the use of students as experimental subjects, several works suggest that, to a great extent, the results can be generalised to industry practitioners [16]. In any case, we are aware that more experiments with a larger number of subjects are necessary and, that using practitioners would lead to more reliable results. We thoughtfully selected a representative problem statement. However, more empirical studies with other requirements specifications are necessary. Although it is a big effort, we are currently building a repository of cases to use in experiments.

VI. DISCUSSION

The experiment has allowed us to test the conceptual model derivation technique with subjects that have learnt the OO-Method for the first time. The subjects teamed up in pairs and were trained in the OO-Method. Since this method, although it is based in the UML, contains specific languages such as the Functional Model, its training takes several sessions. Our many-years experience in OO-Method education and training shows that not all students/practitioners grasp the concepts and guidelines easily. This is a risk when attempting to use inexperienced analysts as experimental subjects. We are not actually investigating the features of the OO-Method in the experiment, but of the compared derivation techniques; however, the conceptual modelling method is the foundation stone of the experiment and the subjects are required at least a minimum expertise in it. For this reason, we assessed the OO-Method modelling competence of each team after the training. We differentiate two blocks: teams with a high competence and teams with an average competence. Then the teams were allocated into two treatment groups, preserving the blocking. The allocation procedure achieved more balanced groups. The treatment consisted in training the teams in one of the two conceptual model derivation techniques being compared: either text-based derivation or communication-based derivation. During the experimental task, the teams derived an OO-Method conceptual model from a requirements model. The semantic quality attributes measured in the models were completeness and validity.
The analysis of results has shown that, with regards to high-
competence teams, there is a significant difference between
semantic completeness of the OO-Method conceptual models
created by teams that applied the communication-based
derivation technique and those created by teams applying the
text-based derivation technique. The more requirements from
the Photography Agency that are supported by the conceptual
model, the higher the semantic completeness. This statistical
result coincides with our appreciation that the teams applying
the technique were acting more thoroughly and systematically
than those of the other block, which we interpret that has led to
the higher completeness.

The difference in model completeness within the average-
competence teams, although again favourable to the
communication-based derivation technique, is not statistically
significant. A hypothesis for this phenomenon is that, unless a
team has enough competence to feel comfortable with OO-Method modelling, the derivation guidelines entail an
excessive cognitive workload. Some teams with an average
competence level got stuck during the experimental task, trying
to find out how to express a precondition or a transaction
formula. Their inexperience seems to have led them to spend
more time than highly competent teams to construct the
conceptual model. Since the duration of the experimental task
was restricted, some average-competence teams reported
having to prioritise in order to first attempt to deal with
requirements that they were confident to be able to express.

In any case, we claim that the teams of the high-
competence block have more similarities to real OO-Method
practitioners than the teams of the average-competence block
(actually real practitioners are far more competent that the most
competent team participating in the experiment). Thus, we are
confident that their performance is closer to the one that would
have resulted if subjects had been real practitioners.

Besides the quantitative analysis, some interesting
qualitative information was also gathered during the
experiment. We noted down which aspects of the technique
were more difficult to grasp by the subjects, which
misconceptions were more frequent. Also, after the task, an
open discussion was held so as to gather some feedback
regarding the experience. Additionally, a survey with open
questions asked for benefits and drawbacks of the derivation
technique. This qualitative information has been analysed and
categorised in order to conclude which conceptual modelling
primitives the derivation technique facilitates identifying, and
which are the primitives for which some improvement is
needed. For each subject, it was identified which modelling
primitives were mentioned (in a positive or in a negative
sense), and totalisations were made. Figure 8 shows the
perceptions of the 14 subjects that applied the communication-
based derivation technique with regards to the strengths and
weaknesses of the approach. The green bars indicate how many
subjects mentioned that the technique helped them to identify a
given conceptual modelling primitive. For instance, 10 subjects
reported that it facilitated the identification of classes; the
modelling of attributes, relations and agents was also benefited.
For instance, one subject commented that “It is clear how to
structure classes and attributes.” On the other hand, it became
evident that there was still space for improvement, as shown by
the red bars. 5 subjects reported to have difficulties in
identifying restrictions such as relationship cardinalities and
service preconditions; the subjects also experienced trouble
identifying services. Many subjects also appreciate that the
approach is systematic: “the communicative event diagram is
very useful to define the order in which the model is created.”

Figure 8. Subjective feedback regarding the communicative-based technique

Another interesting remark made by the subjects was the fact
that the “derivation rules restrict many design decisions”
and limit the modeller creativity. On the other hand, many
subjects expressed that they had found the technique intuitive.

Figure 9. Fragment of a model resulting from the incorrect application
of class-extension derivation rules

According to our observations and the subjects’ feedback,
the guidelines that the teams had more trouble to grasp and
apply correctly are, by far, those related to class extension.
Note for instance the fragment of a model shown in Figure 9
where the team created a separate class for each communicative
event affecting the exclusive report business object, instead of
extending the class derives in the locally
initiatory event; namely ExclusiveReport. They were unable to
figure out how to relate the classes. This of course, is an
extreme, infrequent, case. Figure 10 shows a fragment of a
correct model, in which class ExclusiveReport has been
properly extended while processing the communicative events.
It corresponds to the reference model created with Integranova.

Figure 10. Fragment of a model in which the rules have been correctly applied
VII. CONCLUSIONS AND FUTURE WORK

Model driven development methods and tools have already achieved a level of maturity that enables industrial adoption. Thus, current research efforts aim at increasing the modelling abstraction from a platform-independent to a computation-independent level. For instance, a technique for deriving OO-Method object-oriented conceptual models from Communication Analysis business process and requirements models has recently been proposed [4-6]. However, if model-driven requirements engineering proposals are to be transferred to industry, the research community faces the challenge of assessing the impact of applying model-to-model transformations instead of applying traditional text-based techniques. For instance, OO-Method practitioners currently apply the latter and are reluctant to adopt novel approaches.

This paper presents a controlled experiment designed to compare a model-driven, communication-based derivation technique with a text-based derivation technique that resembles how OO-Method industrial practice. The results show that, for the teams of analysts with high-competence in the OO-Method, there is a significant difference in the conceptual model completeness that can be attributed to the applied derivation technique (measured as the percentage of requirements that have been represented in the conceptual model). The teams applying the communication-based derivation technique produced models that are 9.22% more complete than those produced by the teams applying the text-based technique. No significant results have been found among average-competence teams, or with regards to model validity (measured in negative terms as the amount of invalid elements included in the conceptual model).

We have already run a replication of the experiment to verify the results; however, due to the time-consuming procedure for the assessment of the conceptual model quality, the results are not available yet. Also, building upon the Eclipse-based modelling tool described in [17], we are completing the implementation of automatic transformation rules using ATL Transformation Language (see [18] for further detail). It is also part of our future work to compare outcomes of the manual and the automatic derivation, in order to shed more light into the mechanisms of model-driven requirements engineering.

VIII. REFERENCES


[10] M. Fortuna, C. Werner, and M. Borges, "Info Cases: integrating use cases and domain models." In: 16th International Requirements Engineering Conference (RE'08), Barcelona, Spain, 2008, pp. 81-84


You will find further information (including many of our papers available for download) here:
http://www.citeulike.org/user/sergioespana
https://www.researchgate.net/profile/Sergio_Espana/
http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/e/Espa=ntilde=a-Sergio.html
We are always open to collaborations; if you share our research interests, please write us an email.