Fraunhofer Institut Experimentelles Software Engineering

Influence of Requirements Specification Notation on Change Impact Analysis

An Empirical Investigation

Authors:
Tom Koenig
Thomas Olsson
Klaus Schmid
Sebastian Adam

IESE-Report No. 106.07/E
Version 1.0
May 2007

A publication by Fraunhofer IESE

Fraunhofer IESE is an institute of the Fraunhofer Gesellschaft.
The institute transfers innovative software development techniques, methods and tools into industrial practice, assists companies in building software competencies customized to their needs, and helps them to establish a competitive market position.

Fraunhofer IESE is directed by
Prof. Dr. Dieter Rombach (Executive Director)
Prof. Dr. Peter Liggesmeyer (Director)
Fraunhofer-Platz 1
67663 Kaiserslautern

Abstract

This report describes the results of an empirical investigation in which the effectiveness and efficiency of change impact analysis using the notations of EPC and Use Cases were compared.

Keywords: empirical evaluation, effectiveness, efficiency, business process modeling, use case, ReqMan

.

Copyright © Fraunhofer IESE 2007

Table of Contents

Copyright © Fraunhofer IESE 2007

Copyright © Fraunhofer IESE 2007

1 Introduction

Several sources [NiSK00] have identified “Handling requirements changes” as one of the key issues with regard to requirements engineering research areas. A survey recently performed by Fraunhofer IESE underpins this statement, as the results showed that the change issue is one of the two most important problem fields encountered in the field of requirements engineering. On a more fine-grained level, companies indicated that especially the instability of their requirements document and the cost estimation regarding a change were the most crucial problems experienced in this area. Several reasons can be identified regarding these results. Our experience shows that one of the key aspects regarding the instability of requirements specifications relates to the neglecting of Non-Functional Requirements (NFR) that may have a significant influence on both Functional Requirements (FR) and Architectural Requirements (AR). The latter reason especially relates to the task of impact analysis regarding a given change, i.e., how effectively and efficiently can the overall impact of change be analyzed regarding a given requirements specification. Several factors influence this question, such as the traceability of the requirements artifacts and the technique or notation used to specify the artifacts.

Across the requirements process, especially while refining requirements from abstract, coarse-grained requirements to very detailed and fine-grained requirements, different specification techniques exist to address the respective specification needs regarding each abstraction level (e.g., the business process may be specified using Activity Diagrams, while user interaction may be specified using Use Cases). The possibility for and availability of different specification techniques that may be used to document requirements is further amplified, as laying the focus on one level of abstraction (e.g., business processes) different notation techniques are available that may be used to specify the requirements (e.g., Activity Diagram, Event driven Process chains). To some extent, this choice can only be guided by personal preferences, as no other objective criteria can be identified. Focusing on the change impact analysis point of view though, the decision regarding the notational choice may have a significant influence. A notation providing a simple and fast overview of the system might support analysts in performing a change impact analysis activity, while a notation that is very detailed and hard to read may negatively influence the impact analysis. Regarding this factor, we wanted to analyze the impact of two specific notations used to specify requirements on a change impact analysis activity based on these documents.

At Fraunhofer IESE, we developed a requirements engineering framework, TORE (Task and Object-oriented Requirements Engineering) [PaKo04], to inte-

Copyright © Fraunhofer IESE 2007

grate RE and OO methods that support the elicitation and specification of system requirements across different levels of abstraction. We addressed two of the abstractions levels and evaluated the impact of using two different specification techniques on a subsequent change impact analysis activity. This document presents the context and the results of this evaluation.

The document is structured as follows. In Chapter 2, we describe the empirical study in more detail, especially introducing the TORE framework. In Chapter 3, we describe the context of the empirical study as well as the experimental design. Chapter 4 presents and discusses the empirical results and, finally, Chapter 5 concludes with a summary and future work identified.

Copyright © Fraunhofer IESE 2007

2 Context of the empirical study

This chapter describes the context of the study, especially explaining the TORE framework.

2.1 The TORE framework

During a requirements engineering process and its related activities, several decisions have to be taken regarding the system under discussion, e.g., decisions about supported processes, functional behavior, etc. Each of the different requirements engineering activities conducted throughout a requirements engineering process thus leads to decisions that, on the one hand constrain the solution space for the subsequent development activities, and, on the other hand, have to be documented using a specific, suitable notation.

TORE (Task and Object-Oriented Requirements), illustrated in Figure 1, is a framework that integrates requirements engineering and object orientation for User Interface Systems, presenting in total sixteen decision points that have to be addressed throughout the requirements engineering activities. The sixteen decision points of the framework are grouped by levels, namely task level, domain level, interaction level, system level, and GUI level.

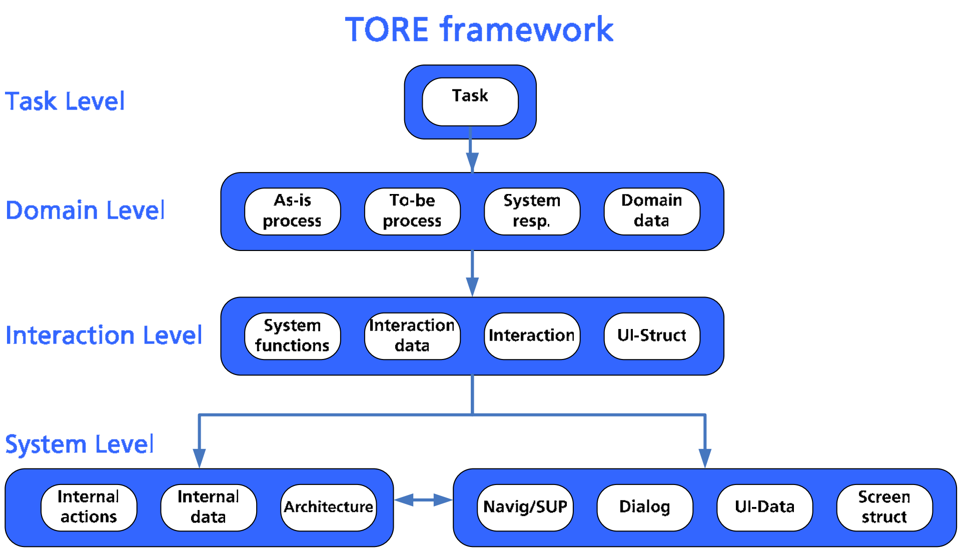


Figure 1: TORE framework

The following is a brief description of each of the decision points:

Copyright © Fraunhofer IESE 2007

- Tasks (T):

  The decision point determines the user roles and the tasks of these roles to be supported by the system.

- As-is/To-be activities (D1, D2):

  Each of the user tasks consists of different activities. In a first step, the as-is activities decision point specifies how these different steps are currently (i.e., without the new system) being performed as part of the user's work. The to-be decision point documents the decision.

- System responsibilities (D3):

  Usually a system is not able to/ should not support all to-be activities, but only a subset. This subset is called system responsibilities.

- Domain data (D4):

  System responsibilities, i.e., process activities, manipulate data. Decisions have to be made on which domain data is relevant for the system responsibilities.

- System functions / User system interaction (I1, I2):

  The different system responsibilities imply a certain interaction between user and system in order to achieve the activity goal. A system responsibility is realized by one or more system functions. The interaction clarifies how the user can use the system function to achieve the system responsibilities.

- Interaction data (I3):

  The different system functions use input data provided by the user and produce output data provided by the system.

- UI Structure (I4):

  Decisions about the grouping of data and system functions in different workspaces have to be made. System functions and data grouped into one workspace will be closely together in the GUI. This means that users need less navigation effort in the interface to invoke system functions and view data within the workspace. Through the UI structure, the rough architecture of the user interface is defined. This structure has a big influence on the usability of the system.

- Application architecture (C1):

  The code realizing the system functions is modularized into different components. In the decision about the component architecture, existing compo-

Copyright © Fraunhofer IESE 2007

nents and physical constraints as well as quality constraints such as performance have to be taken into account. During requirements, only a preliminary decision concerning the architecture is made. This is refined during design and implementation.

- Internal system actions (C2):

  Decisions have to be made regarding the internal system actions that realize the system functions. The system actions define the effects of the system function on the data. These decisions also define an order between the system actions as far as this is necessary to understand the behavior of the system function. In OO, the system actions are grouped within classes. This is only a preliminary decision, which is refined during design and implementation.

- Internal system data (C3):

  The internal system data refines the interaction data to the granularity of the system actions. The decisions about the internal system data reflect all system actions. In OO, system data is grouped within classes. Again, this is only a preliminary decision, which is refined during design and implementation.

- Navigation and support functions (G1):

  It has to be decided how the user can navigate between different screens during the execution of system functions. This determines the navigation functions. In addition, support functions that facilitate the system functions have to be defined. These functions realize parts of system functions that are visible to the user, for example, by processing chunks of data given by system functions in a way that can be represented in the user interface. Another example are support functions that make the system more tolerant towards user mistakes.

- Dialog interaction (G2):

  For each interaction, the detailed control of the user has to be decided. This determines the dialog. It consists of a sequence of support and navigation function executions. These decisions also have a strong influence on the usability of the system.

- Detailed UI data (G3):

  For each navigation and support function, the input data provided by the user as well as the output data provided by the system have to be defined. These decisions determine the UI data visible on each screen.

Copyright © Fraunhofer IESE 2007

- Screen structure (G4):

  The separation of workspaces as defined in (I4) into different screens that support the detailed dialog interaction as described in (G2) has to be decided. The screen structure groups navigation and support functions as well as UI-data. The decisions to separate the workspaces into different screens are influenced by the platform of the system.

For each of the decision points, a set of different notation techniques exist, which may be used to document the respective decision. TORE does so far not present any specific guidelines on which technique should be used for a specific project setting. TORE only presents a set of alternatives for each of the decision points. Practical experience has shown that for certain decision points, even though guidelines are present, it is very hard to judge which technique from a set of alternatives is most suited to document the decision. This choice might have an influence on different factors not taken into consideration at that moment. For example, the change impact analysis activity as sketched in the following section may be influenced by the prior choice of the requirements specification notation.

2.2 Does the requirements specification technique influence the change impact analysis activity?

Handling unstable requirements is a serious issue in industry. Requirements are highly unstable and the emerging changes are especially hard to estimate with regard to cost and time. A crucial activity regarding these problems thus is the change impact analysis activity. In order to estimate the costs of a change, it is essential to accurately identify which parts of a requirements specification are affected by a specific change. It could be observed that the characteristics of the requirements specification present factors especially influencing the outcome of this activity, e.g., the degree of traceability or the notation used to document the requirements. Regarding these characteristics, the following question was raised: Does the notation used to specify the requirements (i.e., in TORE, the decision to document a given decision point) have an influence on the change impact analysis activity?

We identified two notations commonly used to document some of the decision points of the task/domain levels. The decision point tasks and as-is/to-be process present the abstract and refined business processes of a company and are commonly documented using either Event-driven Process Chains (EPC) or textual Use Cases (TUC). Each of the two notations has its advantages and disadvantages recommending its respective usage. It was nevertheless not possible to definitely assess which of the two notations was best suited. Therefore, we conducted a quasi-experiment to address this question. The following presents a brief introduction to each of the two notations, additionally indicating the ob-

Copyright © Fraunhofer IESE 2007

jective expert assessment of each of the two techniques leading to our hypotheses documented in chapter 3.

2.3 Introducing EPC and Use Cases

2.3.1 Use Cases

Use Cases are a technique that can be used to capture the requirements of a system to be developed. In more detail, a use cases explicitly captures the user-system interaction for a specific situation, a so-called scenario. In its common form, a Use Case is a textual description that usually adheres to a specific template. For more background information on Use Cases, see [Cock01]. We at Fraunhofer IESE use the following textual description to specify a Use Case:

- Use Case
  Name of the Use Case (Identifier)

- Actors
  List of actors that participate in this Use Case. Actors are not necessarily humans; actors can be other systems as well.

- Intent
  Goal that should be reached through this Use Case.

- Preconditions
  Condition that has to be fulfilled prior to executing this Use Case.

- Flow of events
  This area documents the flow of the Use Case indicating human interaction and system reaction. The different events (interactions and reactions) are numbered.

- Exceptions
  This area indicates exceptions to the usual flow of events documented above. An Exception refers to the specific number (by specific event) in the flow of events and indicates under which circumstances the exception occurs.

- Rules
  This area documents rules that have to be guaranteed for this Use Case.

Copyright © Fraunhofer IESE 2007

- Quality requirements
  Non-functional requirements (NFRs) that have to be ensured. NFRs can apply to the whole Use Case or to specific parts of the flow of events.
- Input data
  Data serving as input to the Use Case.
- Output data
  Data produced as output of the Use Case.
- Functions
  System functions that have to be provided in order to implement the given flow of events.
- Postconditions
  Condition that has to be met after Use Case execution.

We use the following template structure to document the different Use Cases:

UseCase	
Actors	
Intent	
Preconditions	
Flow of events	
Exceptions	
Rules	
Quality requirements	
Input data	
Output data	
Functions	
Postconditions	

Table 1: Use Case Template

2.3.2 EPC

Event-driven process chains (EPC) are a semi-formal graphical notation used primarily to represent business processes. EPCs were developed in the 1990s by the Institut für Wirtschaftsinformatik (IwI) at the University of Saarland and integrated into their so called ARIS (Architecture of Integrated Information Systems) framework.

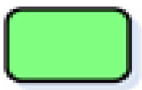
Copyright © Fraunhofer IESE 2007

Due to the fact that, on the one hand, EPCs are one of the central components of the SAP R/3 system and the underlying framework and, on the other hand, they are integrated into the ARIS framework and the homonymous widespread ARIS toolset developed by IDS-Scheer, EPCs have become a very widespread, commonly used method to model business processes in industrial and academic practice.

The EPC notation contains three basic notational elements used to model business processes as well as further additional notational elements. The notational elements are briefly explained in the following section:

- Event

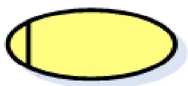
- Function



- Logical Operator

- Data object

- Organizational unit



These notational elements are used to model business processes, potentially across different levels of detail. The different elements are related to each other, e.g., an organizational unit may be responsible for a certain task, an information object may serve as input/output or may be influenced by an activity, events may trigger functions or may be seen as states reached after functions have been executed. This description reflects that functions (and process elements) can be seen as the central object of the EPC flow, as they are related to all other elements. The big picture operationalizing the relationships between the notational elements via connectors can bee seen in the following figure:

Copyright © Fraunhofer IESE 2007

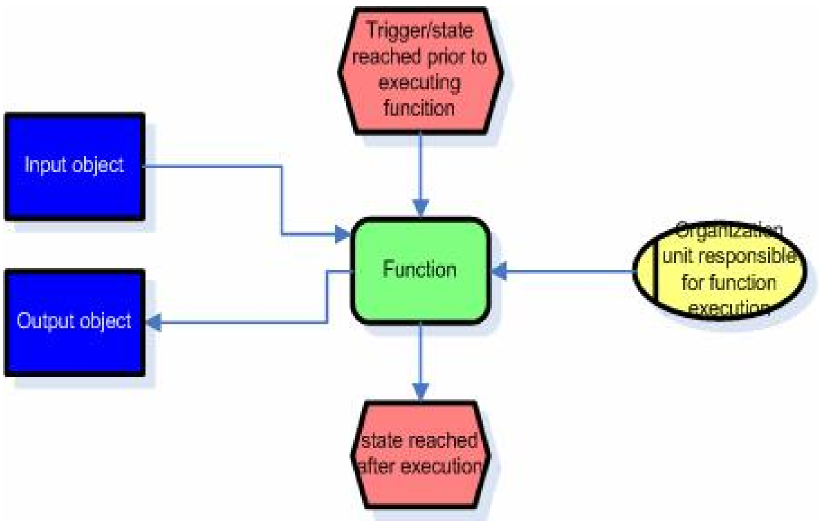


Figure 2: Elements of EPC

2.3.3 Expert Assessment

This section gives a brief expert assessment regarding the aforementioned techniques, TUC and EPC, with regard to their suitability as specification notations for the decision points “tasks” and “as-is/to-be process”.

Use Cases are a textual notation. The template is quite simple and easy to read. We applied Use Cases in several projects, especially also with stakeholders who had never seen or read a Use Cases before, and the experience showed that TUC were very easy to learn and to apply. One of the problems we encountered was that for a larger amount of Use Cases, people lost the overall system overview and were not able to easily relate (trace) the different Use Cases to each other. It was hard to retrace the big picture based on the textual links between Use Cases.

EPCs are a graphical notation. EPCs in their basic form as described above do not contain many elements. They are easy to learn and to understand. Based on their graphical representation, relations (traces) between different elements are modeled and are therefore immediately visible and understandable.

With regard to TORE, the level following the domain level is the interaction level and especially the decision point interaction. This decision point is commonly documented using Use Cases. Therefore, using TUC on the domain level would prevent a notational break, facilitating the transition. Using EPCs would imply a notational break, which may be difficult to trace.

Copyright © Fraunhofer IESE 2007

Based on this assessment, we were not able to conclude which of the two notations is best suited to document the decision points; therefore, further empirical research is necessary to evaluate the influence of each of the two notations, especially on the change impact analysis activity.

Copyright © Fraunhofer IESE 2007

3 The Empirical Investigation

3.1 Introduction to the empirical study - Questions & broad hypotheses

With regard to the two techniques EPC and TUC, we would like to address the goal described in Table 2 and answer the questions described in Table 3.

Object of study	Domain Level Notation techniques broad Use Cases versus Event-driven process chains to describe as-is/to-be process decision points in a requirements document
Purpose	Evaluate impact of the two notations w.r.t conducting change impact analysis task (for several changes) on different levels (domain and interaction level)
Quality focus	Effectiveness, Efficiency
Perspective	Researcher
Context	Objects: Requirements document (Taxi System) and Size (Number of requirements), Change requirements (Type) Subjects: Students (Background, Background Requirements Engineering, Background Use Cases, Background Business Processes)

Table 2: Goal of empirical study

How fast can a group using EPCs as requirements notation perform their change impact analysis task compared to a group using broad Use Cases? How many of the impacts can a group using EPCs find compared to a group using broad Use Cases? Does the type of change (high level change versus specific change) have an impact on the above results?

Table 3: Questions of the empirical study

Based on our prior assessment, we inferred broad hypotheses with regard to the questions in Table 3.

Basically, we think that the change impact analysis task on the domain and interaction level is divided into two subtasks. First of all, the change impact has to

Copyright © Fraunhofer IESE 2007

be analyzed and tracked with regard to the domain level (defined as subtask "determine high-level impacts"). Once this task has been executed, the tracing of the different impacts to the interaction level and the tracking of all impacts on the interaction level based on this analysis has to be done (defined as sub-task "determine low-level impacts").

We think that the tracing and thorough tracking of change impacts is more efficient and effective using EPC with regard to the domain level. EPC are a graphical notation facilitating the overview of a system and thus the impact analysis as opposed to textual Use Cases. The transition towards the interaction level and thus the tracking of changes on the interaction level will be slightly more efficient and effective using TUC, as with TUC, no notational break occurs and the tracing is slightly better.

Thus, the overall hypothesis based on this assessment is that EPC will be more efficient and effective than TUC throughout the change impact analysis activity on the domain and interaction levesl.

Furthermore, we differentiate between two types of changes:

- High-level change: Change having an impact on a process activity or a complete Use Case.
- Specific change: Change having an impact on parts of Use Cases or specific parts of an activity.

Regarding this distinction, we estimate that for high-level changes, a good overview is needed to track the impacts, thus EPC will perform better than TUC with regard to efficiency and effectiveness.

With regard to specific changes, we think that none of the two notations will show a better performance.

3.2 Experimental design

We conducted the experiment in three sessions, an introductory session (session 0) and two change impact analysis sessions (sessions 1 & 2).

Copyright © Fraunhofer IESE 2007

	Group 1	Group 2
Session 0	Introductory session for all participants	
Session 1	Use Cases	EPC
Session 2	EPC	Use Cases

Figure 3: Experiment setting

In the first session (session 0), a brief introduction to TUC and EPC was given, thereby assuring that each participant had a basic and sufficient knowledge of both techniques. Based on this introduction, all participants were able to conduct the change impact analysis activity, as no further knowledge with regard to the techniques was considered necessary.

We divided the participants into two groups. The participants were randomly assigned to the groups in order to reduce a potential human performance effect.

The following describes the material handed out to the participants and the procedure of the experiment.

3.2.1 Material

During the change impact analysis sessions, both groups were handed out the following documents:

- Guidelines

  This document describes which activities the participants have to execute and based on which documents decisions have to be made.

- Requirements documents

  The requirements document specified the requirements for a "Taxi system". It is divided into three sections:

  - Introduction

    Textual description broadly describing the taxi system.

  - Domain Level

Copyright © Fraunhofer IESE 2007

The domain level specifies the decision points as-is/to-be process. The notation used is either TUC or EPC.

- o Interaction Level

  The interaction level specifies the decision point interaction using Use Cases.

In addition to these documents, the participants were handed out a set of change requests:

- Change requests

  The change requests are described textually. In total, 10 change requests had to be analyzed. Change requests 1-5 were analyzed in session 1 and change requests 6-10 were analyzed in session 2.

- Questionnaire

  After the experiment sessions, the participants filled out a *questionnaire* regarding their experience with change impact analysis, requirements engineering, and the two notations used, EPC and TUC. Additionally, the participants were asked to give their subjective assessment with regard to our hypotheses.

3.2.2 Experimental setting

The experimental setting, which also describes the differences between the sessions and the groups, is depicted in Figure 3.

Session 1:

In the first session, both groups had to analyze change impacts #1-#5. Group 1 analyzed the change impacts based on the requirements document type defined as "TUC" or "Requirements document setting A" depicted in Figure 4, using TUC as specification notation on the domain level. Group 2 analyzed the change impacts based on the requirements document type defined as "EPC" or "Requirements document setting B" depicted in Figure 5, using EPC as specification notation on the domain level.

Copyright © Fraunhofer IESE 2007

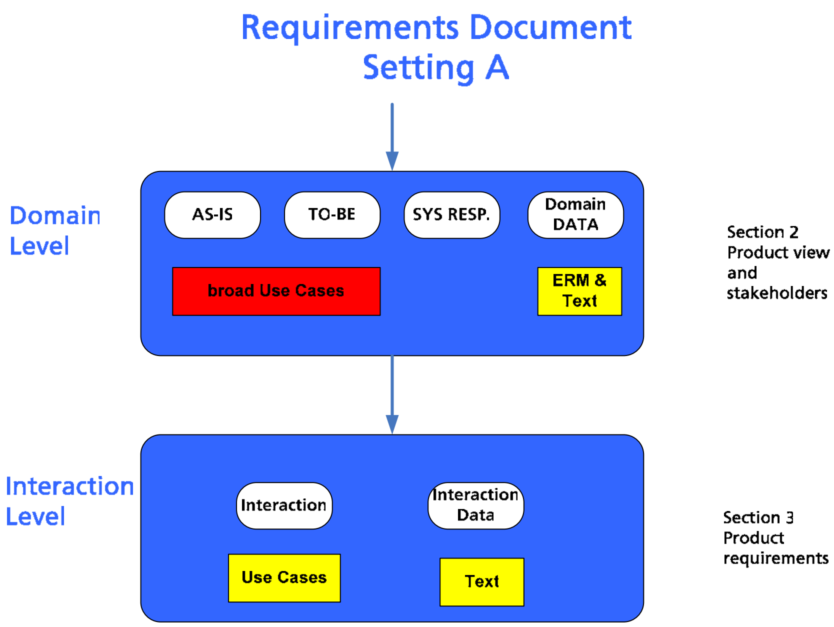


Figure 4: Requirements document setting A "TUC"

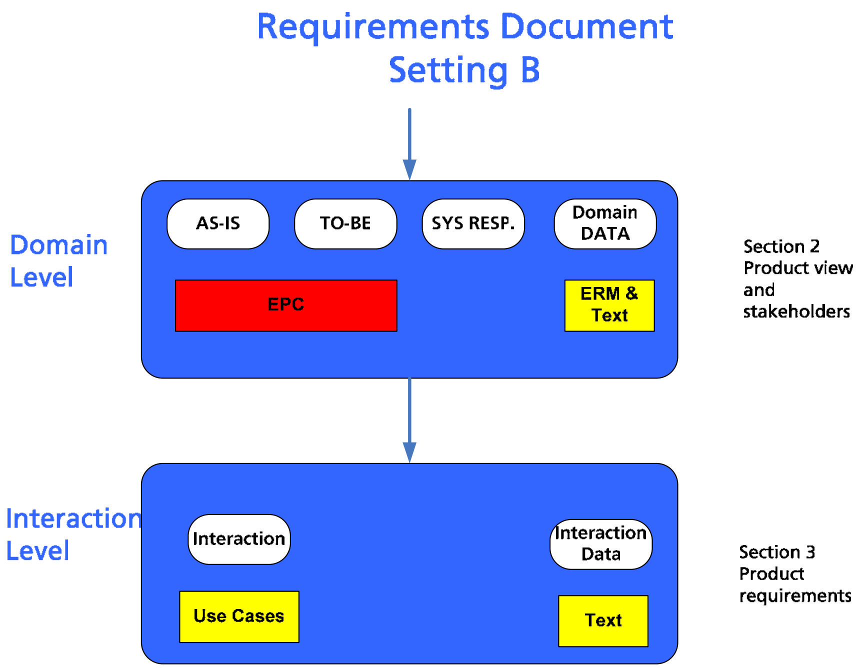


Figure 5: Requirements document setting B "EPC"

Copyright © Fraunhofer IESE 2007

Session 2:

In the second session, both groups had to analyze change impacts #6-#10. Group 2 analyzed the change impacts based on the requirements document type defined as "TUC" or "Requirements document setting A" depicted in Figure 4, using TUC as specification notation on the domain level. Group 1 analyzed the change impacts based on the requirements document type defined as "EPC" or "Requirements document setting B" depicted in Figure 5, using EPC as specification notation on the domain level.

Both groups thus worked with EPC as well as with TUC (session 1 and session 2, respectively) to analyze the requirements document.

3.3 Refinement of Hypotheses

Based on the experimental setting, we elaborated the broad hypotheses. The requirements document (see 3.2.1) specifies the different decision points, (i.e. as-is-, to-be process as well as interaction) with regard to the TORE abstraction levels domain and interaction which differ in terms of their level of abstraction. The domain level specifies the requirements on a higher level (process), while the interaction level refines these high-level requirements and specifies requirements on a lower level. With regard to these abstraction levels, we thus divided the requirements document into high-level (Chapter 2) and low-älevel (Chapter 3) as depicted in Figure 6.

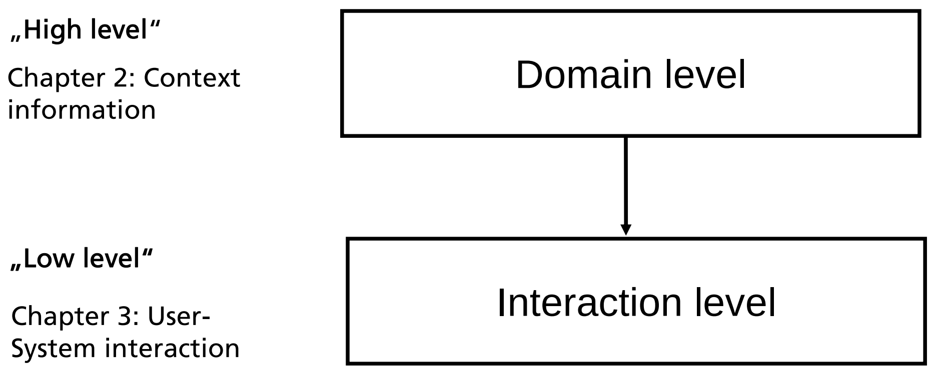


Figure 6: Requirements document classification

This requirements hierarchy (domain- and interaction level) has an impact on the change impact analysis task, as also stated in section 3.1. The change impact analysis task can be split into two subtasks. On a higher level of abstraction, participants have to track the impacts with regard to the domain level decision points. Based on the results of this first subtask, participants are able to trace the located impacts to the lower level of the specification (i.e., interaction level) and track the change impacts on that level. The change impact analysis

Copyright © Fraunhofer IESE 2007

task is thus divided (see Figure 7) into determining impacts (high level) and determining impacts (low level), including the trace step.

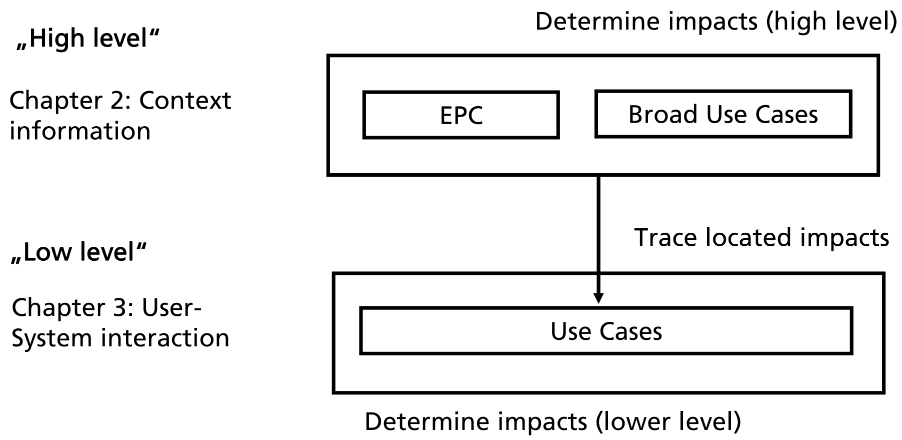


Figure 7: Change impacts analysis task

Concerning this distinction, we refined (based on our assessment given in section 3.1) the broad hypotheses (see section 3.2):

- Efficiency (High Level): H1.1: Individuals using EPC perform better than individuals using TUC with regard to the time needed to perform the task.
- Effectiveness (High Level): H2.1: Individuals using EPC perform better than individuals using TUC with regard to the impact detection rate.
- Efficiency (Low Level): H1.2: Individuals using Use Cases perform better than individuals using EPC with regard to the time needed to perform task.
- Effectiveness (Low Level): H2.2: Individuals using Use Cases perform better than individuals using EPC with regard to the impact detection rate.
- Efficiency (OVERALL): H1: Individuals using EPC perform better than individuals using TUC with regard to the time needed to perform task.
- Effectiveness (OVERALL): H2: Individuals using EPC perform better than individuals using TUC with regard to the impact detection rate.
- Efficiency (High level change): H3.1: Individuals using EPC perform better than individuals using broad Use Cases with regard to the time needed to the perform task.
- Efficiency (Specific change): H3.2: There is no difference between individuals using EPC and individuals using broad Use Cases with regard to the time needed to the perform task.

Copyright © Fraunhofer IESE 2007

- Effectiveness (High level change): H4.1: Individuals using EPC perform better than individuals using broad Use Cases with regard to the impact detection rate.
- Effectiveness (Specific change): H4.2: There is no difference between individuals using EPC and individuals using broad Use Cases with regard to the impact detection rate.

As the experiment was conducted in two sessions, we formulated another hypothesis regarding this fact:

- Efficiency & Effectiveness (Overall - Session 1): H5.1 & H5.2: Individuals using EPC perform better than individuals using broad Use Cases with regard to the time needed to perform the task and with regard to the impact detection rate.
- Efficiency & Effectiveness (Overall - Session 2): H5.3 & H5.4: Individuals using EPC perform better than individuals using broad Use Cases with regard to the time needed to perform the task and with regard to the impact detection rate.

3.4 Procedure

In a preliminary step, a short introductory tutorial (session 0) was held. The intent of the tutorial was to give a short overview of the experiment procedure on the one hand, but, on the other hand, to also give a brief introduction to the topic of change impact analysis as well as to the two notations Use Cases and EPCs.

The change impact analysis task was performed in two sessions as described in section 3.2. In each of the sessions, the participants were handed out the guidelines as well as the requirements document (only Part1: Overview and Part2: Domain level description) as well as the changes to be analyzed. Based on this information, the participants proceeded to perform the first subtask "determine high-level impacts". After a fixed time period, the change impact analysis results were collected and the participants were handed out Part3: Interaction Level of the requirements document. The participants then proceeded to perform the second subtask "determine low-level impacts" based on Part3. After a fixed time period, the results were collected and the session ended. Hereafter, the participants had to fill out a questionnaire assessing their personal experience and their personal evaluation of our hypotheses.

Copyright © Fraunhofer IESE 2007

3.5 Experiment variables

To investigate the hypotheses listed in section 3.3, different types of variables have to be considered. First of all one has to differentiate between independent and dependent variables. With regard to the independent variables, the distinction between controlled and uncontrolled variables is important.

The following lists all controlled and uncontrolled independent variables as well as all dependent variables that are essential for the empirical study. For each variable, a brief description as well as the respective scale is indicated.

Controlled independent variables	
Notation technique (domain level)	Notation used to specify the requirements on the domain level. Nominal scale: Two levels: EPC, TUC
Notation technique (interaction level)	Notation used to specify the requirements on the domain level. Fixed: TUC
System	System for which the requirements are specified in the requirements documents Fixed: Taxi System
Traceability	How is traceability assured throughout the requirements document? Fixed: By Names
Type of change	Type of a specific change request Nominal scale: Two levels: high level, specific

Table 4: Controlled independent variables

Uncontrolled independent variables	
Requirements engineering experience	Indicates the participants' experience in the field of requirements engineering Ordinal scale
Requirements document experience	Indicates the participants' experience in using requirements documents Ordinal scale
Change integration experience	Indicates the participants' experience with integrating changes into a requirements document. Ordinal scale
EPC experience	Indicates the participants' experience regarding the EPC notation

Copyright © Fraunhofer IESE 2007

	Ordinal scale
TUC experience	Indicates the participants' experience regarding the TUC notation Ordinal scale
Language	Indicates the participants' experience using the English language Ordinal scale
Motivation	Indicates the participants' motivation Ordinal scale

Table 5: Uncontrolled independent variables

Regarding the dependent variables, we have identified two high-level variables, efficiency and effectiveness, of the change impact analysis activity we want to measure.

We use the following variables as operationalizations of effectiveness and efficiency, respectively.

Effectiveness will be expressed through the mean impact detection rate (IDR), while efficiency will be expressed through the mean time per correct change impact (TPC):

Effectiveness:

$$IDR = \frac{\#correct\ impacts}{\#overall\ impacts}$$

Efficiency:

$$TPC = \frac{time\ needed}{\#correct\ impacts}$$

IDR and TPC will be determined with respect to the notation technique, as we want to compare the effectiveness and efficiency of the two notations EPC and TUC.

Additionally the two variables will be refined to address the different hypotheses sketched. We will evaluate the IDR and TPC for the two subtasks "determine high level impacts" and "determine low level impacts" to address hypotheses 1.x and 2.x. To address hypotheses 3.x and 4.x the variables will also be evaluated for the types of changes (high level and specific change). At last a differentiation of the two sessions will be made to address hypotheses 5.x

Copyright © Fraunhofer IESE 2007

3.6 Threats to validity

This section discusses the threats to validity that are important in the setting of the empirical study. Threats to validity describe factors that influence the dependent variables but might not be controlled immediately by the experimenter [Camp63]. As the influence of the factors on the dependent variables should be minimized, each of the threats to validity should be addressed upfront by the experiment design. This section lists the different threats to validity we identified and describes the measures we took to identify them and to limit their influence upfront, if possible.

Maturation

Maturation deals with different reactions/behavior of the participants over time affecting the outcome. This especially regards motivation (negative) and learning effects (positive). The questionnaire showed that motivation could be preserved throughout the experiment. In order to compensate learning effects, other changes were used for the second session. As the same requirements document was used in both sessions, the time to familiarize with documents (which only applied for the first session) was not part of the evaluation. We assume that a certain learning effect still remains for the second session.

Instrumentation

Instrumentation deals with the fact that a difference in the experiment artifacts has an influence on the outcome. Both groups analyzed the requirements for the same system (i.e., Taxi system). The requirements document differed only with regard to the notation used to specify the requirements for the domain level (EPC versus TUC), i.e., Chapter 2. In order to specify the same requirements, the requirements for both documents were matched against each other and missing facts were documented in the respective artifact. Additionally, a thorough review was performed by an external person in order to assure that both documents capture and express the same requirements.

Selection

Selection addresses the influence of variation on human performance. The risk is that one group is significantly stronger than the other with respect to the tasks to be performed. This threat was (as is usually done) controlled by assigning the participants randomly to the two groups. In addition, we tried to evaluate the participants' experience regarding the different critical aspects. The questionnaire evaluation does not indicate any significant strength differences between the two groups.

Process conformance

Process conformance addresses the threat that the participants might not follow the intended procedure. In order to control this factor, we decided to hand

Copyright © Fraunhofer IESE 2007

out the different document parts one after the other and not to hand out the complete documentation upfront. Thus, the participants were not able to track changes in the complete document at once.

Mortality

Mortality addresses the fact that participants may drop out during the course of the experiment. We were not able to prevent drop-out. However, subjects that did not participate in both sessions were not taken into account during evaluation.

Interaction of selection and treatment

This threat addresses the aspect of having participants who are not representative of the population. As the subjects are not software professionals, generalization of the results is limited.

Interaction of setting and treatment

This threat addresses the aspect of the experiment setting and the material not being representative of industrial practice. The material, i.e., the requirements document for the Taxi system, is a requirements document from industrial practice, which was slightly adapted with regard to the size of the document. Nonetheless, it can be seen as representative of industrial practice. The session setting, i.e., the university setting, cannot be considered representative of industrial practice.

Copyright © Fraunhofer IESE 2007

4 Analysis of the results of the empirical study

In this chapter, we sketch the results of our empirical study.

Efficiency and effectiveness with regard to the change impact analysis task and its respective subtasks

This section presents the results of the empirical evaluation with regard to hypotheses 1, 2, as well as 1.x and 2.x.

4.1 Descriptive Statistics

The following section summarizes the data gathered via the concluding questionnaire.

The questionnaire addresses the uncontrolled independent variables listed in Table 1, as well as factors potentially influencing the results, and, finally, subjective estimations of each participant with regard to the performance of the notation.

Regarding their experience with requirements documents we asked participants to subjectively assess their experience and further specify how many requirements documents they have read so far.

Copyright © Fraunhofer IESE 2007

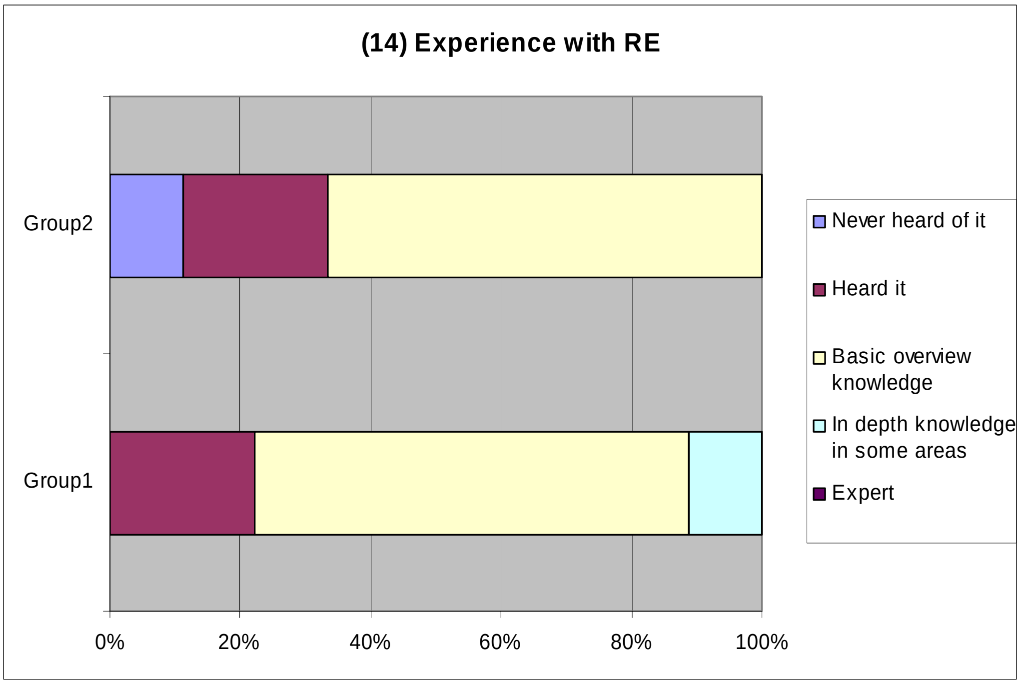


Figure 8: Experience with requirements engineering

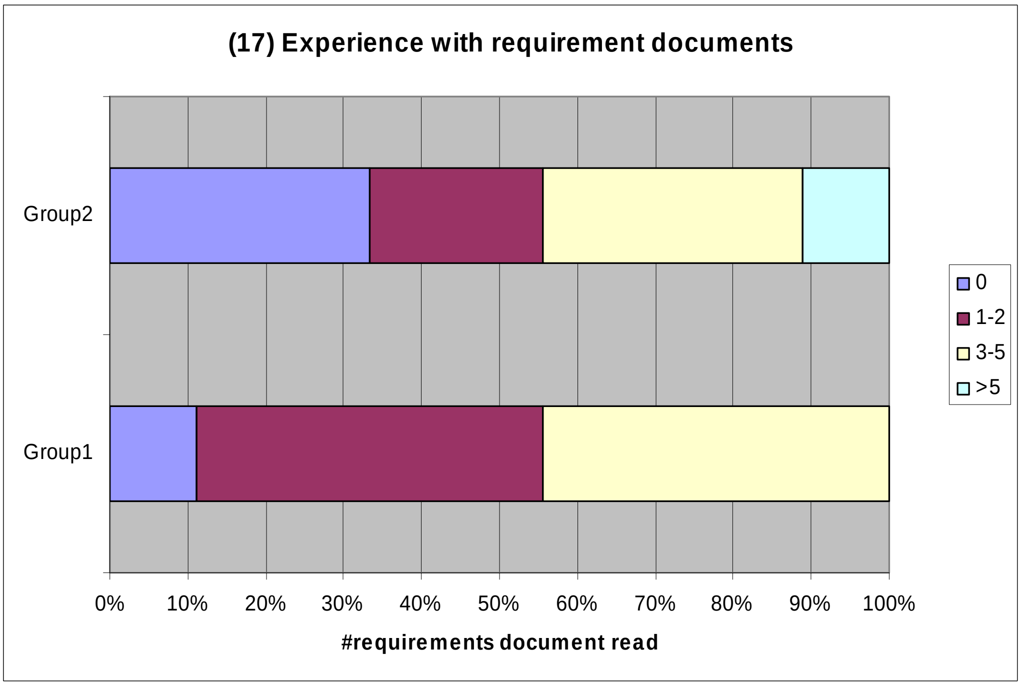


Figure 9: Experience with requirements documents

Copyright © Fraunhofer IESE 2007

Figure 8 and Figure 9 show that both groups do not differ significantly regarding their requirements engineering background.

With regard to the used notations TUC and EPC, we can observe differences between the two groups. Group 1 subjectively seems to be slightly more experienced with both the EPC and the TUC notation (see Figure 10 and Figure 11).

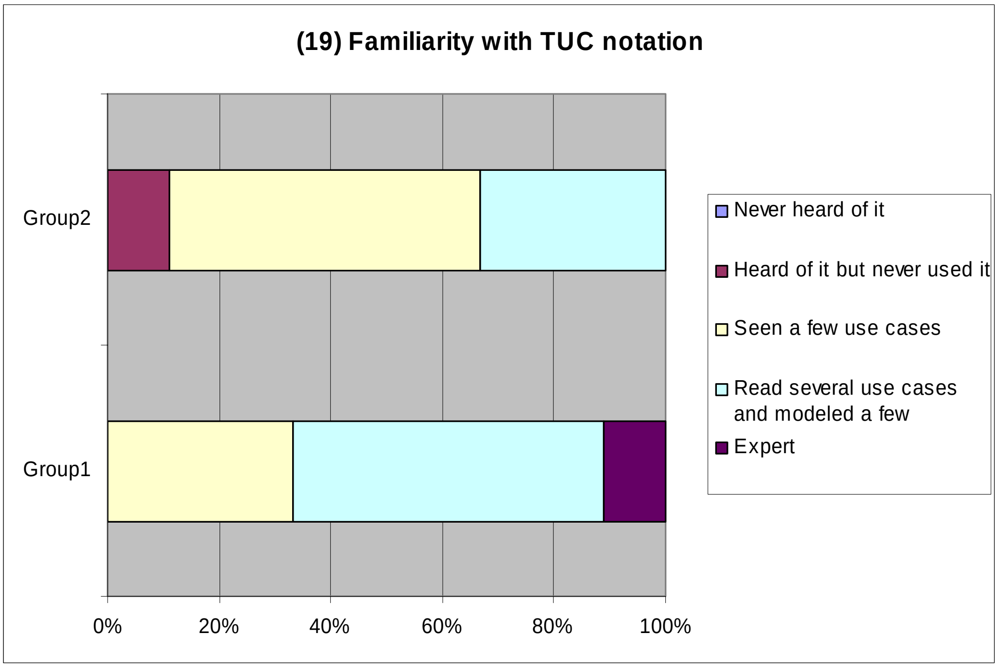


Figure 10: Familiarity with TUC notation

Copyright © Fraunhofer IESE 2007

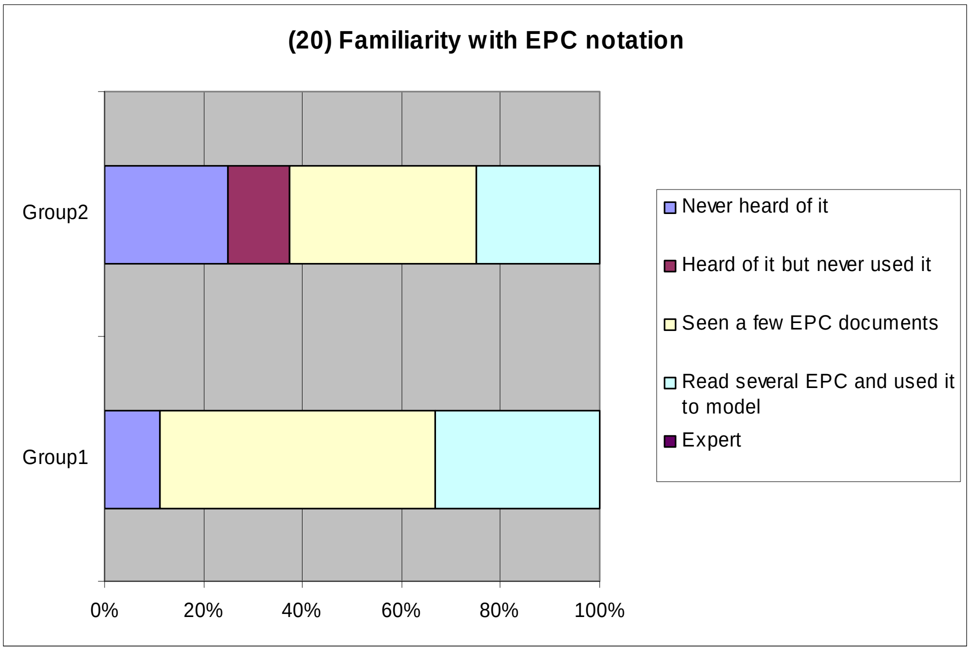


Figure 11: Familiarity with EPC notation

Regarding the task of change impact analysis (see Figure 12), most of the participants had not analyzed any requirements document for change impact. Group 1 shows a little bit more experience than Group 2 in this area.

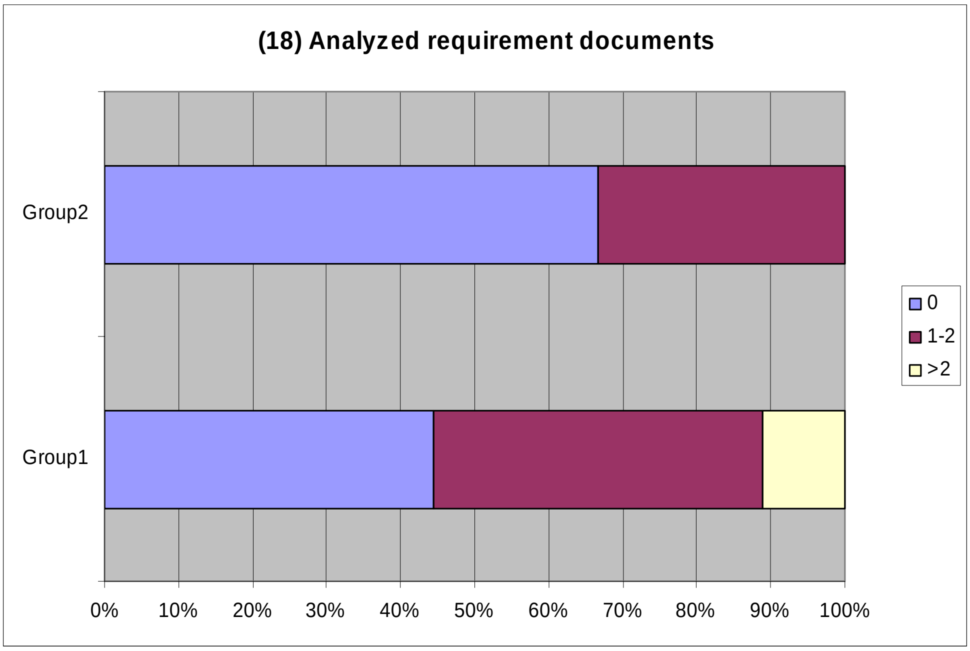


Figure 12: Analyzed requirements documents

Copyright © Fraunhofer IESE 2007

Regarding the experience of understanding documents in English, most of the participants indicated (see Figure 13) not having any no problems at all.

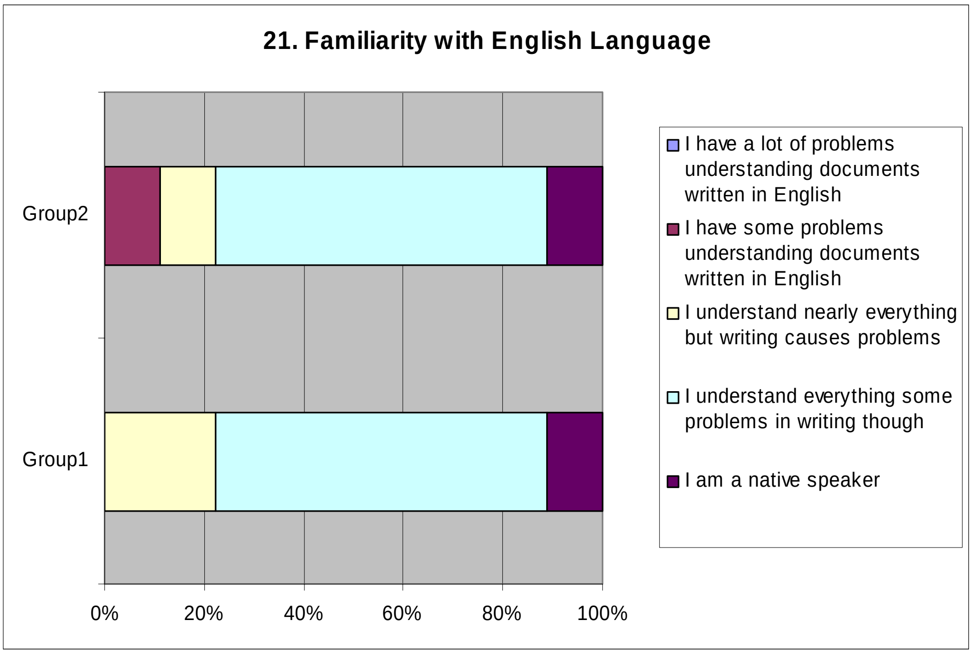


Figure 13: Experience with English language

Concerning the uncontrolled independent variables, the questionnaire did not give any indication of problems.

As we wanted to assess the participants' subjective assessment regarding our hypotheses, we asked participants to indicate what their subjective opinion was regarding the following facts:

- It was easy to find impacts using Use Cases
- It was easy to find impacts using EPC
- I was able to find impacts fast using Use Cases
- I was able to find impacts fast using EPC
- It was easier to find impacts using Use Cases than using EPC
- I was able to find impacts faster using Use Cases than using EPC
- I was able to find more impacts using Use Cases than using EPC
- I was able to find a higher percentage of impacts (number of impacts found versus number of overall impacts contained) using Use Cases than using EPC

The following figure (see Figure 14) shows the results of this assessment.

Copyright © Fraunhofer IESE 2007

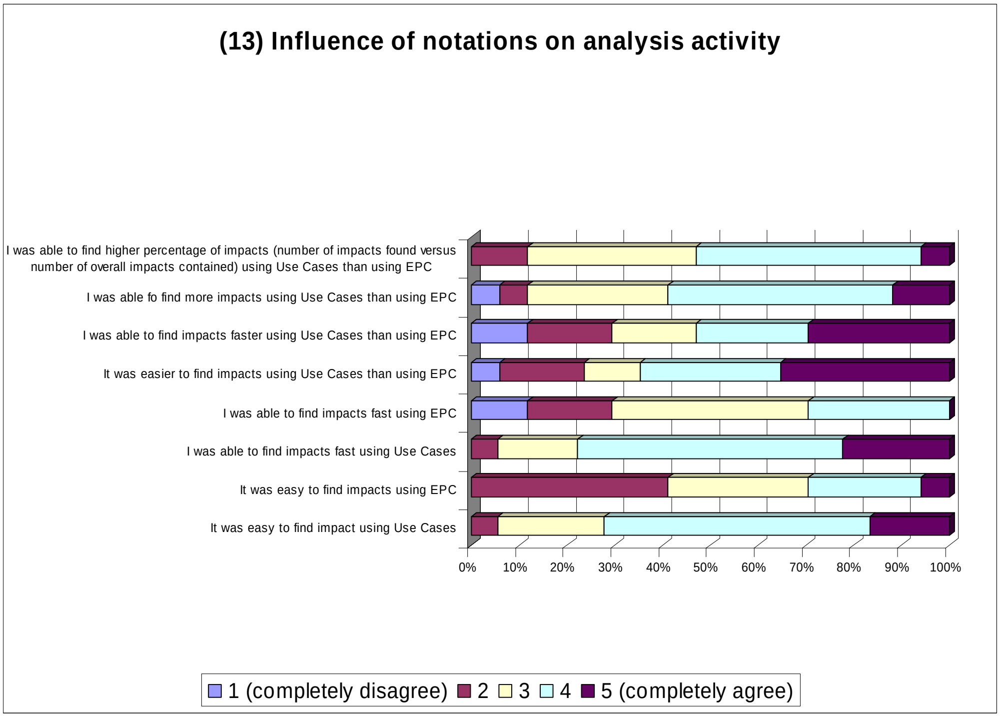


Figure 14: Subjective hypothesis assessment

The results show that the participants tend to favor TUC over EPC, as a higher percentage of participants ranked TUC higher than EPC with regard to the ease of finding impacts, the speed and the effectiveness of finding impacts.

4.2 Overview and interpretation of the empirical findings

In this section, we briefly sketch the results of the empirical finding and discuss the empirical findings, relating the results to the hypotheses. The detailed results for each hypothesis can be found in section 4.3.

Copyright © Fraunhofer IESE 2007

Hypotheses	Efficiency	Effectiveness
Overall: H1 & H2	p=0,046 $H_0$1 can be rejected H1 can be accepted	p=0,111 $H_0$2 cannot be rejected H2 cannot be accepted
High Level: H1.1 & H2.1	p=0,9 $H_0$1.1 cannot be re-jected H1.1 cannot be ac-	p=0,28 $H_0$2.1 cannot be re-jected H2.1 cannot be ac-
Low Level: H1.2 & H2.2	p=0,006 $H'_0$1.2 can be rejected H (a) 1.2 can be ac-cepted	p=0,046 $H'_0$2.2 can be rejected H (a) 2.2 can be ac-cepted

Figure 15: Empirical findings for hypotheses H1, H2

As can be seen in Figure 15, hypothesis H1 can be accepted, while H2 is not acceptable. Thus we can say that as a general statement regarding the overall task of change impact analysis, EPC are significantly better than TUC with regard to the efficiency of the task. The time needed to find impacts using EPC is significantly lower than using TUC. As briefly sketched in section 3.1, we think that the reason for this difference lies in the graphical form of the EPC, allowing a much better and faster overview of the system and tracing to the lower levels, which is needed for the change impact analysis, thus speeding up the task significantly. No statement can be made regarding effectiveness, though; the results do not differ significantly.

When analyzing the impact analysis in more detail (i.e., regarding the subtasks-high and low-level impact analysis as described in section 3.3), we saw that no significant difference could be observed for high-level change impact analysis. Here our hypotheses did not hold, thus on a higher level, the graphical overview may not be as helpful as expected. However, a significant difference was observable with regard to low-level change impact analysis, i.e., tracing of changes to the lower level and the impact analysis on the lower level. In these cases, our initial hypothesis (i.e., TUC are more effective and more efficient than EPC) did not hold, as the descriptive statistic data (see Table 6 and Table 7) indicate an opposite trend, with the values for TPC and IDR being better for EPC than for TUC. Tracing and impact analysis thus benefit from the EPC overview, which facilitates the impact analysis activity.

TPC_{TUC}	TPC_{EPC}
4.39	2.73

Table 6: TPC subtask “determine low-level impacts”

Copyright © Fraunhofer IESE 2007

IDR_{TUC}	IDR_{EPC}
45%	55%

Table 7: IDR for subtask "determine low-level impacts"

Regarding the influence of the change type (i.e., high-level versus low-level changes), the results of the empirical finding are sketched in Figure 16

Hypotheses	Efficiency	Effectiveness
High level change (H3.1 & H4.1)	p=0,49 H_0 3.1 cannot be rejected H3.1 cannot be accepted	p=0,37 $H_0$4.1 cannot be rejected H4.1 cannot be accepted
Specific change (H3.2 & H4.2)	p=0,36 H3.2 cannot be rejected	**p=0,0038** **H4.2 can be rejected** **Significant difference!**

Figure 16: Empirical finding for hypotheses H3 and H4

With one exception, the results here do not prove our hypotheses. We expected to see a difference for high-level changes, as here a good overview of the system is needed. It seems as though both TUC and EPC offer this overview, as the data does not differ significantly for TUC and EPC. With regard to specific changes, we did not expect to find differences. Regarding effectiveness, however, a significant difference was observable, as EPC did find a higher rate of defects compared to TUC. Taking into account the results for H1.2 and H2.2, it also seems that for specific changes that especially affect the lower-level specification parts, the tracing and tracking offered by EPC seem to have a positive effect on the analysis task.

The final hypothesis we analyzed regarded the session influence on the results. As sketched, we expected to see a positive influence regarding EPC for both sessions, alleviated for session 2 as learning effects apply. The results are sketched in Figure 17.

Copyright © Fraunhofer IESE 2007

Hypotheses	Efficiency	Effectiveness
Session 1	p=0,049 $H'_0$5.1 can be rejected H5.1 can be accepted	p=0,032 $H'_0$5.2 can be rejected H5.2 can be accepted
Session 2	p=0,23 $H'_0$5.3 cannot be rejected H5.3 cannot be accepted	p=0,42 $H'_0$5.4 cannot be rejected H5.4 cannot be accepted

Figure 17: Empirical finding for hypothesis H5

The hypotheses could be proven for session 1; regarding session 2, though, the hypotheses did not hold any longer. We assume that the learning effect regarding the system may have significantly influenced the results in such a way that the notational effect (EPC versus TUC) did not hold any longer. The question thus raised was: "Does the notational influence only have significant influence for change impact analysis activities regarding unknown system specification, i.e., without knowledge regarding the system?"

Roughly addressing the captured data, it has to be noted that in most of the cases (H1-H5), the times and detection rates acquired are lower (regarding time), respectively higher (regarding rate), for EPC than for TUC. As shown before, we could only prove this to be of significant for some of the hypotheses.

An interesting observation can be made regarding the subjective assessment of the participants (see Figure 14). The participants assessed TUC to be more efficient and effective than EPC. We assume that this assessment stems, on the one hand, from the fact that TUC happens to be a more widespread and commonly notation than EPC and, on the other hand, that participants feel it is harder to learn a graphical notation introducing new concepts (EPC) than a textual notation (TUC). Thus students feel more at ease using TUC than EPC and assessed this to be more effective and efficient. This assumption could briefly be confirmed by the students during presentation of the experimental results at a later stage.

4.3 Detailed results of statistical findings

The following sections address each hypothesis in detail and describe the results found. Each section is structured as follows.

- Hypothesis
- Descriptive results (TPC/IDR; Box plot)
- Statistical test results

Copyright © Fraunhofer IESE 2007

- Final conclusion

4.3.1 Hypothesis 1.1 - Efficiency (High Level):

Hypothesis

(High level) Individuals using EPC perform better than individuals using TUC with regard the time needed to perform task "determine high level impacts".

Descriptive results

Table 8 lists the TPC for each of the notations TUC and EPC, Figure 18 shows the respective box plot.

TPC_{TUC}	TPC_{EPC}
4.18	4.07

Table 8: TPC for subtask "determine high level impacts"

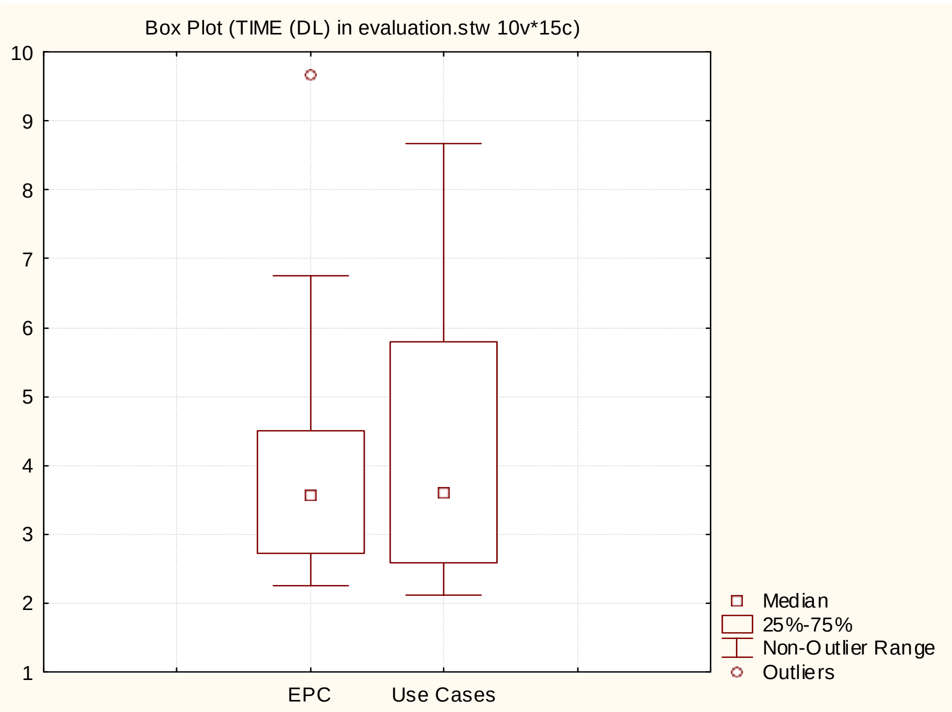


Figure 18: Box plot TPC subtask "determine high level impacts"

Copyright © Fraunhofer IESE 2007

Interpretation

The descriptive statistics do not indicate a strong difference in TPC values (the difference between the TPC values also being very small ~2%.

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 19) confirm the statistical findings and indicate that we cannot accept out hypothesis H1.1.

	Wilcoxon Matched Pairs Test (TIME (DL) in evaluation.stw) Marked tests are significant at p <,05000			
Pair of Variables	Valid N	T	Z	p-level
EPC & Use Cases	15	58,00000	0,113592	0,909561

Figure 19: Hypothesis testing H1.1

Final result

H1.1 cannot be accepted

4.3.2 Hypothesis 1.2 - Efficiency (Low Level):

Hypothesis

(Low level) Individuals using Use Cases perform better than individuals using EPC with regard the time needed to perform task "determine low level impacts".

Descriptive results

Table 9 lists the TPC for each of the notations TUC and EPC, Figure 20 shows the respective box plot.

TPC_{TUC}	TPC_{EPC}
4.39	2.73

Table 9: TPC for subtask

Copyright © Fraunhofer IESE 2007

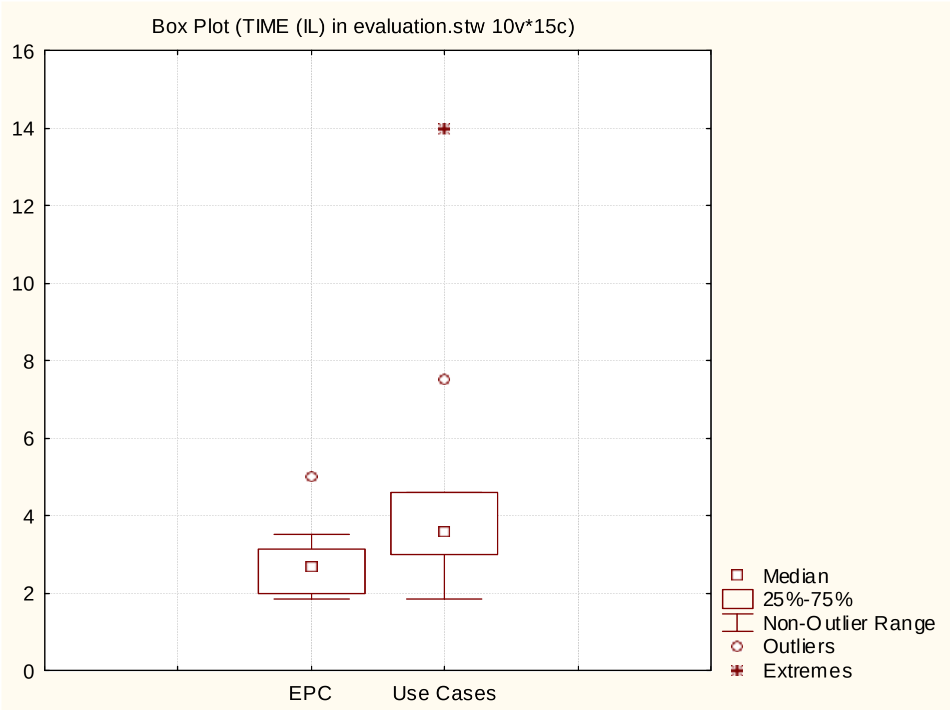


Figure 20: Box plot TPC subtask "determine low level impacts"

Interpretation

The difference between the TPC values is quite large (~60%). The statistical findings indicate a strong difference regarding the TPC value, but the difference is converse to the expected behavior. Statistical findings show that individuals using EPC perform better than individuals using TUC. We formulate this finding as alternative hypothesis H1.2 (a).

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 21) confirm the statistical findings and indicate that we can accept our hypothesis H1.2 (a).

	Wilcoxon Matched Pairs Test (TIME (IL) in evaluation.stw) Marked tests are significant at p <,05000			
Pair of Variables	Valid N	T	Z	p-level
EPC & Use Cases	15	12,00000	2,726217	0,006407

Figure 21: Hypothesis testing H1.1

Final result

H1.2 (a) can be accepted

Copyright © Fraunhofer IESE 2007

4.3.3 Hypothesis 1 - Efficiency (Overall):

Hypothesis

(Low level) Individuals using EPC perform better than individuals using TUC with regard to the time needed to perform change impact analysis task

Descriptive results

Table 10 lists the TPC for each of the notations TUC and EPC, Figure 22 shows the respective box plot.

TPC_{TUC}	TPC_{EPC}
3.32	3.87

Table 10: TPC for change impact analysis task

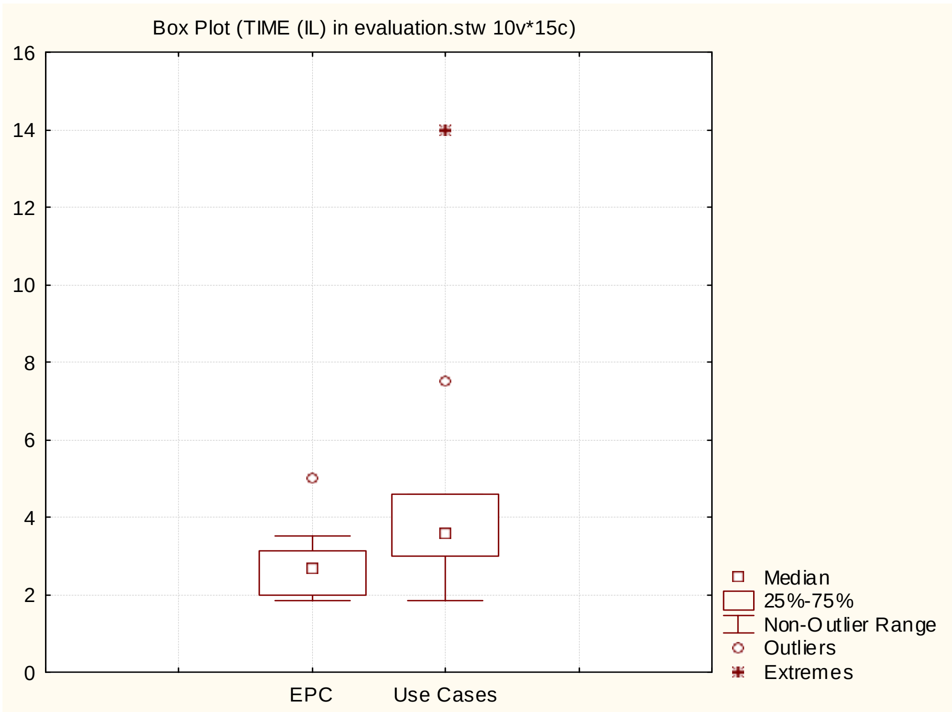


Figure 22: Box plot TPC change impact analysis task

Interpretation

The statistical data indicate a difference regarding the TPC value.

Copyright © Fraunhofer IESE 2007

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 23) confirm the statistical findings and indicate that we can accept our hypothesis H1.

Pair of Variables	Wilcoxon Matched Pairs Test (TIME (OVERALL) in evaluation.stw) Marked tests are significant at p <,05000			
	Valid N	T	Z	p-level
EPC & Use Cases	15	25,00000	1,987866	0,046827

Figure 23: Hypothesis testing H1

Final result

H1 can be accepted

4.3.4 Hypothesis 2.1 - Effectiveness (High Level):

Hypothesis

(High level) Individuals using EPC perform better than individuals using TUC with regard to impact detection rate regarding the task "determine high level impacts".

Descriptive results

Table 11 lists the TPC for each of the notations TUC and EPC, Figure 24 shows the respective box plot.

TPC_{TUC}	TPC_{EPC}
32%	28%

Table 11: TPC for subtask "determine high level impacts"

Copyright © Fraunhofer IESE 2007

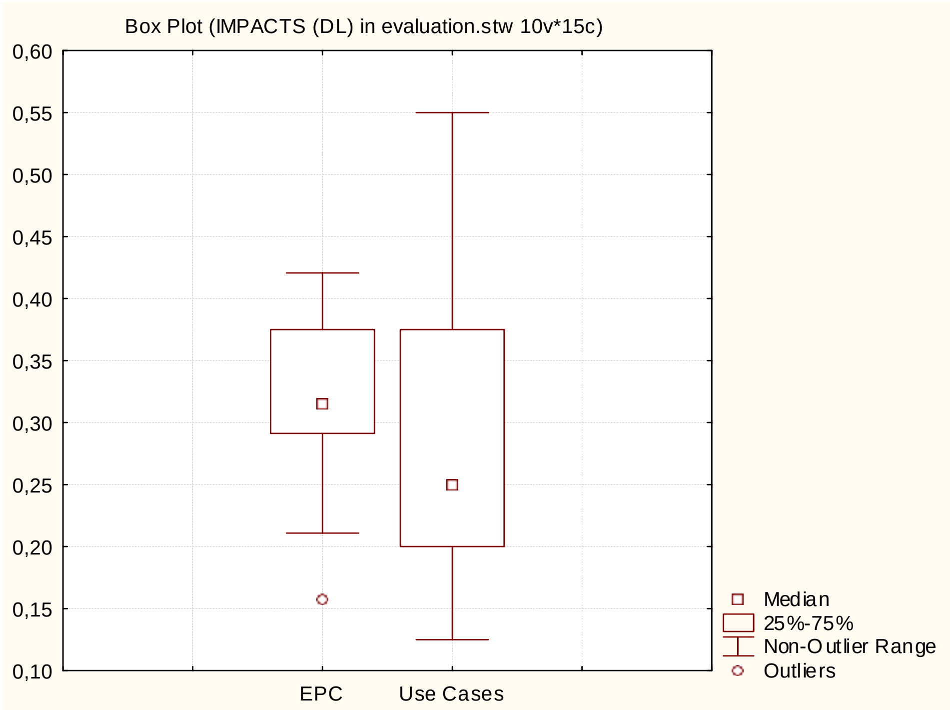


Figure 24: Box plot IDR subtask "determine high level impacts"

Interpretation

The descriptive statistics do not indicate a strong difference in TPC values (the difference between the TPC values also being very small ~2%.

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 25) confirm the statistical findings and indicate that we cannot accept out hypothesis H2.1.

	Wilcoxon Matched Pairs Test (IMPACTS (DL) in evaluation.stw) Marked tests are significant at p <,05000			
Pair of Variables	Valid N	T	Z	p-level
EPC & Use Cases	15	41,00000	1,079127	0,280532

Figure 25: Hypothesis testing H2.1

Final result

H2.1 cannot be accepted

Copyright © Fraunhofer IESE 2007

4.3.5 Hypothesis 2.2 - Effectiveness (Low Level):

Hypothesis

(Low level Individuals using Use Cases perform better than individuals using EPC with regard to impact detection rate regarding the subtask "determine low level impacts".

Descriptive results

Table 12 lists the IDR for each of the notations TUC and EPC, Figure 26 shows the respective box plot.

IDR_{TUC}	IDR_{EPC}
45%	55%

Table 12: IDR for subtask "determine low level impacts"

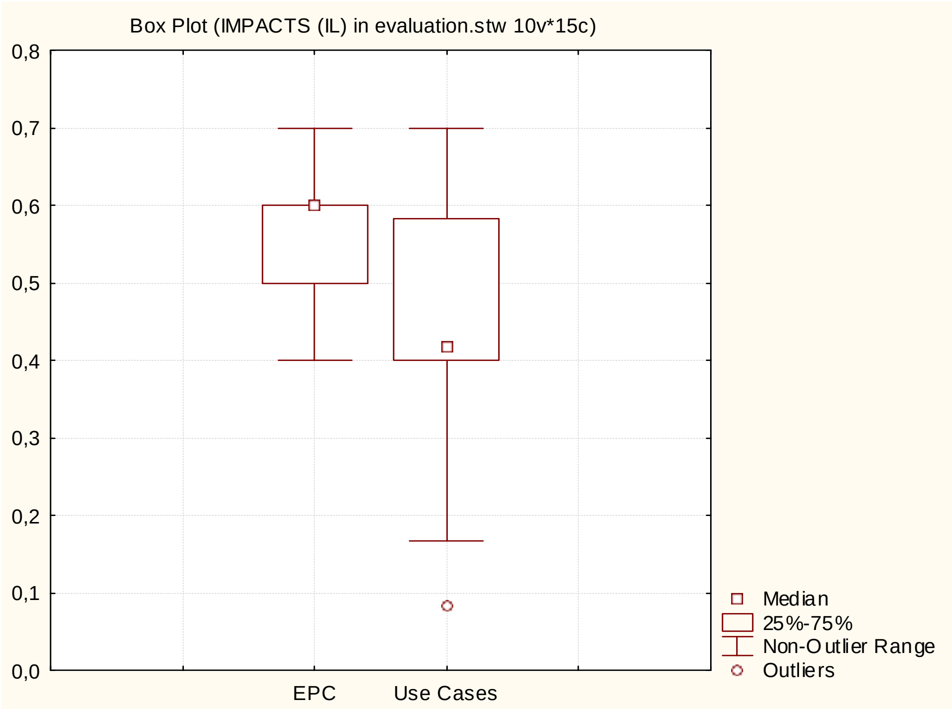


Figure 26: Box plot IDR subtask "determine low level impacts"

Interpretation

A difference between the TPC values can be observed. The statistical evaluation has to show whether this difference can be statistically proven. The difference is

Copyright © Fraunhofer IESE 2007

converse to the expected behavior (H2.2). Statistical findings show that individuals using EPC perform better than individuals using TUC with regard o IDR at this level. We formulate this finding as alternative hypothesis H2.2 (a)

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 27) confirm the statistical findings and indicate that we can accept our hypothesis H2.2 (a).

	Wilcoxon Matched Pairs Test (IMPACTS (IL) in evaluation.stw) Marked tests are significant at p <,05000			
Pair of Variables	Valid N	T	Z	p-level
EPC & Use Cases	15	17,00000	1,991741	0,046400

Figure 27: Hypothesis testing H2.2

Final result

H2.2 (a) can be accepted

4.3.6 Hypothesis 2 - Effectiveness (Overall):

Hypothesis

(Low level Individuals using EPC perform better than individuals using TUC with regard to impact detection rate regarding change impact analysis task

Descriptive results

Table 13 lists the IDR for each of the notations TUC and EPC, Figure 28 shows the respective box plot.

IDR_{TUC}	IDR_{EPC}
34%	40%

Table 13: IDR for change impact analysis task

Copyright © Fraunhofer IESE 2007

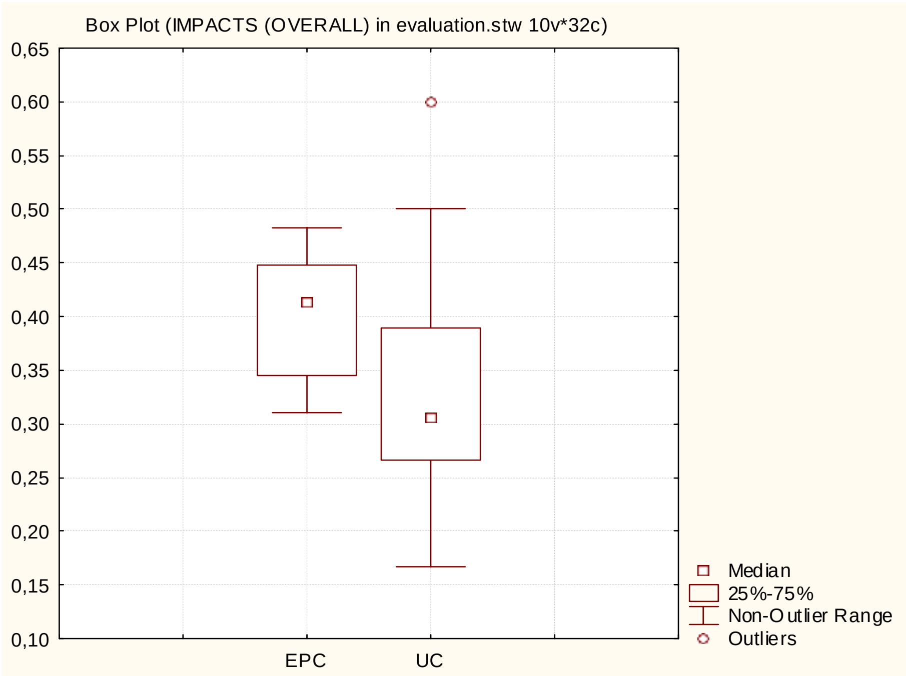


Figure 28: Box plot IDR change impact analysis task

Interpretation

The statistical data indicate a weak difference regarding the IDR value.

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 29) confirm the statistical findings and indicate that we cannot accept our hypothesis H2.

	Wilcoxon Matched Pairs Test (IMPACTS (OVERALL) in evaluation.stw) Marked tests are significant at p <,05000			
Pair of Variables	Valid N	T	Z	p-level
EPC & UC	15	32,00000	1,590293	0,111770

Figure 29: Hypothesis testing H2

Final result

H2 cannot be accepted

Copyright © Fraunhofer IESE 2007

4.3.7 Hypothesis 3.1 - Efficiency (High level change):

Hypothesis

(High level change) Individuals using EPC perform better than individuals using broad Use Cases with regard to the time needed to perform change impact analysis task

Descriptive results

Table 14 lists the TPC for each of the notations TUC and EPC, Figure 30shows the respective box plot.

TPC_{TUC}	TPC_{EPC}
3,3	3.58

Table 14: TPC for high level change

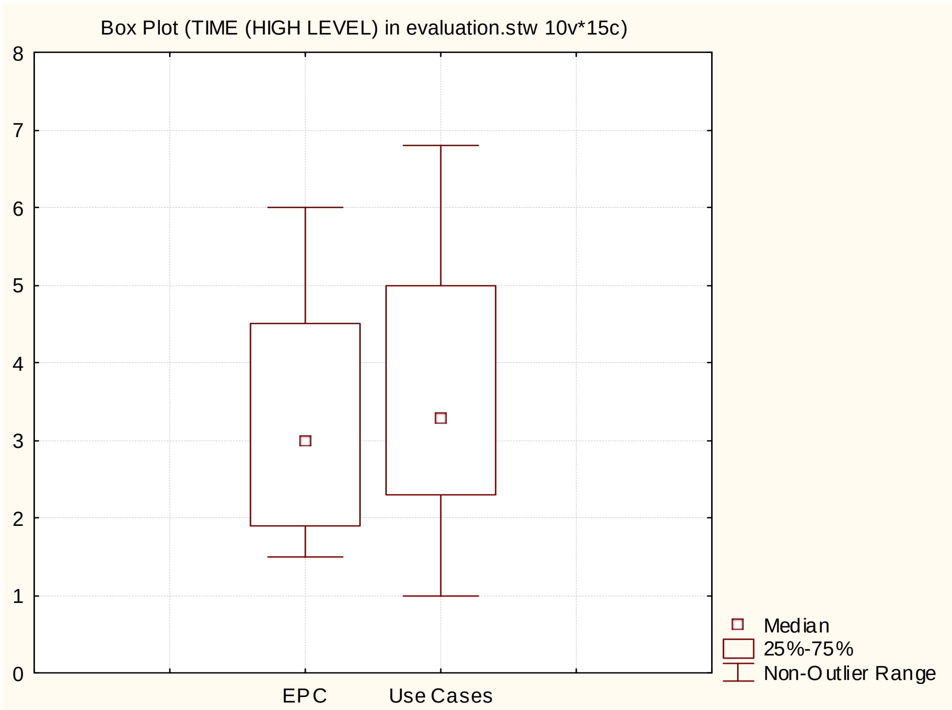


Figure 30: Box plot TPC for high level change

Interpretation

The statistical data indicate a very weak difference regarding the TPC value.

Copyright © Fraunhofer IESE 2007

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 31) confirm the statistical findings and indicate that we cannot accept our hypothesis H3.1.

	Wilcoxon Matched Pairs Test (TIME (HIGH LEVEL) in evaluation.stw) Marked tests are significant at p <,05000			
Pair of Variables	Valid N	T	Z	p-level
EPC & Use Cases	15	48,00000	0,681554	0,495521

Figure 31: Hypothesis testing H3.1

Final result

H3.1 cannot be accepted

4.3.8 Hypothesis 3.2 - Efficiency (Specific change):

Hypothesis

(Specific change) There is no difference between individuals using EPC and individuals using broad Use Cases regarding the time needed to perform change impact analysis task.

Descriptive results

Table 15 lists the TPC for each of the notations TUC and EPC, Figure 32 shows the respective box plot.

TPC_{TUC}	TPC_{EPC}
4	3.6

Table 15: TPC for high level change

Copyright © Fraunhofer IESE 2007

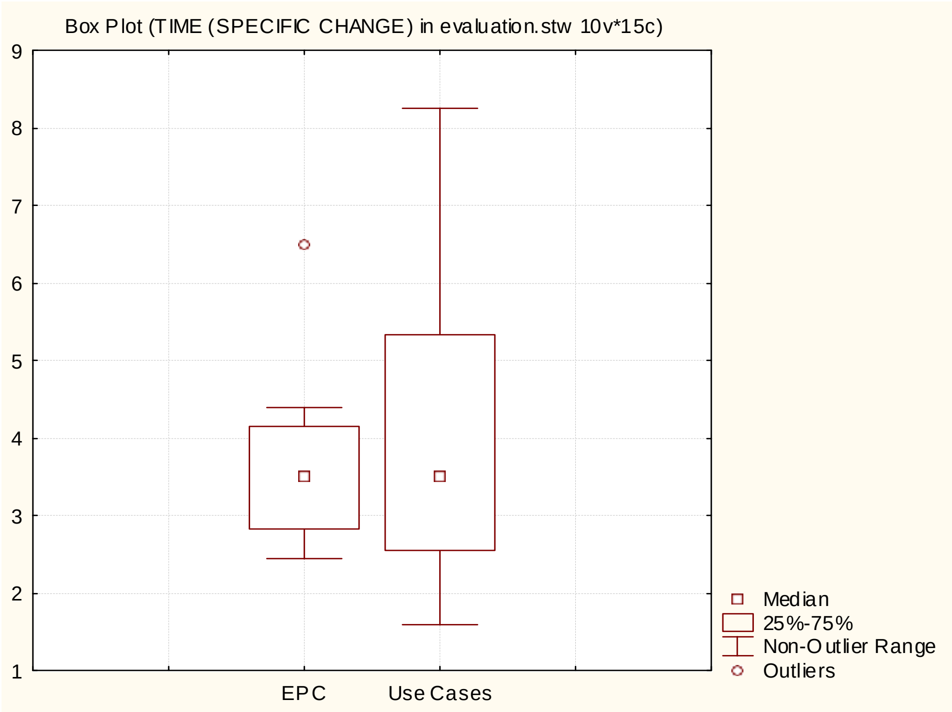


Figure 32: Box plot TPC for high level change

Interpretation

The statistical data indicate a very weak difference regarding the TPC value.

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 33) do not allow any reasoning as we cannot reject hypothesis 3.2.

	Wilcoxon Matched Pairs Test (TIME (SPECIFIC CHANGE) in evaluation.stw) Marked tests are significant at p <,05000			
Pair of Variables	Valid N	T	Z	p-level
EPC & Use Cases	15	44,00000	0,908739	0,363489

Figure 33: Hypothesis testing H3.2

Final result

H3.2 cannot be rejected

Copyright © Fraunhofer IESE 2007

4.3.9 Hypothesis 4.1 - Effectiveness (High level change):

Hypothesis

(High level change) Individuals using EPC perform better than individuals using broad Use Cases with regard to the impact detection rate regarding the change impact analysis task

Descriptive results

Table 16 lists the IDR for each of the notations TUC and EPC, Figure 34 shows the respective box plot.

IDR_{TUC}	IDR_{EPC}
35,8%	32,1%

Table 16: IDR for high level change

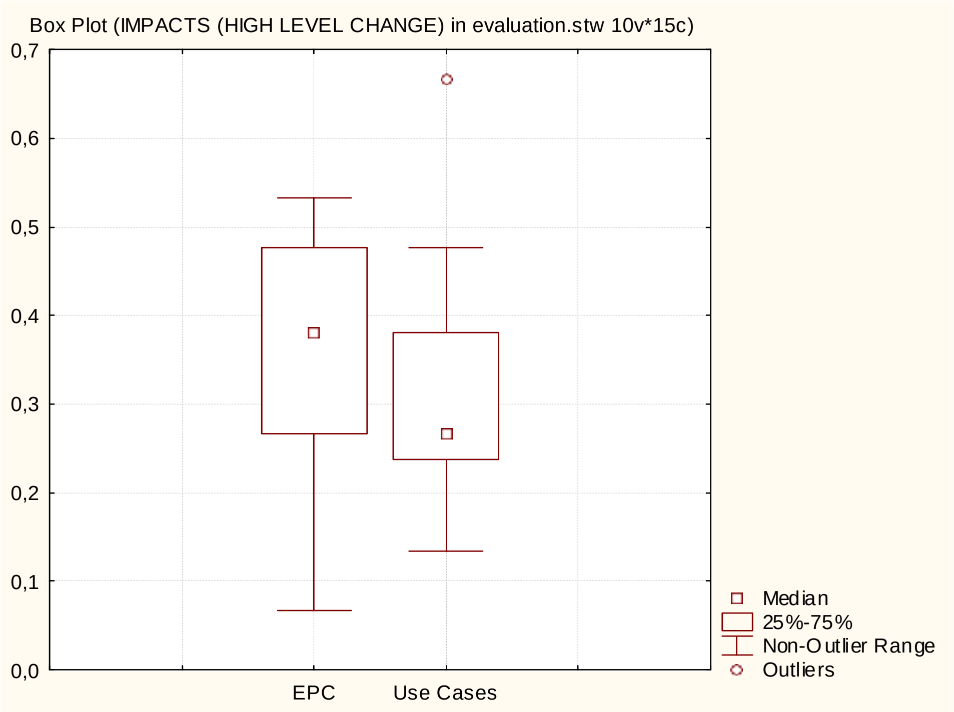


Figure 34: Box plot IDR for high level change

Interpretation

The statistical data indicate a very weak difference regarding the IDR value.

Copyright © Fraunhofer IESE 2007

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 35) confirm the statistical findings and indicate that we cannot accept our hypothesis H4.1.

Pair of Variables	Wilcoxon Matched Pairs Test (IMPACTS (HIGH LEVEL CHANGE) in evaluation.stw) Marked tests are significant at p <,05000			
	Valid N	T	Z	p-level
EPC & Use Cases	15	44,50000	0,880341	0,378675

Figure 35: Hypothesis testing H4.1

Final result

H4.1 cannot be accepted

4.3.10 Hypothesis 4.2 - Effectiveness (Specific change):

Hypothesis

(Specific change) There is no difference between individuals using EPC and individuals using broad Use Cases regarding the impact detection rate regarding the change impact analysis task.

Descriptive results

Table 17 lists the IDR for each of the notations TUC and EPC, Figure 36 shows the respective box plot.

IDR_{TUC}	IDR_{EPC}
45,1%	32,2%

Table 17: IDR for high level change

Copyright © Fraunhofer IESE 2007

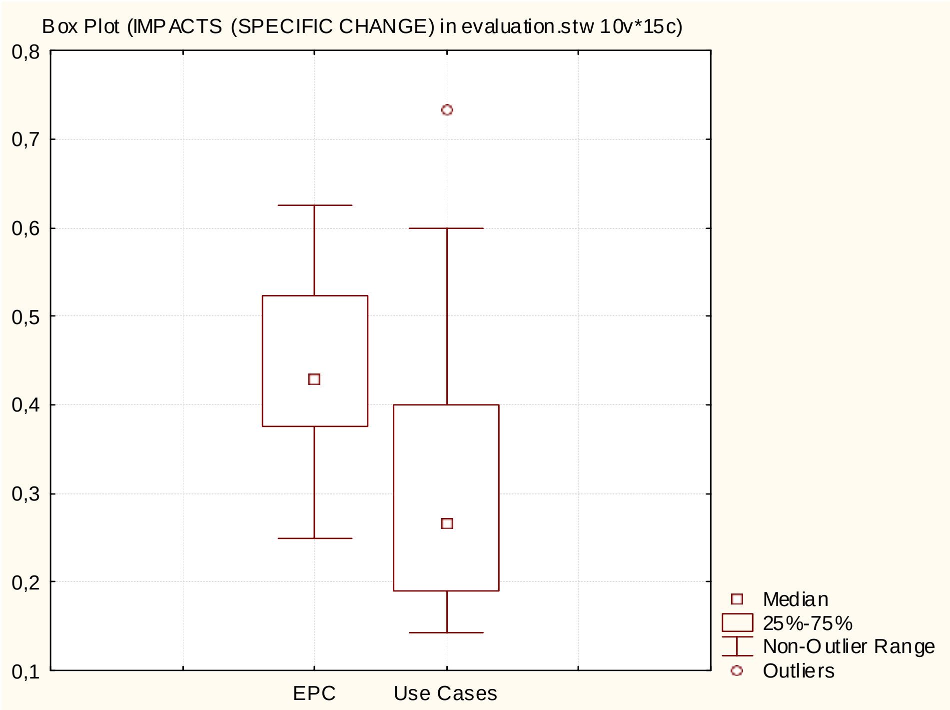


Figure 36: Box plot IDR for high level change

Interpretation

The statistical data indicate a difference regarding the IDR value.

Statistical test

The results of the non-parametric Wilcoxon-Test performed (see Figure 37) do confirm that we can reject hypothesis 4.2 indicating that there is a difference regarding IDR with regard to detecting specific changes.

	Wilcoxon Matched Pairs Test (IMPACTS (SPECIFIC CHANGE) in evaluation.stw) Marked tests are significant at p <,05000			
Pair of Variables	Valid N	T	Z	p-level
EPC & Use Cases	15	23,50000	2,073061	0,038167

Figure 37: Hypothesis testing H4.2

Final result

H4.2 cannot be rejected

4.3.11 Hypothesis 5.1 - Efficiency (session 1):

Hypothesis

(Session 1) Individuals using EPC perform better than individuals using broad Use Cases with regard to the time needed to perform change impact analysis task

Descriptive results

Table 18 lists the TPC for each of the notations TUC and EPC, Figure 38shows the respective box plot.

TPC_{TUC}	TPC_{EPC}
3.94	2,99

Table 18: TPC for session 1

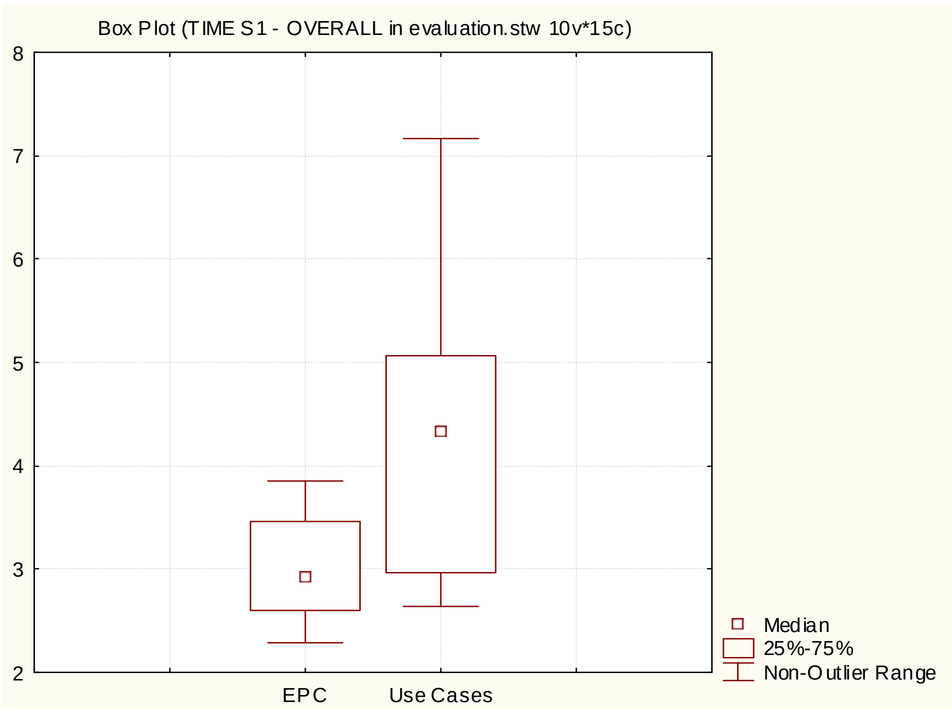


Figure 38: Box plot TPC for session 1

Interpretation

The statistical data indicate a difference regarding the TPC value.

Copyright © Fraunhofer IESE 2007

Statistical test

The results of the non-parametric Mann-Whitney-U-Test performed (see Figure 39) confirm the statistical findings and indicate that we can accept our hypothesis H5.1.

Mann-Whitney U Test (TIME S1 - OVERALL in evaluation.stw)
By variable Notation
Marked tests are significant at p <,05000

variable	Rank Sum EPC	Rank Sum UC	U	Z	p-level	Z adjusted	p-level	Valid N EPC	Valid N UC	2
Time needed	39,00000	81,00000	11,00000	-1,96737	0,049142	-1,96737	0,049142	7	8	0

Figure 39: Hypothesis testing H5.1

Final result

H5.1 can be accepted

4.3.12 Hypothesis 5.2 - Efficiency (Session 2):

Hypothesis

(Session 2) Individuals using EPC perform better than individuals using broad Use Cases with regard to the time needed to perform change impact analysis task

Descriptive results

Table 19 lists the TPC for each of the notations TUC and EPC, Figure 40 shows the respective box plot.

TPC_{TUC}	TPC_{EPC}
3,3	3.58

Table 19: TPC for session 2

Copyright © Fraunhofer IESE 2007

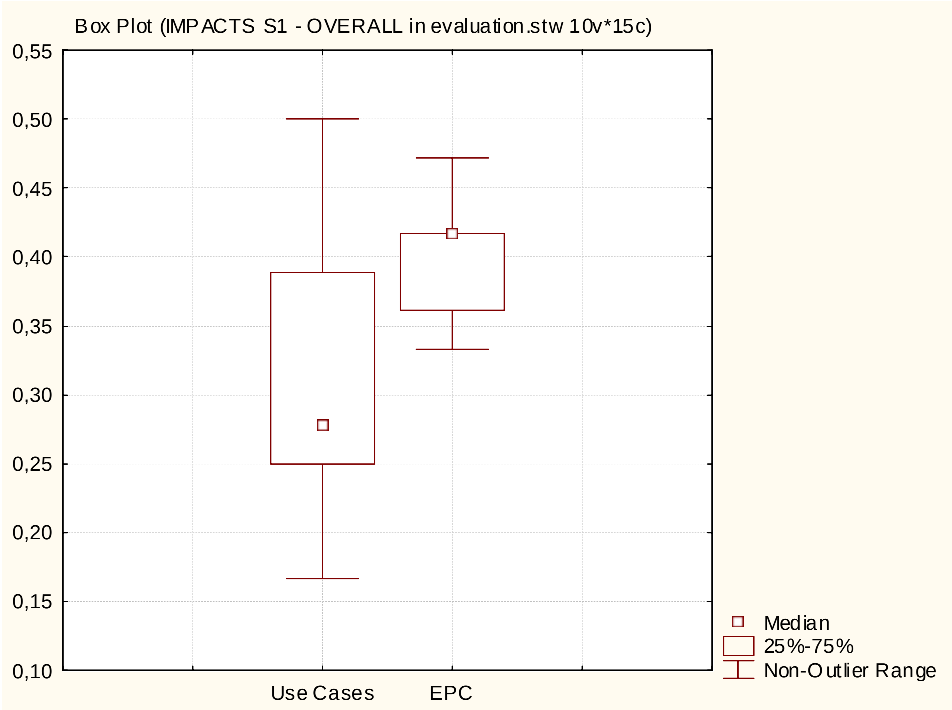


Figure 40: Box plot TPC for session 2

Interpretation

The statistical data indicate a difference regarding the TPC value.

Statistical test

The results of the non-parametric Mann-Whitney-U-Test performed (see Figure 41) confirm the statistical findings and indicate that we can accept our hypothesis H5.2.

	Mann-Whitney U Test (IMPACTS S1 - OVERALL in evaluation.stw) By variable Notation Marked tests are significant at p <,05000									
variable	Rank Sum EPC	Rank Sum UC	U	Z	p-level	Z adjusted	p-level	Valid N EPC	Valid N UC	2*1sided exact p
Impact detection ra		45,50000	9,500000	2,140959	0,032278	2,154467	0,031204	7	8	0,028904

Figure 41: Hypothesis testing H5.2

Final result

H5.2 can be accepted

Copyright © Fraunhofer IESE 2007

4.3.13 Hypothesis 5.3 - Effectiveness (Session 1):

Hypothesis

(Session 1) Individuals using EPC perform better than individuals using broad Use Cases with regard to the impact detection rate regarding the change impact analysis task

Descriptive results

Table 20 lists the IDR for each of the notations TUC and EPC, Figure 42 shows the respective box plot.

IDR_{TUC}	IDR_{EPC}
35,8%	32,1%

Table 20: IDR for session 1

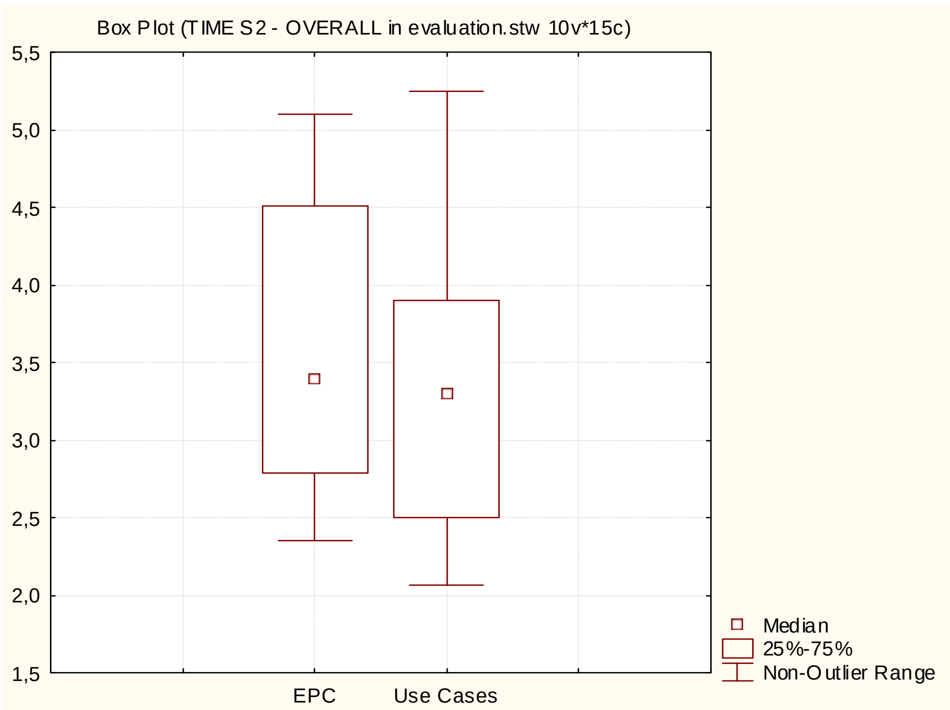


Figure 42: Box plot IDR for session 1

Interpretation

The statistical data indicate a very weak difference regarding the IDR value.

Copyright © Fraunhofer IESE 2007

Statistical test

The results of the non-parametric Mann-Whitney-U-Test performed (see Figure 43) confirm the statistical findings and indicate that we cannot accept our hypothesis H5.3.

Mann-Whitney U Test (TIME S2 - OVERALL in evaluation.stw)
By variable Notation
Marked tests are significant at p <,05000

variable	Rank Sum EPC	Rank Sum UC	U	Z	p-level	Z adjusted	p-level	Valid N EPC	Valid N UC	2 e
Time needed	66,00000	54,00000	26,00000	0,231455	0,816961	0,231455	0,816961	8	7	0

Figure 43: Hypothesis testing H5.3

Final result

H5.3 cannot be accepted

4.3.14 Hypothesis 5.4 - Effectiveness (session 2):

Hypothesis

(Session 1) Individuals using EPC perform better than individuals using broad Use Cases with regard to the impact detection rate regarding the change impact analysis task

Descriptive results

Table 21 lists the IDR for each of the notations TUC and EPC, Figure 44 shows the respective box plot.

IDR_{TUC}	IDR_{EPC}
35,8%	32,1%

Table 21: IDR for session 1

Copyright © Fraunhofer IESE 2007

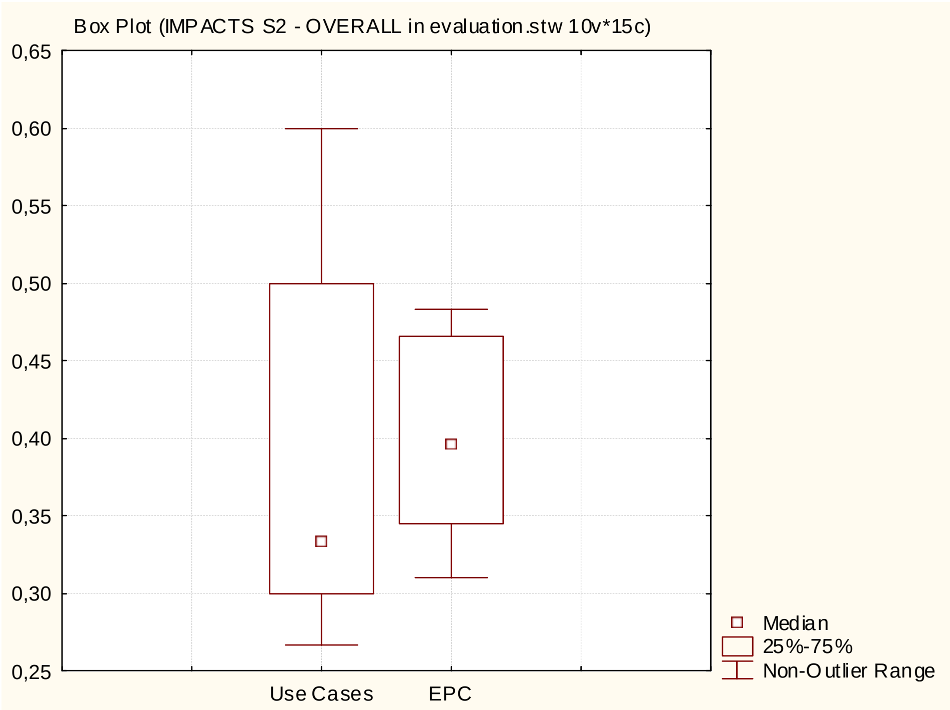


Figure 44: Box plot IDR for session 1

Interpretation

The statistical data indicate a very weak difference regarding the IDR value.

Statistical test

The results of the non-parametric Mann-Whitney-U-Test performed (see Figure 45) confirm the statistical findings and indicate that we cannot accept our hypothesis H5.4.

	Mann-Whitney U Test (IMPACTS S2 - OVERALL in evaluation.stw) By variable Notation Marked tests are significant at p <,05000								
variable	Rank Sum EPC	Rank Sum UC	U	Z	p-level	Z adjusted	p-level	Valid N EPC	V
Impact detection rate	71,00000	49,00000	21,00000	0,810093	0,417888	0,812271	0,416637	8	

Figure 45: Hypothesis testing H5.4

Final result

H5.4 cannot be accepted

Copyright © Fraunhofer IESE 2007

5 Summary

Our hypothesis that “Individuals using EPC perform better than individuals using TUC with regard to the time needed to perform changing tasks” can be accepted while the hypothesis that “Individuals using EPC perform better than individuals using TUC with regard to the impact detection rate” is not acceptable.

Thus we can say that a general statement regarding the overall task of change impact analysis the EPC are significantly better as opposed to TUC with regard to the efficiency of the task. The time needed to find impacts using EPC is significantly lower compared to TUC. As briefly sketched in section 3.1 we think that the reason for this difference lies in the graphical form of the EPC allowing a much better and faster overview over the system and the tracing to the lower levels, which is needed for the change impact analysis thus speeding up the task significantly. No statement can be made regarding effectiveness though; the results do not differ significantly.

Analyzing the impact analysis in more detail (i.e. regarding the subtasks high and low level impact analysis as described in section 3.3) we saw that no significant difference could be observed for high level change impact analysis. Here our hypotheses did not hold, thus on a higher level the graphical overview may not be as helpful as expected. However a significant difference was observable with regard to low level change impact analysis, i.e. the tracing of changes to the lower level and the impact analysis on the lower level. In these cases our initial hypothesis (i.e. TUC are more effective and more efficient than EPC) did not hold, as the descriptive statistic data indicate an opposite trend: the values for TPC and IDR are better for EPC as they are for TUC. The tracing and impact analysis thus benefit from the EPC overview facilitating the impact analysis activity.

Furthermore, we expected to see a difference for high level changes, as here a good overview over the system is needed. It seems that both, TUC and EPC, offer this overview, as the data does not differ significantly for TUC and EPC. With regard to specific changes we did not expect to find differences. Regarding the effectiveness however a significant difference was observable as EPC found a higher rate of defects compared to TUC. It seems that for specific changes that especially affect the lower level specification parts, the tracing and tracking offered through EPC have a positive effect on the analysis task.

The final hypothesis we analyzed regarded the session influence on the results. As sketched, we expected to see a positive influence regarding EPC for both sessions, alleviated for session 2 as learning effects apply. The hypotheses could

Copyright © Fraunhofer IESE 2007

be proven for session 1; regarding session 2, the hypotheses did not hold any longer. We assume that the learning effect regarding the system may have significantly influenced the results in such a way that the notational effect (EPC versus TUC) did not hold any longer. The question thus raised was: "Does the notational influence only have significant influence for change impact analysis activities regarding unknown system specification, i.e. without knowledge regarding the system?"

Roughly addressing the captured data, it has to be noted, that in most of the cases (H1-H5) the times and detection rates acquired are lower (regarding time) respectively higher (regarding rate) for EPC than for TUC. As shown afore we could only prove this to be of significant for some of the hypotheses.

An interesting observation can be made regarding the subjective assessment of the participants. The participants assessed TUC to be more efficient and effective than EPC. We assume that this assessment stems at the one hand from the fact that TUC happen to be a more widespread and common notation than EPC and at the other hand that participants feel it be harder to learn a graphical notation introducing new concepts (EPC) than a textual notation (TUC). Thus students feel more at ease using TUC than EPC and assessed this to be more effective and efficient. This assumption could briefly be confirmed by the students during presentation of the experimental results at a later stage.

Copyright © Fraunhofer IESE 2007

References

[Camp63] Campbell, D., Stanley, J.: Experimental and Quasi-Experimental Design for Research. Houghton Mifflin Company, 1963

[Cock01] Cockburn, A.: Writing Effective Use Cases. Addison-Wesley, 2001

[NiSK00] Nikula, U., J. Sajaniemi and H. Kälviäinen (2000). A State-of-the-Practice Survey on Requirements Engineering: Industry Expectations on Education and Technology Transfer. Fifth International Conference on Software Process Improvement Research, Education and Training - INSPIRE V, The University of North London, London, Great Britain, The British Computer Society. pp. 13-24.

[PaKo04] Paech, B., Kohler, K.: Task-Driven Requirements in Object-Oriented Development. In: Perspectives on Software Requirements. Kluwer Academic Publisher, 2004

Copyright © Fraunhofer IESE 2007

Document Information

Title:	Influence of requirements specification notation on change impact analysis - An Empirical Investigation
Date:	May 2007
Report:	IESE-106.07/E
Status:	Final
Distribution:	Public

Copyright 2007, Fraunhofer IESE.
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means including, without limitation, photocopying, recording, or otherwise, without the prior written permission of the publisher. Written permission is not needed if this publication is distributed for non-commercial purposes.