|  |  |
| --- | --- |
| Workpackage 4: EVALUATION  D4.4.1  eVALUATION mETHODOLOGY FOR THE EVALUATION OF THE VALIDATORS | |
| UsiXML_Transparent2.png | Project acronym: UsiXML  Project full title: User interface eXtensible Mark-up Language  ITEA label n° 08026 |

**DOCUMENT CONTROL**

|  |
| --- |
| **Deliverable N°**: D4.4  **Due Date**: 08/2012  Delivery Date: 02/2013 |

|  |
| --- |
| **Short Description**: This documents describes the methodology of evaluation proposed for the evaluation of the validators developed in the Workpackage 4.1 of the Itea2 UsiXML project. |

|  |
| --- |
| **Lead Partner**: SymbiaIT  **Contributors:** UCLM, Thales  **Made available to**: Workgroup |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rev | Date | Author | Checked by | Internal Approval | Description |
| Draft | 28/08/12 | Arturo S. García (SymbiaIT)  Francisco Montero (UCLM) |  |  | Initial version |
|  | 01/10/12 | Charles Robinson (Thales) |  |  | revision |
|  | 18/10/12 | Arturo S. García (SymbiaIT)  Francisco Montero (UCLM) |  |  | revision |
|  | 19/10/12 | Charles Robinson (Thales) |  |  | revision |
|  | 16/11/12 | Arturo S. García (SymbiaIT)  Francisco Montero (UCLM) |  |  | revision |

CONTENTS

[1. INTRODUCTION 5](#_Toc354566814)

[1.1. Scope of the evaluation 5](#_Toc354566815)

[2. Glossary of terms 5](#_Toc354566816)

[3. Strategy of evaluation. The big picture. 7](#_Toc354566817)

[4. Conclusions and further work 8](#_Toc354566818)

[5. References 8](#_Toc354566819)

[Appendix A. Template for the evaluation of validator at runtime 9](#_Toc354566820)

[A.I. Purpose and Objectives 9](#_Toc354566821)

[A.II. Audience 9](#_Toc354566822)

[A.III. Template 10](#_Toc354566823)

[A.IV. Metrics for the runtime evaluation of the validator 17](#_Toc354566824)

[Appendix B. Template for the evaluation of validator at design-time 19](#_Toc354566825)

[B.I. Purpose and Objectives 19](#_Toc354566826)

[B.II. Audience 19](#_Toc354566827)

[B.III. Template 20](#_Toc354566828)

[B.IV. Metrics for the evaluation of the modelling activities 27](#_Toc354566829)

# INTRODUCTION

The main aim of this document is to set out the approach that will be followed for the evaluation of the validators developed under the scope of the UsiXML project. This task, led by SymbiaIT, is to prove the viability of the model-driven engineering approach for μ7-compliant UI. Quantitative and qualitative metrics will be selected in order to evaluate the results of the demonstrators developed within this project. These metrics will concern designers, developers, and end users. For instance, coverage of the μ7 capabilities, total design time, portion of code produced from the models, subjective user satisfaction and task completion time.

Assessment procedures and protocols are provided as appendices at the end of this document, these being:

* A template for reporting the runtime evaluation of the validator (Appendix A).
* A template for reporting the design-time evaluation of the validator (Appendix B).

## Scope of the evaluation

The main aim of the evaluation described in this document and its appendices is to evaluate the quality of the process defined by UsiXML based on the quality of the tools conforming to this methodology. In other words, the evaluation is aimed at assessing UsiXML through the use of their associated products.

These evaluation activities are related to the Task 4.4 of the ITEA2 UsiXML project:

Task 4.4 Evaluation: in order to prove the viability of the model-driven engineering approach for µ7-compliant UI suggested in this project, quantitative and qualitative metrics will be selected in order to evaluate the results of the validators developed in the aforementioned tasks. These metrics will concern designers, developers, and end users. For instance, coverage of the µ7 capabilities, total design time, portion of code produced from the models, subjective user satisfaction, and task completion time.

This document introduces informative considerations related to the evaluation at runtime and design-time and relevant to UsiXML.

# Glossary of terms

For application to all parts of this document, the following definitions are introduced.

|  |  |
| --- | --- |
| **Concept** | **Description** |
| Demonstrator | This is a functional example that shows the way in which concepts of UsiXML may be applied to products in industry, commerce and academia. These are used to present the benefits of the UsiXML approach. |
| Design-time evaluation | Evaluations carried out during the engineering process of UsiXML-based products in order to assess how well the UsiXML models and tools were used to create these products. |
| Designer/modeller | These terms will refer to the person that has used the UsiXML tools to model the validator. |
| Developer | A person that uses the models to code the final application or part of it. This person may or may not also be a modeller. The models are assessed during implementation of the code and these may require amendments. |
| Effectiveness | Measures of effectiveness relate the goals or sub-goals of the user to the accuracy and completeness with which these goals can be achieved. |
| Efficiency | Measures of efficiency relate the level of effectiveness achieved to the expenditure of resources. |
| Error at runtime: user error | Errors made by the users while carrying out tasks during a run time test. |
| Error in the design-time: specification error | This represents when the designer needs to return to edit earlier models due to “misrepresentation of the specification”. |
| Expert designer | A person with a high degree of skill in or knowledge of designing, in the context of UsiXML. |
| Expert user | User who has previous experience with the validator. With respect to some validators, it can be considered that a user with long experience in the domain in which the validator is applied could be an expert user if the interface and functionality replicates a previous one. |
| Internal and external quality | These terms are related to the ISO/IEC 9126 standard, and refer to the quality of the software product: “The model specifies six characteristics for internal and external quality, which are further subdivided into subcharacteristics. These subcharacteristics are manifested externally when the software is used as a part of a computer system, and are a result of internal software attributes." |
| Metric | The defined measurement method and the measurement scale. |
| Quality in use | The extent to which a product used by specified users meets their needs to achieve specified goals with effectiveness, productivity and satisfaction in specified contexts of use. |
| Quality model | The set of characteristics and the relationships between them which provide the basis for specifying quality requirements and evaluating quality. |
| Quality evaluation | Systematic examination of the extent to which an entity is capable of fulfilling specified requirements. |
| Requirements | Includes the functional and non-functional requirements of the validators. |
| Runtime evaluation | Evaluation carried out with end users in order to test the overall usability of the validator developed in the context of the UsiXML ITEA2 project. |
| Satisfaction | Measures of satisfaction describe the comfort and acceptability of the use. |
| Satisfaction questionnaire | Set of questions arranged in order to know the satisfaction of the user/modeller with the validator/use of UsiXML. |
| SUS questionnaire | System Usability Scale (SUS) is a simple, ten-item attitude Likert scale giving a global view of subjective assessments of usability. The scores for the individual items are not meaningful on their own, but the importance is on the composite measure. |
| Task in the design-time context | Task to be performed by the designer when modelling the validator. They will be related to the use of UsiXML models and tools. |
| Task in the runtime context | Task to be performed by users in the run time tests. They will be related to tasks of the domain of the validator. |
| User | Someone who uses the interface at run time. |
| Validation | Confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled. |
| Validator | This is a functional example that is used to prove concepts within UsiXML. |
| µ7 capabilities | The UsiXML language relies on the µ7 concept, which is defined as multi-device, multi-platform, multi-user, multi-linguality / culturality, multi-organisation, multi-context, multi-modality. |
| Verification | Confirmation by examination and provision of objective evidence that specified requirements have been fulfilled. |

Table 1. Glossary of terms.

# Strategy of evaluation. The big picture.

The strategy followed for the design of the evaluation of the validators is based on the standard ISO/IEC 25010:2011 (this is a revision of the ISO/IEC 9126-1:2001). This international standard, ISO/IEC 25010, describes a two-part model for software product quality: a) internal quality and external quality, and b) quality in use. The first part of the model specifies six characteristics for internal and external quality, which are further subdivided into subcharacteristics. These subcharacteristics are manifested externally when the software is used as a part of a computer system, and are a result of internal software attributes.

The scope of our evaluation proposal consists of two dimensions (see Figure 1):

* The process (process quality)
* The effect of software product quality (quality in use quality)

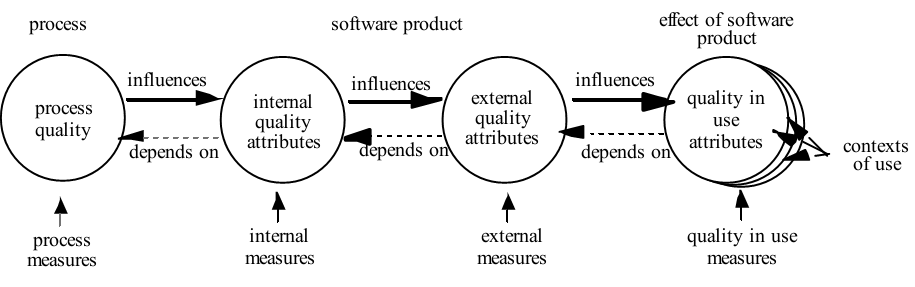


Fig. 1. Quality in the lifecycle

(source: ISO/IEC 25010:2011).

Figure 1 depicts the quality lifecycle defined in the standard ISO 25010. The first circle refers to the quality of the process. In order to evaluate the process quality two online questionnaires were developed by SymbiaIT and they were sent to the relevant developers. This activity was conducted in 2011 during the months of May and November. These two questionnaires were used to gather information about the validators. The first one about the companies and personnel behind them, and the second one on topics such as their previous experience with UsiXML and their attitude about Model-driven Development (MDD). These questionnaires were based on [Hutchinson, Whittle et al. 2011]. The results of both questionnaires were sent to the partners using the WorkPackage mailing list. In any case, additional information related to the development processes could be gathered from partners involved in this project. In this sense, a set of metrics related to design-time metrics were identified and suggested by SymbiaIT.

The second and third circles refer to the quality of the software product (internal and external quality attributes). Due to time constraints and the importance of the runtime evaluation, it is believed that it is more important to focus this work on the last circle.

The last circle covers the quality of the product at runtime. In this sense, Appendix A of this document shows a template based on the Common Industry Format (CIF) (additional information in ISO/IEC 9126-4:2004, see Appendixes F and G of that international standard) to provide the partners with a simple way to create the usability report of the evaluations they will perform on their validators, and an example evaluation will be provided. There are also some general user testing recommendations included.

The quality in use evaluation of validators is being addressed by using user testing techniques. In order to analyse the activities of the users in these testing activities, a set of metrics were identified and included in this deliverable (see Appendix A section IV).

Also, some evaluation is performed on UsiXML language and its associated tools (design-time evaluation). Since some of the tools are not mature enough to be assessed in their entirety, the design-time evaluation will focus on effectiveness, productivity, satisfaction and µ7 coverage. This strategy intentionally leaves aside any comparison between a validator developed following traditional methodologies and an MDD approach, since this would assume that two versions of the validators need to be developed (or that a previous version of them should exist), and it is not likely to happen in all cases.

To sum up:

* Getting to know the validators and the development groups of each partner (Questionnaire 1).
* Use of UsiXML, knowledge, predisposition and opinion about it and MDE (Questionnaire 2).
* Runtime evaluation of the validators (Appendix A).
  + Template for the evaluation.
  + Set of metrics.
* Design-time evaluation (Appendix B).
  + Template for the evaluation.
  + Set of metrics.

# Conclusions and further work

The methodology proposed has been followed by several partners that finished their evaluation successfully. Future work for this methodology of evaluation could include the comparison of traditionally developed applications with applications developed using UsiXML in order to extract the benefits of the UsiXML approach.

# References

|  |  |
| --- | --- |
| [Brooke, J., 1996] | John Brook 1996. SUS: A 'quick and dirty' usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., & McClelland, I. L. (Eds.), Usability evaluation in industry, 189-194. London, UK: Taylor & Francis. |
| [Hutchinson, Whittle et al. 2011] | John Hutchinson, Jon Whittle, Mark Rouncefield, and Steinar Kristoffersen. 2011. Empirical assessment of MDE in industry. In Proceedings of the 33rd International Conference on Software Engineering (ICSE ’11). ACM, New York, NY, USA, 471-480. |
| [ISO/IEC 25010:2011] | Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models. http://www.iso.org. |

1. Template for the evaluation of validator at runtime
   1. Purpose and Objectives

A template is proposed for the generation of a report of end user evaluations. This template is based on the Common Industry Format (ISO/IEC 9126-4:2004). The overall purpose of the Common Industry Format (CIF) for Usability Test Reports is to promote incorporation of usability as part of the procurement decision-making process for interactive products. Examples of such decisions include purchasing, upgrading and automating. It provides a common format for human factors engineers and usability professionals in supplier companies to report the methods and results of usability tests to customer organizations.

* 1. Audience

The people using this template will be those in charge of performing the evaluation of the validators at each partner.

* 1. Template

Title Page

|  |  |
| --- | --- |
| Common Industry Format Usability Test  Validator name and version number  Evaluation at runtime  Report by: Evaluator name  Company name/logo  Date when the test was conducted  The date the report was prepared | |
| UsiXML_Transparent2.png | Project acronym: UsiXML  Project full title: User interface eXtensible Mark-up Language  ITEA label n° 08026 |

Contents (mostly extracted from ISO/IEC 9126-4:2004)

1. Introduction

1.1. Executive Summary

1.2. Full Validator Description

1.3. Test Objectives

2. Method

2.1. Participants

2.2. Context of Validator Use in the Test

2.3. Experimental Design

2.3.1 Procedure

2.3.2 Participant General Instructions

2.4. Runtime evaluation Metrics

3. Results

3.1. Treatment of data

3.2. Presentation of the Results

3.2.1 Performance Results

3.2.2 Satisfaction Results

4. Appendices

1. Introduction
   1. Executive Summary

This section provides a high level overview of the test. Possible content is:

* Identity and description of the validator
* Summary of the method(s) of the test including the number of and type of participants and their tasks
* Results expressed as mean scores or other suitable measure of central tendency
  1. Full Validator Description

This section identifies the formal validator name and release or version. It describes what parts of the validator were evaluated. This section should also specify:

* The user population for which the validator is intended
* A brief description of the environment in which it should be used
* The type of user work that is supported by the validator
  1. Test Objectives

This section describes all of the objectives for the test and any areas of specific interest. Possible objectives include testing user performance of work tasks and subjective satisfaction in using the product. This section should include:

* The functions and components of the validator with which the user directly and indirectly interacted in this test.
* If the validator component or functionality that was tested is a subset of the total validator, explain the reason for focusing on the subset.

1. Method

This section must provide sufficient information to allow an independent tester to replicate the procedure used in testing.

* 1. Participants

This section describes the users who participated in the test in terms of demographics, professional experience, computing experience and special needs.

A general description should include important facts such as:

* The total number of participants tested.
* Segmentation of user groups tested. Example: novice and expert programmers.
* The key characteristics and capabilities expected of the user groups being evaluated.
* How participants were selected and whether they had the essential characteristics and capabilities.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Gender** | **Age** | **Education** | **Occupation/role** | **Professional experience** | **Computer experience** | **Product experience** |
| **P1** |  |  |  |  |  |  |  |
| **P2** |  |  |  |  |  |  |  |
| **Pn** |  |  |  |  |  |  |  |

Depending on the validator and the evaluation performed, the last three columns may need categories/choices specified (years: 0-5, 5-10, >10) in order to compare between participants.

Table 2. Table to be used to collect information regarding participants in runtime evaluation.

* 1. Context of Validator Use in the Test

This section describes the tasks, scenarios and conditions under which the tests were performed, the tasks that were part of the evaluation, the platform on which the application was run, and the specific configuration operated by test participants. Any known differences between the evaluated context and the expected context of use should be noted in the corresponding subsection.

* 1. Experimental Design

Describe the logical design of the test. Define independent variables and control variables. Briefly describe the measures for which data were recorded for each set of conditions.

* + 1. Procedure

This section details the test protocol:

* Give operational definitions of measures and any presented independent variables or control variables.
* Specify the steps that the evaluation team followed to execute the test sessions and record data.
* State whether participants were paid or otherwise compensated.
  + 1. Participant General Instructions

Include here or in an appendix all instructions given to the participants.

* 1. Runtime evaluation Metrics

Explain what measures have been used for each category of usability metrics: effectiveness, efficiency and satisfaction. Use the generic set of metrics selected for the UsiXML project and extend it if the evaluation of the validator needs it. Additional information is available in section A.IV.

1. Results

This is the second major technical section of the report. It includes a description of how the data were scored, reduced, and analysed. It provides the major findings in quantitative formats.

* 1. Treatment of Data

In order to treat gathered data, several activities are suggested:

* Data Scoring: The method by which the data collected were scored should be described in sufficient detail to allow replication of the data scoring methods by another organization if the test is repeated.
* Data Reduction: The methods by which the data were reduced should be described in sufficient detail to allow replication of the data reduction methods by another organization if the test is repeated.
* Statistical Analysis: The method by which the data were analysed should be described in sufficient detail to allow replication of the data analysis methods by another organization if the test is repeated.
  1. Presentation of the Results

Both tabular and graphical presentations of results should be included. The data may be accompanied by a brief explanation of the results but detailed interpretation is discouraged.

* + 1. Performance Results

It is recommended that efficiency and effectiveness results be tabulated across participants on a per unit task basis. A summary table showing total mean task times and completion rates across all tasks should be presented.

Task A

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **User #** | **Unassisted Task Effectiveness [(%)Complete]** | **Assisted Task Effectiveness**  **[(%)Complete]** | **Task Time (min)** | **Effectiveness / Mean Time-On-Task** | **Errors** | **Assists** |
| **1** |  |  |  |  |  |  |
| **2** |  |  |  |  |  |  |
| **N** |  |  |  |  |  |  |
| **Mean** |  |  |  |  |  |  |
| **Standard Deviation** |  |  |  |  |  |  |
| **Min** |  |  |  |  |  |  |
| **Max** |  |  |  |  |  |  |

Table 3. Table to be used to collect information about one task.

Summary

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **User #** | **Total Unassisted Task Effectiveness [(%)Complete]** | **Total Assisted Task Effectiveness**  **[(%)Complete]** | **Total Task Time (min)** | **Effectiveness / Mean Time-On-Task** | **Total**  **Errors** | **Total**  **Assists** |
| **1** |  |  |  |  |  |  |
| **2** |  |  |  |  |  |  |
| **N** |  |  |  |  |  |  |
| **Mean** |  |  |  |  |  |  |
| **Standard Deviation** |  |  |  |  |  |  |
| **Min** |  |  |  |  |  |  |
| **Max** |  |  |  |  |  |  |

Table 4. Table showing the summary of results for one task.

* + 1. Satisfaction Results

SUS [Brooke, J., 1996] is recommended to be used as a satisfaction questionnaire because it is well-known, straightforward and publicly available.

**Satisfaction questionnaire**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Questions** | **SD1** |  |  |  | **SA2** |
| 1. I think that I would like to use this system frequently |  |  |  |  |  |
| 2. I found the system unnecessarily complex |  |  |  |  |  |
| 3. I thought the system was easy to use |  |  |  |  |  |
| 4. I think that I would need the support of a technical person to be able to use this system |  |  |  |  |  |
| 5. I found the various functions in this system were well integrated |  |  |  |  |  |
| 6. I thought there was too much inconsistency in this system |  |  |  |  |  |
| 7. I would imagine that most people would learn to use this system very quickly |  |  |  |  |  |
| 8. I found the system very cumbersome to use |  |  |  |  |  |
| 9. I felt very confident using the system |  |  |  |  |  |
| 10. I needed to learn a lot of things before I could get going with the system |  |  |  |  |  |

1SD = Strongly Disagree

2SA = Strongly Agree

Table 5. SUS questionnaire.

**Scoring SUS**

SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own. To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1,3,5,7,and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU.

SUS scores have a range of 0 to 100.

1. Appendices of this document

Custom questionnaires, Participant General Instructions and Participant Task Instructions are appropriately submitted as appendices.

* 1. Metrics for the runtime evaluation of the validator

The purpose of this section is to provide the people in charge of performing the runtime evaluation of the validators with a set of metrics that will help them to focus on the important aspects of the evaluation they have to perform.

The metrics listed in this document are organized by the quality characteristics from International Standards documents (i.e.: ISO/IEC 9126-1, ISO/IEC 9126-4). The following information is given for each metric in the table:

* Quality subcharacteristics: ISO subcharacteristic to which the metric belongs.
* Metric name.
* Purpose of the metric: This is expressed as the question to be answered by the application of the metric.
* Method of application: Provides an outline of the application.
* Measurement, formula and data element computations: Provides the measurement formula and explains the meanings of the used data elements[[1]](#footnote-1).
* Interpretation of measured value: Provides the range and preferred values.

**RUNTIME METRICS FOR VALIDATORS**

| Metric name |  | Quality subcharacteristics | Purpose of the metrics | Method of application | Measurement, formula and data element computations | Interpretation of measured value |
| --- | --- | --- | --- | --- | --- | --- |
| Task effectiveness |  | Effectiveness | What proportion of the goals of the task is achieved correctly? | User test | M1 = |1-∑A|  A = proportional value of each missing or incorrect component in the task output. | 0 <= M1 <= 1  The closer to 1.0 the better. |
| Task completion |  | Effectiveness | What proportion of the tasks is completed? | User test | X = A / B  A = number of tasks completed  B = total number of tasks attempted | 0 <= X <= 1  The closer to 1.0 the better. |
| Error frequency |  | Effectiveness | What is the frequency of errors? | User test | X = A / T  A = number of errors made by the user  T = time or number of tasks | 0 <= X  The closer to 0 the better. |
| Task time |  | Productivity | How long does it take to complete a task? | User test | X = Ta  Ta = task time | 0 <= X  The smaller the better. |
| Productivity proportion |  | Productivity | What proportion of the time is the user performing productive actions? | User test | X = Ta / Tb  Ta = productive time or task time – help time – error time – search time.  Tb = task time | 0 <= X <= 1  The closer to 1.0 the better. |
| Relative user efficiency |  | Productivity | How efficient is a user compared to an expert? | User test | X = A / B  A = ordinary user’s task efficiency  B = expert user’s task efficiency | 0 <= X <= 1  The closer to 1.0 the better. |
| Satisfaction scale |  | Satisfaction | How satisfied is the user? | User test | X = A / B  A = questionnaire producing psychometric scales  B = population average | 0 < X the larger the better. |
| Satisfaction questionnaire |  | Satisfaction | How satisfied is the user with specific software features? | User test | X = ∑(A) / n  A = response to a question  n = number of responses | Compare with previous values, or with population average. |

1. Template for the evaluation of validator at design-time
   1. Purpose and Objectives

A template is proposed for the generation of a report at design-time evaluation. This kind of evaluation is focused on the following aspects:

* **Effectiveness**: Effectiveness metrics measure the accuracy and completeness with which modelling activities supported by the UsiXML language and its associated tools can be achieved.
* **Productivity**: Measures of productivity relate the level of effectiveness achieved to the expenditure of resources.
* **Satisfaction**: Satisfaction measures the extent to which developers/designers are free from discomfort and their attitudes towards the use of the UsiXML language and tools.
* **µ7 coverage**: This metrics measure the perception that developers/designers have about the level of coverage µ7 aspects of their validator, their subjective vision about how the UsiXML language and tools have helped them to cover those aspects.
  1. Audience

The people using this template will be those in charge of performing the modelling activities related to the development of the validators at each partner. The information regarding the companies and the involved stakeholders were previously collected by two questionnaires:

* <https://docs.google.com/spreadsheet/viewform?formkey=dHZGczB4TzkyMUwta1lIblM2WDBZX0E6MA>
* <https://docs.google.com/spreadsheet/viewform?formkey=dGI0MG5JZ0ZNRDBWTXNVZ3BDSjR4ZUE6MA>.
  1. Template

Title Page

|  |  |
| --- | --- |
| Common Industry Format Usability Test  Validator name and version number  Evaluation at design-time  Report by: Evaluator name  Company name/logo  Date when the test was conducted  The date the report was prepared | |
| UsiXML_Transparent2.png | Project acronym: UsiXML  Project full title: User interface eXtensible Mark-up Language  ITEA label n° 08026 |

**Contents (mostly extracted from ISO/IEC 9126-4:2004)**

1. Introduction

1.1. Executive Summary

1.2. Full Validator Description

1.3. Test Objectives

2. Method

2.1. Participants

2.2. Context of Validator Use in the Test

2.3. Experimental Design

2.3.1 Procedure

2.3.2 Participant General Instructions

2.4. Design-time evaluation Metrics

3. Results

3.1. Treatment of data

3.2. Presentation of the Results

3.2.1 Performance Results

3.2.2 Satisfaction Results

3.2.3. µ7 coverage results

4. Appendices

1. Introduction
   1. Executive Summary

This section provides a high level overview of the test. Possible content is:

* Identity and description of the validator
* Summary of the method(s) of the evaluation including the number of and type of participants (in this template participants are the developers and designers of the validator) and their tasks
* Results expressed as mean scores or other suitable measure of central tendency
  1. Full Validator Description

This section identifies the formal validator name and release or version.

* 1. Test Objectives

This section describes all of the objectives for the test and any areas of specific interest. Possible objectives are effectiveness, productivity, satisfaction and µ7 coverage.

1. Method

This section must provide sufficient information to allow an independent tester to replicate the procedure used in testing.

* 1. Participants

This section describes the users who participated in the development of the validator in terms of demographics, professional experience, computing experience and special needs.

A general description should include important facts such as:

* The total number of participants tested.
* Segmentation of user groups tested. Example: novice and expert programmers.
* The key characteristics and capabilities expected of the user groups being evaluated.
* How participants were selected and whether they had the essential characteristics and capabilities.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Gender** | **Age** | **Education** | **Occupation/ role** | **Professional experience** | **UIDLs1 experience** | **UIDL tools experience** |
| **P1** |  |  |  |  |  |  |  |
| **P2** |  |  |  |  |  |  |  |
| **Pn** |  |  |  |  |  |  |  |

UIDL: User Interface Description Language.

Table 6. Table to collect information about the participants in the design-time evaluation.

* 1. Context of Validator Use in the Test

This section describes the UsiXML models and tools used during the development of the validator.

* 1. Evaluation design
     1. Procedure

This section details the evaluation protocol:

* Give operational definitions of measures and any presented independent variables or control variables.
* Specify the steps that the evaluation team followed to execute the test sessions and record data.
* State whether participants were paid or otherwise compensated.
  + 1. Participant General Instructions

Include here or in an appendix all instructions given to the participants.

* 1. Design-time evaluation Metrics

Explain what measures have been used for each category of usability metrics: effectiveness, efficiency and satisfaction. Use the generic set of metrics selected for the UsiXML project and extend it if the evaluation of the validator needs it. Additional information can be found in the section B.IV of this document.

1. Results

This is the second major technical section of the report. It includes a description of how the data were scored, reduced, and analysed. It provides the major findings in quantitative formats.

* 1. Treatment of Data

In order to treat gathered data, several activities are suggested:

* Data Scoring: The method by which the data collected were scored should be described in sufficient detail to allow replication of the data scoring methods by another organization if the test is repeated.
* Data Reduction: The methods by which the data were reduced should be described in sufficient detail to allow replication of the data reduction methods by another organization if the test is repeated.
* Statistical Analysis: The method by which the data were analysed should be described in sufficient detail to allow replication of the data analysis methods by another organization if the test is repeated.
  1. Presentation of the Results

Both tabular and graphical presentations of results should be included. The data may be accompanied by a brief explanation of the results but detailed interpretation is discouraged.

* + 1. Performance Results

It is recommended that efficiency and effectiveness results be tabulated across participants based on the different kind of specification models (domain, task, abstract presentation and context). A summary table showing total mean task times and completion rates across all modelling activities should be presented.

Task A (for instance domain modelling if the UsiXML language and tools have been used for this purpose)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **User #** | **Task effectiveness** | **Errors** | **Errors frequency** | **Specification Time (hours)** | **Extra time** |
| **1** |  |  |  |  |  |
| **2** |  |  |  |  |  |
| **N** |  |  |  |  |  |
| **Mean** |  |  |  |  |  |
| **Standard Deviation** |  |  |  |  |  |
| **Min** |  |  |  |  |  |
| **Max** |  |  |  |  |  |

Table 7. Table to be used to collect information about one modelling activity.

Summary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **User #** | **Task effectiveness** | **Errors** | **Errors frequency** | **Specification Time (hours)** | **Extra time** |
| **1** |  |  |  |  |  |
| **2** |  |  |  |  |  |
| **N** |  |  |  |  |  |
| **Mean** |  |  |  |  |  |
| **Standard Deviation** |  |  |  |  |  |
| **Min** |  |  |  |  |  |
| **Max** |  |  |  |  |  |

Table 8. Table showing the summary of results for one modelling activity.

* + 1. Satisfaction Results

SUS [Brooke, J., 1996] is recommended to be used as a satisfaction questionnaire because it is well-known, straightforward and publicly available.

**Satisfaction questionnaire**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Questions** | **SD1** |  |  |  | **SA2** |
| 1. I think that I would like to use the UsiXML language and its related tools frequently |  |  |  |  |  |
| 2. I found the UsiXML language and its related tools unnecessarily complex |  |  |  |  |  |
| 3. I thought the UsiXML language and its related tools was easy to use |  |  |  |  |  |
| 4. I think that I would need the support of a technical person to be able to use the UsiXML language and its related tools |  |  |  |  |  |
| 5. I found the UsiXML language and its related tools contributions well integrated |  |  |  |  |  |
| 6. I thought there was too much inconsistency in the UsiXML language and its related tools |  |  |  |  |  |
| 7. I would imagine that most developers and designers would learn to use the UsiXML language and its related tools very quickly |  |  |  |  |  |
| 8. I found the UsiXML language and its related tools very cumbersome to use |  |  |  |  |  |
| 9. I felt very confident using the UsiXML language and its related tools |  |  |  |  |  |
| 10. I needed to learn a lot of things before I could get going with the UsiXML language and its related tools |  |  |  |  |  |

1SD = Strongly Disagree

2SA = Strongly Agree

Table 9. SUS questionnaire adapted to evaluate UsiXML.

**Scoring SUS**

SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own. To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1,3,5,7,and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU.

SUS scores have a range of 0 to 100.

* + 1. µ7 coverage results

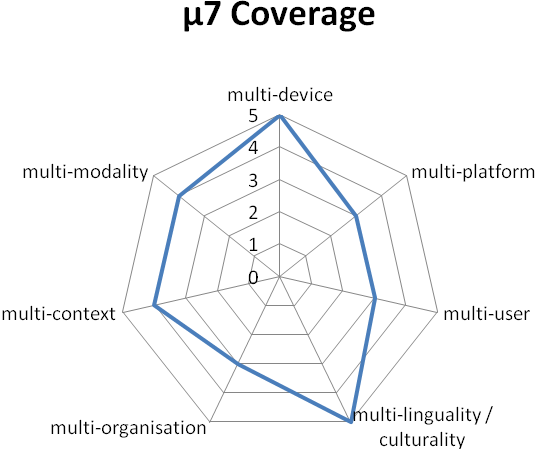
The following questionnaire is recommended to be used as a µ7 coverage questionnaire.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **µ7 coverage** | **1** | **2** | **3** | **4** | **5** |
| multi-device |  |  |  |  |  |
| multi-platform |  |  |  |  |  |
| multi-user |  |  |  |  |  |
| multi-linguality / culturality |  |  |  |  |  |
| multi-organisation |  |  |  |  |  |
| multi-context |  |  |  |  |  |
| multi-modality |  |  |  |  |  |

The 1 to 5 scale refers to the level of coverage of the µ7 aspect, being 1 the lowest and 5 the highest.

Table 10. Table used to collect data from the evaluation of the µ7 coverage.

In order to graphically show the results, a radar chart is proposed.



1. Appendices of this document

Custom questionnaires, Participant General Instructions and Participant Task Instructions are appropriately submitted as appendices

* 1. Metrics for the evaluation of the modelling activities

The purpose of this section is to provide the people in charge of performing the evaluations of the validators with a set of metrics that could help them focusing on the important aspects of the evaluation they have to perform.

The metrics listed in this document are organized by the quality characteristics from International Standards documents (i.e.: ISO/IEC 9126-1, ISO/IEC 9126-4). The following information is given for each metric in the table:

* Quality subcharacteristics: ISO subcharacteristic to which the metric belongs.
* Metric name.
* Purpose of the metric: This is expressed as the question to be answered by the application of the metric.
* Method of application: Provides an outline of the application.
* Measurement, formula and data element computations: Provides the measurement formula and explains the meanings of the used data elements[[2]](#footnote-2).
* Interpretation of measured value: Provides the range and preferred values.

QUALITY IN USE METRICS FOR the designer’s and modeller’s work

| Metric name |  | Quality subcharacteristics | Purpose of the metrics | Method of application | Measurement, formula and data element computations | Interpretation of measured value |
| --- | --- | --- | --- | --- | --- | --- |
| Task effectiveness |  | Effectiveness | What proportion of the requirements were you able to specify using a particular UsiXML model? | Test with designers.  Calculate the proportion of requirements that have been specified using a particular UsiXML model. | X = A / T  A = number of requirements specified.  T = total number of requirements. | 0 <= X <= 1  The closer to 1.0 the better. |
| Error frequency |  | Effectiveness | What is the frequency of errors in a particular model? | Test with designers. Count the number of specification errors detected. | X = A  A = frequency of specification errors (seldom, regularly, very frequent). | The less frequent the better. |
| Specification time |  | Productivity | How long does it take to complete a particular specification model? | Test with designers. Measure the time that the designer takes to complete a particular specification model. | X = Ta  Ta = specification time | 0 <= X  The smaller the better. |
| Relative user efficiency1 |  | Productivity | How efficient is a new designer compared to an expert? | Test with designers. Compare the task time of each designer to the ideal time that a UsiXML expert would take. | X = A / B  A = ordinary user’s task time  B = expert user’s task time | 0 <= X <= 1  The closer to 1.0 the better. |
| Extra time |  | Productivity | How much work is needed to get the validator working (adding functionality that could not be modelled)? | Test with designers. Measure the time needed to complete the implementation of the validator functionality that was not modelled because it was not possible or because errors of the designer. | X = A / B  A = time needed to add functionality that could not be modelled  B = total time | 0 <= X <= 1  The closer to 0.0 the better. |
| Satisfaction questionnaire |  | Satisfaction | How satisfied is the user with UsiXML? | Test with designers. Prepare a questionnaire using the System Usability Scale (SUS). | X = A / B  A = questionnaire producing psychometric scales  B = population average | 0 < X  The larger the better. |
| Discretionary usage |  | Satisfaction | Will the designers use UsiXML in the future? | Questionnaire or interview. | X = A/ B  A = positive responses  B = total amount of people questioned | 0 <= X <= 1  The closer to 1.0 the better. |

1Relative user efficiency metric can be derived based on the experience of the participants (see section 2.1 of this appendix) and the information gathered.

1. In some situations more than one formula is proposed for a metric. [↑](#footnote-ref-1)
2. In some situations more than one formula is proposed for a metric. [↑](#footnote-ref-2)