



D 7.4

THE UX-TOOL: MEASUREMENT DEVICE AND DOCUMENTATION FOR EMPIRICAL ASSESSMENT OF THE USER EXPERIENCE IN INTERACTIVE STORYTELLING

Project Number	FP7-ICT-231824
Project Title	Integrating Research in Interactive Storytelling (NoE)
Deliverable Number	D7.4
Title of Deliverable	The UX-tool: Measurement device and documentation for empirical assessment of the user experience of Interactive Storytelling
Workpackage No. and Title	WP7 - User-centered Evaluation of IS Systems
Workpackage Leader	VUA
Deliverable Nature	Report
Dissemination Level	Public
Status	Finished
Contractual Delivery Date	31 th December 2011
Actual Delivery Date	13 th January 2012
Author(s) / Contributor(s)	Christoph Klimmt, Christian Roth, Ivar Vermeulen, Peter Vorderer
Number of Pages	34

Table of Contents

ABSTRACT	1
1. INTRODUCTION	2
2. UX-TOOL MANUAL: HOW TO OPERATE THE USER EXPERIENCE MEASUREMENT TOOLBOX	10
2.1 HOW TO SET UP A NEW USER ACCOUNT? ACCOUNT CREATION	10
2.2 HOW TO SET UP A NEW STUDY? THE SURVEY WIZARD	10
2.2.1 <i>Study name and description</i>	11
2.2.2 <i>Study design</i>	11
2.2.3 <i>Measurement modules</i>	11
2.2.4 <i>Part A: Preconditions of meaningful user experience</i>	12
2.2.5 <i>Part B: User Experience</i>	13
2.2.6 <i>Part C: System specific user reactions</i>	14
2.2.7 <i>Part D: User information</i>	15
2.2.8 <i>Simple explorative study</i>	17
2.2.9 <i>Group comparison study</i>	17
2.3 ANALYSIS	19
3. REFERENCES	22
4. APPENDIX: SCALES AND ITEMS OF THE UX-TOOL	25
4.1 DIMENSIONS FOR USER EXPERIENCE EVALUATION	25
4.1.1 <i>Part A: Preconditions of meaningful user experience</i>	25
4.1.2 <i>Part B: User Experience - Common and frequent experiential qualities</i>	28
4.1.3 <i>Part C: System specific user reactions</i>	31
4.1.4 <i>Part D: User information / Demographics</i>	32



Abstract

This manual introduces an electronic measurement tool, called UX-tool, which supports easy and rapid user research during or after the development of an interactive storytelling application. With this tool, researchers and developers involved in interactive storytelling can for the first time implement a social-scientific approach to user testing with measurement scales that have been carefully developed and pre-tested with several reference systems. It is designed to support social science non-specialists to easily implement reliable and valid studies by providing an online survey system with a set of over 14 tested user experience scales. Researchers can choose the scales that are suitable for their studies. The complete system is located online at the domain <http://www.ux-tool.com>.

Keywords: User experience measurement, Interactive Storytelling, UX-tool

CITATION Information

The IRIS Consortium (2012). UX tool: Measurement device and documentation for user research in interactive storytelling.

Online: → <http://iris.scm.tees.ac.uk/publications> D-7.4



1. Introduction

This manual introduces an electronic measurement tool that supports easy and rapid user research during or after the development of an interactive storytelling application. Interactive storytelling is receiving increasing attention and interest by computer science communities, and the evolution of technologies as well as design paradigms is going on at a remarkable pace. While many research teams are exploring new technologies for connecting pre-authored narrative with interactive and dynamic story generation, research on end-users is comparatively sparse. A few studies have been published that have confronted (mostly only a few) users with a prototype of an interactive storytelling system and then asked about their experience and evaluation. While such pilot studies can be very helpful in moving forward with a specific story design project, the rationale behind the present UX tool is different.

With this tool, researchers and developers involved in interactive storytelling can for the first time implement a social-scientific approach to user testing with measurement scales that have been carefully developed and pre-tested with several reference systems. Based on theoretical work in media entertainment and expert consultations (Klimmt et al., in press), the UX tool comes with ready-made question sets (“scales”) on a variety of user experiences that can be – or are likely to be – relevant to a given prototype of interactive digital storytelling. Research teams may select from the list of possibly relevant dimensions, add new questions and scales according to their individual requirements, and then run a testing questionnaire that fits elegantly to specific project goals and objectives. Based on an online platform, the UX tool thus enables user research at various stages of system development (from early prototyping to the stage of finishing a beta-version for public release). It reduces the preparation efforts and monetary costs of including elements of user testing in the research and design process, which allows keeping the user perspective in mind even in technologically complex and challenging stages of system development.

The UX tool helps to answer several fundamental questions on user responses to an interactive storytelling system in an empirical, social-scientific way:

- What is the typical **experience profile** that my interactive storytelling system evokes in users? Are there aspects of user experience (e.g., suspense) that are particularly salient in my users?
- To which extent does the average user experience mirror the kind of experience that I **expected** or envisioned when planning the system? Does the system achieve the kind and intensity of user experience that it is intended to achieve?
- How does the experience profile evoked by my interactive storytelling system look like **compared to well-known reference systems** of interactive storytelling?



- In which way and to which extent do **different design versions** of my interactive storytelling system **evoke specific experience** profiles? Does a specific design decision in my interactive story make a difference for the average user experience?
- In which way and to which extent do **different target audiences** go through specific average experiences when using my interactive storytelling system?

To determine the kind and intensity of the average user experience, the UX tool includes concept-based, ready-made question sets (“scales”) for fourteen facets of user responses, all of which have been linked theoretically and/or empirically to interactive storytelling:

- **Usability of system:** Usability is the key question of whether users understand their action possibilities and are capable to translate their action intentions (e.g., interaction with a character, manipulation of a virtual object) into story-relevant behavior effectively. To which extent do they succeed in doing what they want to do? (e.g., Nielsen, 1993).
- **Correspondence with system performance and own expectations:** This dimension refers to whether users find their technical interaction possibilities being violated or fulfilled by the interactive storytelling system. Because no system is perfect, it is important to know if users are ready to accept the technological boundaries and limitations of the given system, whether they had expected the technology to perform better (e.g., in terms of dialogue naturalness) or whether the system actually outperformed their expectations. Disappointment with the underlying technology would be a fundamental threat to any meaningful user experience, and surprise about what a system can do relates to novelty effects, overall enjoyment, and maybe distraction from the intended type of user experience (as people may begin to play around with advanced technology instead of driving the plot forward).
- **Presence:** The concept of Presence has been advanced mostly in the field of virtual reality rather than interactive storytelling (with a few exceptions, e.g., Dow et al., 2007). However, some emerging systems and prototypes of interactive virtual storytelling may render both Spatial Presence in virtual environments (Wirth et al., 2007) and Social Presence with virtual agents essential dimensions of the intended user experience. Feeling Presence implies insusceptibility to external distraction and high levels of engagement with the story world or virtual space.



- **Character believability:** Just as many conventional narratives, interactive storytelling frequently involves characters; many of them serve as interaction partners for users. Their autonomous behavior and the social-emotional responses that they evoke are therefore important determinants of whether the overall user experience can unfold in the way designers have intended it to do.
- **Effectance:** The impact of users' decisions and actions onto the story events must be made salient. In video games research, the term effectance has been used to describe this basic experiential quality (Klimmt et al., 2007). To the extent that users experience a 'sufficient' degree of effectance, experiential qualities that build on salient impact, such as curiosity ("what will happen if I do this to the character?"), are enabled. Therefore, effectance is considered an important precondition of meaningful user experiences; it is mostly connected to technical features such as processing speed and explicitness of feedback given to user inputs (Klimmt et al., 2007).
- **Autonomy:** Given the diverse options to make an interactive story interactive, users may have different perceptions of how much freedom they have in interacting. For instance, "scripted" interaction may limit user possibilities to a minimum (in order to preserve a coherent pre-authored storyline), whereas other design approaches may maximize users' degrees of freedom what to do and how to do it. Because autonomy is often linked to positive overall experience (Ryan, Rigby & Przybylski, 2006), system designers may want to know which level of autonomy users think has been given to them.
- **Curiosity:** Stimulating users' curiosity is most often a positive contribution to overall system appreciation and to sustained exposure motivation. Users can be 'made' curious about the plot development, about specific characters, about the capabilities of the technology and interface, and also about the consequences of their (playful versus serious) decisions concerning the interactive story.
- **Suspense:** Mild or intense forms of suspense are common among consumers of many different types of narratives (Vorderer et al., 1996). Because users of interactive storytelling systems are typically engaged, for instance, by holding some responsibility for the outcome of a conflict, psychological conditions for surprise are frequently met in such environments (Klimmt et al., 2009).
- **Flow:** Clearly, flow (Csikszentmihalyi, 1990) belongs to the key types of user experience of many interactive storytelling applications. Such applications



generate flow experience with the ‘right’ pacing of interaction loops and with stimulating users’ cognitive activity (e.g., thinking about how to resolve problems and to push the plot forward, similar to adventure video games). Flow directly builds on usability (see above), as a continuous stream of user action can only occur if users are comfortable with the devices they apply for interaction with the system.

- **Aesthetic Pleasantness:** Interactive narratives frequently cite or build on pieces of literature with great aesthetic values (e.g., Cavazza et al., 2007). Furthermore, the confluence of computer-generated imagery, interacting characters, and dynamic plot development is likely to result in aesthetically appealing user experiences that hold artistic value. While aesthetic pleasantness can have many specific sources (e.g., Figure s, story elements), its assessment is considered as a summative measure of the overall aesthetic appreciation users feel towards a given system (Rowold, 2008).
- **Enjoyment:** Aesthetic experiences are not necessarily comfortable in nature. Enjoyment – the key concept in entertainment research – clearly is focused on pleasant experiences, although negative emotions such as sadness can be entertaining on a meta-level of emotion (Oliver & Bartsch, 2010). The dimension of enjoyment is a summative experiential facet of positive, pleasant affective state with various manifestations (Vorderer et al., 2004). In interactive storytelling, enjoyment may be the synthesis of curiosity, surprise, and several other affective responses that occur jointly or sequentially during exposure to a system. In this sense, enjoyment is an aggregate experience whose measurement is useful in predicting the overall motivational appeal of a given interactive storytelling application; it also allows to contextualize user research on interactive storytelling with comparisons to the entertaining appeal of other media such as video games or movies.
- **Emotional state:** Different narratives can evoke diverse affective responses in users. They can range from very unpleasant to very pleasant, from very calm to extremely excited. Similar to the Positive and Negative Affect Schedule (PANAS; Watson, Clark & Tellegen, 1988), the UX tool includes items to measure positive and negative affect as general emotional categories.
- **Role adoption / identification:** In interactive storytelling, users typically are offered a role within the narrative setting, and it should be of great interest to designers whether users adopt this role (identify with the role) or keep some distance (‘remain outside of the story’, cf. Oatley, 1994). Customization of



this measure is necessary, however, because the role offered to users may vary greatly. Sometimes, this role may be made explicit (the user is assigned to “be” a hero, for instance), at other times the role may just be conveyed implicitly to users so that merely some properties of the character that users control become salient.

For these concepts of user experience, the UX tool includes ready-made, pretested and validated scales (see appendix for an overview of the dimensions/scales). Their length (number of questions per scale) is limited so that study participants will not be overwhelmed with a huge questionnaire. With a Figure al user interface, researchers can generate a customized, online-based questionnaire that either contains all these scales or is construed by selecting only those among the mentioned scales that fit to the given particular research interest.

The UX tool also enables customization of the user questionnaire by adding new questions. For specific interactive storytelling prototypes, system designers and researchers may be interested in special kinds of user responses that are not well-covered by the ‘mainstream’ dimensions pre-built in the UX tool. For example, testing an interactive comedy with end-users may require to add a measure for exhilaration. Adding rating items on humor and exhilaration (either invented by the researchers or extracted from the literature) is just as simple as selecting ready-made scales from the UX tool repository. This way, the tool combines the advantages of pre-fabricated, pre-tested scales with the option to add new research components. Of course, the tool also allows collecting background information on study participants such as demographics.

The logics of measurement for the user experience scales are based on conventional experimental-psychological research techniques. In order to assess the intensity of a given experience in a given user, the user is requested to rate her/his agreement to a statement (“item”) that manifests the essence of one particular dimension of user experience. For instance, the item “I found myself staring at the screen in anticipation” describes certain features of the experiential quality of “suspense”. Users participating in a study testing an interactive storytelling system are asked to rate the extent to which this statement matches their personal experience during exposure to the interactive storytelling system. Lower scores on the rating scale that ranges from 1 to 5 mark less agreement and thus a lower intensity of the described kind of experience; higher scores, in contrast, reflect higher intensities of the experience. To avoid dependence on one particular verbalization of a user experience, scales are used that combine several statements (items) with diverse formulations of the same kind of experience. Pre-test studies conducted within the IRIS project demonstrated that the items included in the UX tool converge statistically so that there is a solid empirical foundation for taking each scale as a functioning measure of a given user experience. Agreement scores of an individual person for all items that belong to the measure for one user experience (e.g., suspense) are averaged to obtain a robust metric value. These scale (or index) values can then be averaged across study



participants, for example, to compare the average suspense of men and women when confronted with a specific interactive narrative.

Example Study (from Endrass et al., 2011)

User responses were explored to two different dialog-based interaction modes in a soap opera-like interactive narrative. An interactive demonstrator was evaluated using UX tool scales in two different versions: One version of the “virtual beergarden” story provided a continuous interaction-style so that human users could enter verbal contributions to ongoing conversations between characters any time (including interruptions of non-player characters), and a second system version provided system-initiated interaction where the user can only make verbal statements at certain prompts (i.e., when non-player characters have finished their contributions and can ‘listen’ to the user character). The goal of the study was to find out whether this difference in designing user interactivity, which comes with fundamental implications for system technology and programming efforts, would make any relevant – desirable or undesirable – difference for the user experience. If users would find, for instance, one paradigm of interactivity more usable, more enjoyable, and/or more positive in terms of character believability, such differences could be considered in developing the “beergarden” prototype interactive storytelling further towards a more appealing full system.

A total of 42 university students participated in the study. They were randomly assigned to either use the interactive “beergarden” story with the continuous dialogue feature or with the system-initiated dialogue feature. Afterwards, they completed a questionnaire that included UX tool scales on user experience plus several customized questions that were thought to be informative in the present context. Average scores on the scales per user dimension and average rating responses to the additional questions were compared between those users with continuous dialogue and users with system-initiated (round-based) dialogue to determine possible discrepancies in user reactions:



Dimension of user experience	Users confronted with continuous dialog		Users confronted with round-based dialog		t-test
	M	SD	M	SD	p
Autonomy	2.68	0.93	2.17	0.79	0.055
Curiosity	3.78	0.77	3.33	0.92	0.095
Suspense	3.16	0.74	2.82	0.79	n.s
Enjoyment	3.45	0.95	3.07	0.94	n.s
Comparisons with expectations					
"I had the expectation that the experience would be more engaging"	1.75	0.97	2.18	0.59	0.09
Similarities to other experiences					
"The experience reminded me of playing a video game"	4.25	1.33	4.82	0.39	0.06
"The experience reminded me of watching a movie"	2.45	0.99	1.77	1.15	0.049
"The experience reminded me a little of playing improvisation theater"	2.10	1.33	1.36	0.49	0.02

Comparing the scores between the two user groups provides a set of interesting findings: Varying the style of text-based dialogue interaction has substantial impact on the user experience of the overall interactive narrative, for instance, with regard to perceived autonomy, evocation of curiosity or an approximation of an experience of improvisation theatre (which both versions of the interactive storytelling system did not mirror very well, but which is resembled by the continuous dialogue version much better than by the round-based dialogue version). Thus, findings from a relatively small user experiment that employed the scales of the UX tool can produce a variety of helpful insights the system creators and researchers can exploit in different ways.

The UX tool is the product of joint research and development conducted by various partners of the Network of Excellence "Integrating Research in Interactive Storytelling" (IRIS: FP7-ICT-231824). This collaboration among various academic teams with advanced expertise in interactive storytelling allowed to generate a solid conceptual foundation for the tool as well as pre-testing the tool questionnaire with various systems of interactive storytelling. From these pretesting and validation studies, the UX tool has been equipped with reliability data (indicating the robustness of the contained ready-made scales) and reference values produced by UX tool scales with existing interactive narratives. For example, if a research team is working on a prototype of interactive storytelling that is similar in certain aspects to the "Façade" system (www.interactivestory.net), the documentation materials provided by the IRIS consortium would allow to compare UX scale values produced in a user study on this new system with what IRIS researchers have measured in studies employing the "Façade" system. Therefore, direct comparisons of UX scale results with reference data from other systems – either systems that are similar or systems that one may find



systematically different from a current prototype – are enabled, which can contribute to increase the benefit of scientific collaboration and comparative approaches in system design and development. This manual includes a list of materials and resources through which reference values as well as additional information is freely available.

Finally, this conceptual introduction to the UX tool offers a message of modesty. Research teams employing the UX tool will in many cases greatly benefit from acquiring user data, and the tool reduces the costs and efforts for such user testing significantly. However, experimental user research based on averaged scale scores is only one important element in assessing the characteristics and performance of interactive entertainment systems. Most importantly, qualitative approaches of user studies are likely to be helpful in understanding user reactions as well. With open, structured interviews, users can contribute to idea generation and point out important experiential features that would not be detected through a closed-questions-based measurement system such as the UX tool. At a more abstract level, findings from user studies alone will not tell the whole truth about the strengths and weaknesses of an entertainment system – expert reviews and analyses of cost-efficiency, for example, are of course equally relevant. One promising strategy to exploit the power of the UX tool could be, however, to link the post-exposure questionnaire run through the tool with process-based measures, such as server protocols that generate data on user behavior within the storyworld. For example, interesting findings may emerge when the level of user activity (e.g., number of words entered into the dialogue system, or distance covered by moving through the virtual world) is correlated with UX tool scores to see how specific use of interactivity features contributes to individual experiences. This way, the research horizon that can be addressed with the UX tool may be expanded dramatically. However, such study designs require advanced research technology and sophisticated techniques of data analysis. The primary goal of the UX tool is, for the time being, to support research teams with less complex (but equally informative) user studies.



2. UX-Tool Manual: How to operate the User Experience Measurement Toolbox

The User Experience Measurement Toolbox, short UX-tool, is a tailored online survey tool aimed at user experience research in Interactive Digital Storytelling. It is designed to support social science non-specialists to easily implement reliable and valid studies by providing an online survey system with a set of over 14 tested user experience scales. Researchers can choose the scales that are suitable for their studies. The complete system is located online at the domain <http://www.ux-tool.com>

2.1 How to set up a new user account? Account creation

Open your web browser, e.g. Internet Explorer, Firefox, Chrome or Safari, latest versions recommended, with the domain <http://www.ux-tool.com> and click the “Create account” button to start the user account creation process. You are asked to enter your name, affiliation and a valid e-mail address, as well as a password. A confirmation e-mail will be sent to you and you are ready to upload a profile picture, edit your account information and start working with the toolbox.

2.2 How to set up a new study? The Survey Wizard

After logging in with the user account, the main screen is shown that gives an overview over all studies that are created with this account. New users will see an empty list as there are no studies created yet. The survey wizard will help to set up a new study. It is started with a mouse click on the “new survey wizard” button and will proceed to a new screen.

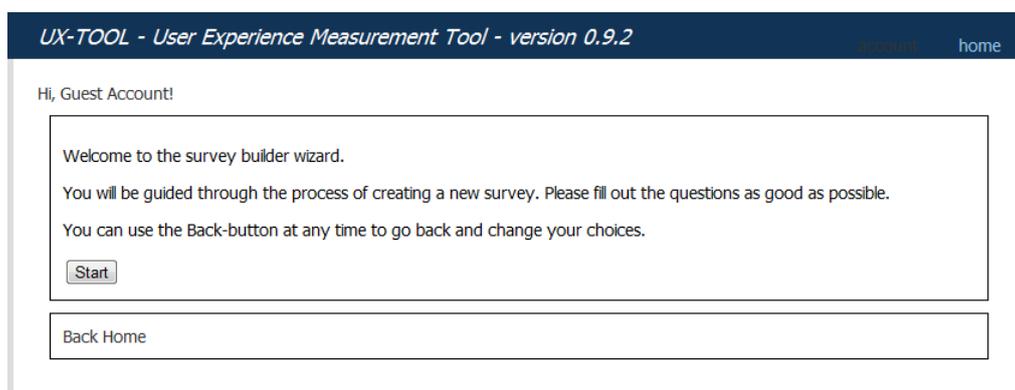


Figure 1: Welcome screen of the survey wizard



2.2.1 Study name and description

First, the new study has to be named. The name should help you to identify the study. Your newest study will always be on top of the list. Nevertheless, when plenty of new studies are created a good name helps you to identify your study within longer lists.

It is useful to name it after the application (plus version number) you are testing and the main research purpose of the study. E.g. “First usability study of IS prototype v 0.6” or “Gender differences for Storybook application”.

A second page will allow for a more detailed study description. Here the very purpose of the study should be briefly described: Information regarding the application that is about to be tested, the main research questions, and planned comparisons for the analysis part.

The next part of the survey wizard deals with the study design.

2.2.2 Study design

Using the UX-tool offers many possibilities to conduct user experience research. It is very important to know exactly what the study should be about since that determines the design of the study which in return determines what is possible to learn from the analysis.

2.2.3 Measurement modules

The user experience measurement modules are organized in four parts:

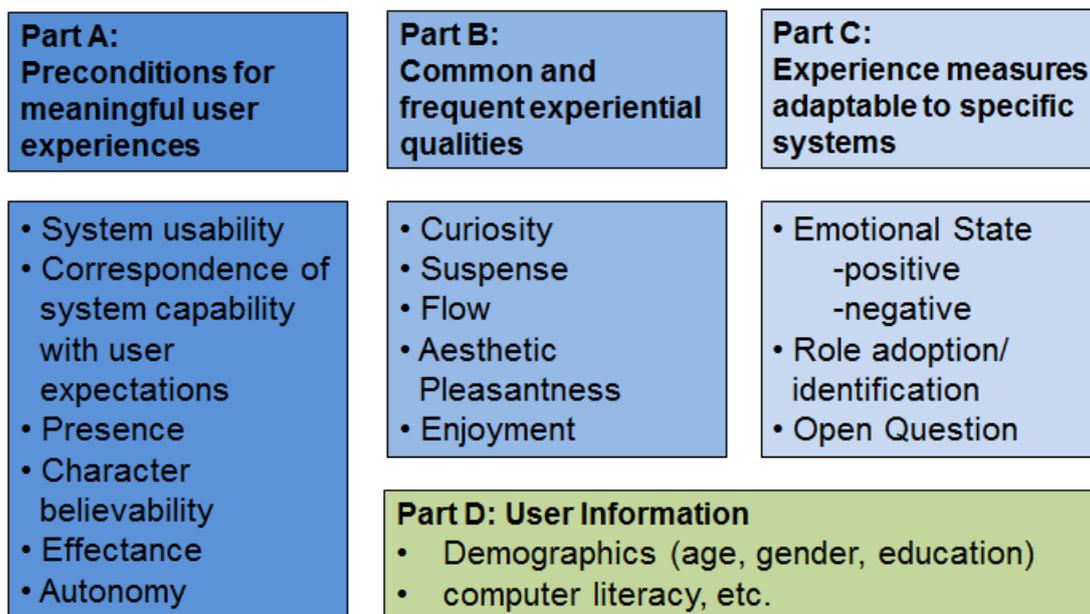


Figure 2: Dimensional architecture of the UX-tool divided in four parts

For each of these parts a separate screen allows to select modules for the questionnaire.



2.2.4 Part A: Preconditions of meaningful user experience

The fundamental part of the dimensional architecture of the evaluation toolkit relates to those aspects of the system that are necessary to enable meaningful experiences for users. These preconditions relate to issues of general system usability, but hold elements that are specific to interactive storytelling as well. The first part of the evaluation instrument includes measures for the following preconditions for user experiences:

System usability

The experience that the interaction with the story is fluent, smooth, and error-free

Correspondence of system capabilities with user expectations

The experience that the system makes realistic offers about how users can influence the story and then keeps its promise so that frustration is avoided

Presence

Users establish a sense of „being in the story world“

Character believability

Virtual agents in the story world contribute to a coherent story experience and do not damage users' illusion

Effectance

Users can easily recognize when and how they have causally affected the story or story world

Autonomy

Users have the perception that can freely choose how to interact with and within the story world

The screenshot shows a web interface for 'UX-TOOL - User Experience Measurement Tool - version 0.9.2'. At the top right, there are links for 'account' and 'home'. The main content area displays a greeting 'Hi, Guest Account!' followed by a survey instruction: 'Indicate which of the following pre-condition Likert scales should be present in the survey:'. Below this instruction is a list of five pre-conditions, each with an unchecked checkbox: 'System Usability', 'Character Believability', 'Effectance', 'Autonomy', and 'Presence'. At the bottom of the list are two buttons: 'Back' and 'Next'. Below the main content area is a separate box containing a 'Back Home' button.

Figure 3: Modules of part A



2.2.5 Part B: User Experience

The second part of the measurement toolbox for user experience evaluation refers to specified experiential qualities that are expectably typical or common design goals across many different manifestations of interactive storytelling. Their typicality stems mostly from the fact that designers will want to motivate users to begin exposure to their product (appeal) and to stay with the experience as long as it intentionally takes or is over by author decision (sustained motivation to use). This general principle of attractive system design qualifies curiosity, suspense, surprise, and flow as common user experiences for which many system designers are likely to have a need for measurement. Two more abstract experiential qualities that build on several (if not multiple) system features and user experiences – aesthetic pleasantness and enjoyment – are added here to generate a more complete account of ‘the typical user response’ to interactive storytelling.

Curiosity

Users maintain interest in what will happen next and how they could affect the story

Suspense

Users develop hopes and expectations about the story progress, but also face uncertainty about that progress

Flow

Users become absorbed in ongoing, continuous interaction with the story world

Aesthetic Pleasantness

Positive experiences of beauty or artistic impressiveness

Enjoyment

An overall sense of positively valenced experiential quality

The screenshot shows a web application interface for 'UX-TOOL - User Experience Measurement Tool - version 0.9.2'. The page has a dark blue header with 'UX-TOOL - User Experience Measurement Tool - version 0.9.2' on the left and 'account home' on the right. Below the header, the main content area is white and contains the following text and elements:

- Greeting: "Hi, Guest Account!"
- Instruction: "Indicate which of the following experiential qualities Likert scales should be present in the survey:"
- List of qualities with checkboxes:
 - Curiosity
 - Flow
 - Suspense
 - Aesthetic Pleasantness
 - Enjoyment
- Navigation buttons: "Back" and "Next" (both in small grey boxes)
- Footer button: "Back Home" (in a larger white box with a grey border)

Figure 4: Modules of part B



2.2.6 Part C: System specific user reactions

The third group of measures of the evaluation toolkit reflects the call for flexibility and customization. Specific manifestations of interactive storytelling can strive for very narrowly defined affective states of users. Consequently, three measurement elements are offered that allow to fine-tune the overall content of the evaluation toolbox for individual purposes. One of these elements addresses the affective state of the user; the second is about the adoption of the role within the story that users are assigned or (can) take over during exposure. The third element, finally, is an open question tool for detailed assessment of individual user comments.

Affect / Emotional state: positive / negative

Emotional reactions to IS application

Role adoption

Degree of identification with the story's protagonist

Open Questions

E.g. “What did you like about the story?”, “Would you recommend the application to a friend?”

The screenshot shows a web interface for the 'UX-TOOL - User Experience Measurement Tool - version 0.9.2'. At the top right, there are links for 'account' and 'home'. Below the header, it says 'Hi, Guest Account!'. The main content area is a form with the instruction: 'Indicate which of the following specific Likert scales should be present in the survey:'. There are four checkboxes: 'Affect', 'Role Adoption/Identification', 'User Satisfaction', and 'Comparison'. Below these are 'Back' and 'Next' buttons. At the bottom of the form is a 'Back Home' button.

Figure 5: Modules of part C



2.2.7 Part D: User information

The fourth part contains modules to gather information about the participants of the study, especially demographics that allow for group comparisons, e.g. to learn about differences between the male and female user experience of a tested system.

Demographics

Age, Gender, Education

Computer Literacy

Favorites (e.g. media and genres)

Open Questions

UX-TOOL - User Experience Measurement Tool - version 0.9.2 [account](#) [home](#)

Hi, Guest Account!

Indicate which of the following demographic questions should be present in the survey:

- Age
- Gender
- Education (primary, highschool/college, higher)
- Education (dutch)

Figure 6: Modules of part D

After all modules have been selected the questionnaire is ready to be created. It is possible to make changes after the compilation of the survey.

UX-TOOL - User Experience Measurement Tool - version 0.9.2 [account](#) [home](#)

Hi, Guest Account!

That concludes the survey builder wizard. Click the Create Survey -button to build the survey, or the Back-button if you want to go back and change something.

Figure 7: Screenshot of the final survey wizard message



When the survey is created it will be shown as in Figure8. It can be activated immediately to start user experience testing. However, it is advised to revise the setup first. Name and description can be edited as well as modules. The order of modules can be changed via drag and drop, just keep the left mouse button pressed to move a module around. Missing modules can still be added. When you are satisfied with the settings for your questionnaire you are ready to activate it.

Figure 8: Screenshot of a questionnaire with name, description and selected modules

On the main screen you will find your list of questionnaires with your recently created survey on top. Hover your mouse button over it, a set of options will appear to the right. Pointing at a symbol reveals a tooltip with the symbol description. Press the activate symbol and then start the survey to obtain a link you can save on the desktop of the test computers in a lab study or distribute in case of an online study. Now, data gathering can start but you might want to do a test run yourself. You can stop a survey at any time with the according stop button to change settings or to end a study. The reset button allows deleting the data from the test run to have a clean start for the real study.

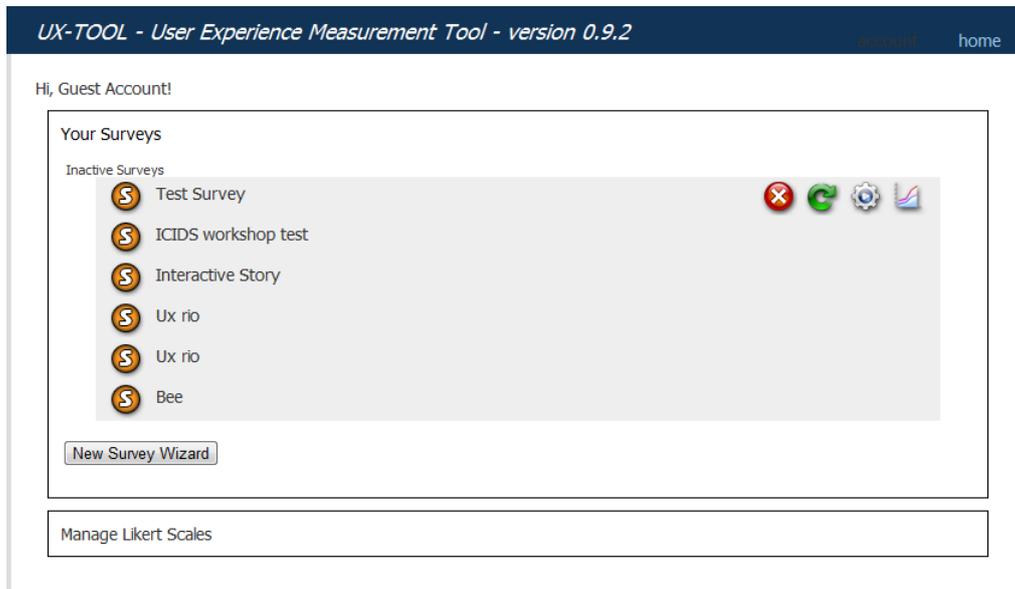


Figure 9: Main screen with study overview and action symbols (on the right)

2.2.8 Simple explorative study

A simple study design is explorative and focuses on a group of participants testing or using an interactive storytelling application. Its purpose is to explore different user experience modalities. Just choose the experience dimensions from the module overview that are interesting to your research and compile the questionnaire.

There is no need to define groups for comparison at this stage. It is possible, however, to compare the user experience of subgroups after gathering the data in the analysis part. For that purpose it is mandatory to choose modules that allow for later subgroup comparisons, such as demographics modules like gender from the user information part. Following the gender example, see the group comparison in Figure 11: group A would consist of all males and group B would be the female sub sample of the participant sample. For explorative studies, we advise to gather as much data from as many participants as possible. One downside of this approach is that the survey will become rather long and tedious to fill out by the participants and testing large amounts of users costs time and money. Online questionnaires should not take longer than a couple of minutes, better not exceeding 10 minutes. Questionnaires used as part of a lab study can be longer. Take always into account that the motivation of participants to fill out the survey will reduce over time. The demographics module should be part of every study though. Usually, this module is placed at the end of a study. For group comparisons please read the next chapter.

2.2.9 Group comparison study

If your study aims at comparing two different versions of a storytelling prototype, two different applications or two different groups using one application (e.g. males and females, high vs. low computer literacy) you will need to have enough participants for every comparison. 20 participants per group are a minimum requirement for



significance tests (t-test) of group differences.

Before setting up a study it is helpful to decide for a within subjects or between subjects study design.

A typical within study design is shown in Figure10. The same group tests e.g. two different applications or versions of an application after another. After each exposure (T1 and T2) participants will be asked to fill out the survey created with the UX-tool.

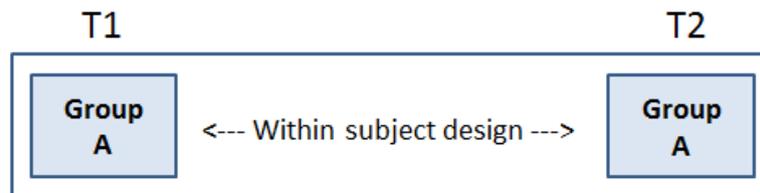


Figure 10: Comparison of two different applications within one group

Whereas in a between subjects study there are two groups of participants which test one application or application version each (see Figure11). Hence, contrary to a within subjects study every participant tests only one application. This has the advantage that lab studies are shorter since there is only one exposure (T1) necessary; however this comes at the cost of the information what the same participants experience with different applications.

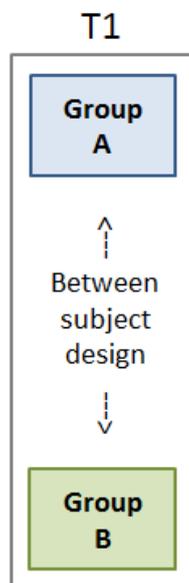


Figure 11: Comparison of the user experience between two different groups regarding the same application

A solution to this lies within even more complex study designs as shown in Figure12 which combines the two study designs and therefore allows for a variety of analysis computations.

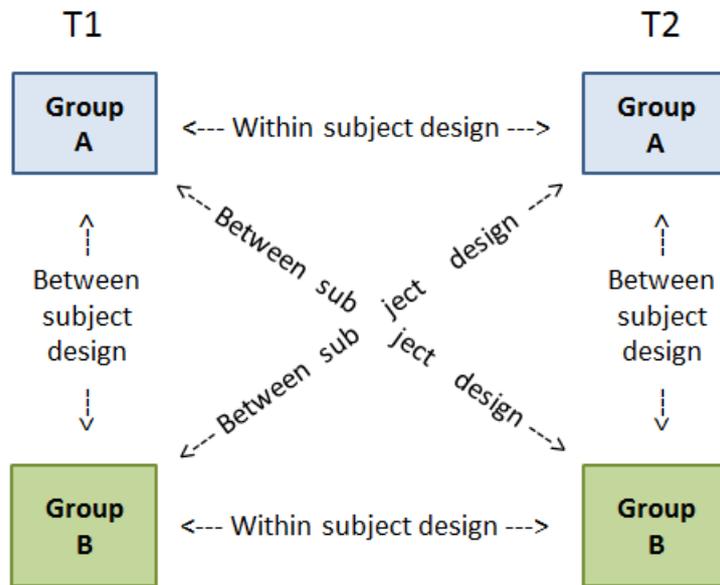


Figure 12: Complex study design with two groups and two exposures

If you want to test more than one application with different groups or subgroups it is mandatory to define which participant belongs to which group / exposure. To distinguish male and female participants just select the premade gender module. For your own group attributions the UX-tool enables you to create your own modules. Select new module in the module overview of your questionnaire and create a multiple choice question. This question will ask participants what group they belong to. In a lab study the study supervisor can advise participants how to answer this question; the names of the possible answers can support the right group selection. Instead of just naming them group A and group B you can name them after the application or version code you are testing.

2.3 Analysis

When a study is finished and stopped in the main window you can click on the analysis symbol to get descriptive information on the scales and items of the modules you used. Histograms show group differences in a convenient way. Just select the criteria to differ groups (e.g. group differentiation or gender module) to get information on how participants of these groups experienced the tested application. Future versions of the UX-tool will extend the functions of the build in analysis part.

To compute statistics and significance tests please save the study data (.csv file) and import it with your favorite analysis tool such as Excel or SPSS (Statistical Package for the Social Sciences). A good introductory read for the use of SPSS and psychological studies is e.g. Dancy and Reidy (2004).

Future versions of the UX-tool aim at incorporating an improved statistical overview and build in significance tests for group differences.



To check group mean value differences for significance the T-test is suggested. For more complex analysis e.g. the comparison of more groups literature we suggest introductory literature as Dancey and Reidy (2004) or Kuehl (1999).

When opening the data file in SPSS aim for the tab called “Statistics” and then “Compare Means”. If you compare means in a *within subjects study* you have to select the “Paired Samples T-test”, if you compare independent groups as part of your *between subjects design*, go for “Independent Samples T-Test” as shown in Figure 13.

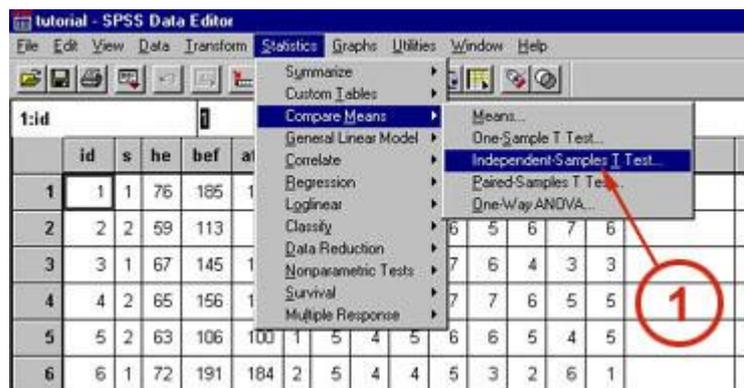


Figure 13: Selecting T-Tests to compare means in SPSS

In both cases you have to define the grouping variable. If you like to investigate group differences based on gender you'll have to select the demographic variable sex as a grouping variable and define male=1 and female=2 as it is recorded in the data set. Then the so called “Test Variables” you are interested in, have to be chosen. These are e.g. all the user experience modules you want to check for group differences. Select them from the list of continuous variables on the left and click on the arrow to send them to the "Test Variable(s)" box. Holding the CTRL-key on your keyboard while clicking allows to select several modules at once.

In the outcome file you will find the group statistics table (see 1 in figure 14) with number of cases (N) the mean values for the selected comparison dimensions (in this case the score of a final exam that is supposed to be influenced by the use of an Interactive Storytelling learning application for one group and no application for the other as comparison). The standard deviation shows how much the set of numbers is spread out around the mean value. The second table shows the actual T-test outcomes. SPSS compares the selected groups for equality of variance (how much the standard deviation spreads out). There are two possible T-tests depending on variance differences. When comparing groups their variances must be relatively similar for the first t-test to be used. The Levene's test checks for this (2). If the significance for Levene's test is higher than 0.05 the "Equal Variances Assumed" will be used. In this example this is the case since the significance shows .704. If the value is 0.05 or below the "Equal Variances Not Assumed" test (the one on the bottom) is used.



➔ T-Test

Group Statistics					
	Application vs. no Application	N	Mean	Std. Deviation	Std. Error Mean
Final Exam Score	1.00	50	74.50	15.228	2.154
	2.00	50	67.70	15.818	2.237

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Final Exam Score	Equal variances assumed	.145	.704	2.190	98	.031	6.80	3.105	.638	12.962
	Equal variances not assumed			2.190	97.859	.031	6.80	3.105	.638	12.962

Figure 14: Outcome of the T-test group comparison in SPSS

Now it is time to check the significance value of the group comparison (3).

Significance tells that the found difference is not caused by chance within a certain confidence interval. To be over 95% sure that differences of the groups are caused by group membership and not randomly or by other influences, the p-value given by the T-test should be lower than $p=0.05$. A p-value of $p=0.01$ equals the high confidence interval of 99%. Values much higher than 0.05 e.g. 2.40 are likely to be caused by other influences than the mere group distinction. One can't attribute insignificant differences to be caused by group differences, e.g. gender differences that effect the experience or exposure to two different treatments. Be aware that large samples (N over 50, especially when over hundreds) have the strong tendency to give the T-test too much "power" so it renders even small group differences significant. Also, groups shouldn't be too small. Typical rules of thumb range from 20 to 50 participants per group. In this example, the found significance value is $p=.031$ which is lower than the threshold of .05 for the 95% confidence interval. Therefore, the outcome is called significant. The Interactive Storytelling application is perceived as having an effect that caused the group differences in the final exam. To know if this observed difference is not only statistically significant but also important or meaningful, you will need to calculate its effect size. Common is Cohen's *d* but correlation Pearson's *r* is also often used as an indicator for effect size. Cohen's *d* is defined as the difference between two means divided by a standard deviation for the data. You can use online effect size calculators for this (as for instance <http://www.uccs.edu/~faculty/lbecker/>) and enter the means and standard deviations of two groups (treatment and control).

To interpret the resulting number, most social scientists use this general guide developed by Cohen (1988):

- < 0.1 = trivial effect
- 0.1 - 0.3 = small effect
- 0.3 - 0.5 = moderate effect
- > 0.5 = large difference effect



3. References

- Brooke J.: SUS - A quick and dirty usability scale. *Usability evaluation in industry*, pp. 189--94 (1996)
- Cavazza, M., Lugrin, J.-L., Pizzi, D., Charles, F.: *Madame Bovary on the Holodeck: Immersive interactive storytelling*. *ACM Multimedia 2007*, pp. 651-660 (2007)
- Chupchik, G. C., Lázló, J.: *The Landscape of Time in Literary Reception: Character Experience and Narrative Action*. *Cognition and Emotion*, 8, pp. 297--312 (1994)
- Cohen, J.: *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum. (1988)
- Dancey, C.P., Reidy, J.: *Statistics Without Maths for Psychology: Using Spss for Windows*, Prentice-Hall, Inc., Upper Saddle River, NJ (2004)
- Deci, R. M., Ryan, E. L.: *Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being*. *American Psychologist*, 55 (1), 68-78. (2000)
- Dow, S., Mehta, M., Harmon, E., MacIntyre, B., Mateas, M.: *Presence and engagement in interactive drama*. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 1475 – 1484. New York: ACM (2007)
- Endrass, B., Klimmt, C., Mehlmann, G., André, E., Roth, C.: *Exploration of user reactions to different dialog-based interaction styles*. In M. Si, D. Thue, E. André, J. Lester, J. Tannenbaum & V. Zammitto (eds.), *Interactive Storytelling - Proceedings of the 4th International Conference on Interactive Digital Storytelling (ICIDS 2011)* (Lecture Notes in Computer Science 7069, S. 243-248). Berlin: Springer. (2011)
- Hefner, D., Klimmt, C., Vorderer, P.: *Identification with the player character as determinant of video game enjoyment*. *Proceedings of ICEC 2007, 6th International Conference on Entertainment Computing*, 39-48. (2007).
- Jackson, S., Martin, A., Eklund, R.: *Long and Short Measures of Flow: The Construct Validity of the FSS-2, DFS-2, and New Brief Counterparts*. *Journal of Sport and Exercise Psychology*, 30, pp. 561--586 (2008)
- Klimmt, C., Hartmann, T., Frey, A.: *Effectance and Control as Determinants of Video Game Enjoyment*. *CyberPsychology & Behavior*, 10 (6), 845--847 (2007)
- Klimmt, C., Rizzo, A., Vorderer, P., Koch, J., Fischer, T.: *Experimental evidence for suspense as determinant of video game enjoyment*. *Cyberpsychology and Behavior*, 12 (1), 29-31 (2009)
- Klimmt, C., Roth, C., Vermeulen, I., Vorderer, P., Roth, F. S.: *Forecasting the experience of future entertainment technology: "Interactive Storytelling" and media enjoyment*. *Games and Culture: A Journal of Interactive Media*. (in press).
- Kuehl, R.O.: *Design of Experiments: Statistical Principles of Research Design and Analysis*. Duxbury Resource Center (1999)



Oatley, K. A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23, 53-74 (1994)

Oliver, M. B. & Bartsch, A.: Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human Communication Research*, 36 (1), 53-81 (2010)

Nielsen, J., *Usability Engineering* (Academic Press, Boston, MA). (1993)

Rowold J.: Instrument development for esthetic perception assessment. *Journal of Media Psychology: Theories, Methods, and Applications*, 20 (6), pp. 35--40 (2008)

Ryan, R. M., Rigby, C.S., Przybylski, A.: The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, 30, 347-363. (2006)

Spielberger, C., Jacobs, G., Crane, R., Russell, S.: Preliminary manual for the state-trait personality inventory (STPI). Unpublished manuscript, University of South Florida, Tampa (1979)

Vorderer P., Klimmt C., Ritterfeld U.: Enjoyment: At the heart of media entertainment. *Communication Theory*, 14 (4), pp. 388--408 (2004)

Vorderer, P., Wulff, H. J., Friedrichsen, M. (eds.): *Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations*. Mahwah, NJ: Lawrence Erlbaum Associates (1996)

Watson, D., Clark, L. A., Tellegen, A.: Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54, pp. 1063--1070 (1988)

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Ribeiro Gouveia, F., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T. Jäncke, P.: A Process Model of the Formation of Spatial Presence Experiences. *Media Psychology*, 9 (3), pp. 493--525 (2007)

Materials and Resources

Further background on documentation of the IRIS UX tool

- The IRIS-Website (<http://iris.scm.tees.ac.uk/publications>)
- IRIS Deliverables D7.2, D7.3 (Documentation of Scale Development and Pre-Testing ; downloadable PDFs)
- UX-tool website: <http://ux-tool.com>

Publications on UX tool theory and measurement

Roth, C., Klimmt, C., Vermeulen, I., Vorderer, P.: The experience of Interactive Storytelling: Comparing „Fahrenheit“ with „Facade“. In J. Anacleto, S. Fels, N. Graham, B. Kapralos, M. Seif El-Nasr & K. Stanley (Eds.), *Entertainment Computing – Proceedings of the 10th International Conference on Entertainment Computing*,



ICEC 2011, Vancouver (Lecture Notes in Computer Science, Vol. 6972, pp. 13-21). Berlin: Springer. (2011)

Klimmt, C., Roth, C., Vermeulen, I., Vorderer, P., Roth, F. S., 2011. Forecasting the experience of future entertainment technology: “Interactive Storytelling” and mediaenjoyment. *Games and Culture: A Journal of Interactive Media*. (in press)

Vorderer, P., Klimmt, C., Roth, C., Vermeulen, I.E.: The Audience Embedded in the Narrative: An Experiment in Interactive Storytelling. In *Proceedings of the Conference of the International Association for Media and Communication Research (IAMCR) - "Audience" Section, Istanbul, Turkey, July 17-21, 2011*. pp.657-658 (2011)

Vermeulen, I., Roth, C., Vorderer, P., Klimmt, C.: Measuring user responses to interactive stories: Towards a standardized assessment tool. In *Proceedings of the International Conference on Interactive and Digital Storytelling (ICIDS 2010)*, 01.-03. November 2010, Edinburgh, UK. LNCS, vol. 6432, Springer Verlag Berlin-Heidelberg. (2010) ISBN 978-3-642-16637-2

Klimmt, C.; Roth, C., Vermeulen, I., Vorderer, P., Roth, F. S.: Forecasting the experience of future entertainment technology: “Interactive Storytelling” and media enjoyment. Full paper presentation (poster) at the Annual Conference of the International Communication Association (ICA), Communication & Technology Division, June 22-26, Singapore (2010)

Roth, C., Vorderer, P., Klimmt, C.: The motivational appeal of interactive storytelling: Towards a dimensional model of the user experience. In: Iurgel, I., Zagalo, N., Petta, P. (eds.) *ICIDS 2009*. LNCS, vol. 5915, Springer, Heidelberg, pp. 38-43.(2009) ISBN 978-3-642-10642-2



4. Appendix: Scales and Items of the UX-tool

Based on the conceptual groundwork, a set of self-report scales was developed. For those experiential dimensions that have already been addressed in past research on video games or conventional entertainment, existing measures were reviewed and adopted, partially also adapted to the specific media context. For some dimensions, all-new scales were created. The application logics of this measurement set is that users are asked to fill in the user experience questionnaire immediately after their exposure to an Interactive Story is over.

4.1 Dimensions for user experience evaluation

Part A: Preconditions of meaningful user experience	Part B: User Experience – common and frequent qualities	Part C: System specific user reactions	Part D: User information / Demographics
System usability	Curiosity	Emotional state: positive / negative	Age
User satisfaction (system corresp. with user expectations)	Suspense	Role adoption / Identification	Gender
Presence	Flow	Comparison with other entertainment offers	Education
Character believability	Aesthetic Pleasantness	Open Questions	Computer Literacy
Effectance	Enjoyment		Open Questions
Autonomy			

Table 2 shows the architecture of the measurement dimensions in four parts

4.1.1 Part A: Preconditions of meaningful user experience

System usability

The experience that the interaction with the story is fluent, smooth, and error-free

The System Usability Scale (SUS) by Brooke (1996) measures users' usability perceptions, is generally applicable to computer-based systems, has good measurement properties, is widely cited, and also widely applied in the assessment of system usability (Lewis & Sauro, 2009). A typical item is "I thought the system was easy to use". Agreement with 10 statements is given on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree". "N" denotes a negatively framed item.



For reasons of questionnaire length, we selected three items for an abbreviated scale.

The abbreviated System Usability Scale (SUS- 3, VUA, adapted from Brooke, 1996) contains the following three items:

1. I thought the system was easy to use.
2. I would imagine that most people would learn to use this system very quickly.
3. I found the system very cumbersome to use. (N)

User satisfaction / Correspondence of system capabilities with user expectations

The experience that the system makes realistic offers about how users can influence the story and then keeps its promise so that frustration is avoided

The relationship between users' prior expectations and expectancy-(dis)confirmation is generally conceptualized as *user satisfaction*. Therefore the following scale aims at measuring user satisfaction with IS systems. The first three items measure general user satisfaction by relating prior expectations to the confirmation of these expectations. The remainder 9 items assess satisfaction with each of the dimensions underlying the user experiences with IS. Agreement with the 13 statements is given on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree". "N" denotes a negatively framed item.

User satisfaction with IS systems scale (SatIS; VUA, 2010)

1. The experience was better than I expected
2. I probably expected to much from the experience (N)
3. I was satisfied with how the system performed
4. I expected the system to be more user-friendly (usability, N)
5. I expected the experience to be more immersing (presence, N)
6. I expected the story's characters to be more believable (character believab., N)
7. I expected to have more control over the experience (effectance, N)
8. I expected the experience to be more surprising (curiosity, N)
9. I expected the experience to be more thrilling (suspense, N)
10. I expected the experience to be more engaging (flow, N)
11. I expected the story to be better (aesthetic pleasantness 1, N)
12. I expected the Figure s to be better (aesthetic pleasantness 2, N)
13. I expected the experience to be more enjoyable (enjoyment, N)

Presence

Users establish a sense of „being in the story world“

To assess perceived presence, the six item sub-dimension "Self-Location" (e.g., experience of being in the mediated environment instead of one's real-physical environment) was used from the MEC Presence questionnaire by Vorderer, Wirth, Gouveia et. al. (2004):



Presence (adapted from MEC Presence questionnaire)

1. I felt like I was a part of the environment in the presentation.
2. I felt like I was actually there in the environment of the presentation.
3. I felt like the objects in the presentation surrounded me.
4. It was as though my true location had shifted into the environment in the presentation.
5. I felt as though I was physically present in the environment of the presentation.
6. It seemed as though I actually took part in the action of the presentation.

Character believability

Virtual agents in the story world contribute to a coherent story experience and do not damage users' illusion

Based on the literature on the subject 4 dimensions underlying character believability have been identified:

- emphatic responses to characters [E]
- cognitive responses to characters [C]
- perceived affect in characters [A]
- perceived intentionality in characters [I]

A four item to match each dimension has been developed:

Character believability (VUA, 2010)

1. I could feel what the characters in the environment were going through [E]
2. I had the impression that the characters in the environment responded in a thoughtful way to what I did [C]
3. I noticed when the characters in the environment displayed strong emotions [A]
4. The characters in the environment seemed to have a strong will of their own [I]

Effectance

Users can easily recognize when and how they have causally affected the story or story world

The two facets of effectance: Immediate impact onto the system [II] and visible impact onto story development [SI] are both conceptually and item-based on the work of Klimmt, Hartmann, and Frey(2007). They have been partly adapted to fit the IS theme better (use of the term *story* instead of *game*).

Effectance (VUA, adapted from Klimmt, Hartmann & Frey 2007)

1. My inputs had considerable impact on the events in the story [II]
2. I had the feeling that I could affect directly something on the screen [II]
3. The consequences of my inputs were clearly visible [II]
4. I could recognize which events in the story I have caused with my inputs. [SI]
5. My decisions clearly influenced how the story went on. [SI]
6. I discovered how my earlier actions influenced what happened later in the story. [SI]



Autonomy

Users have the perception that can freely choose how to interact with and within the story world

All items were created from scratch while orientating a bit on the concept of autonomy as found in the Self Determination Theory (Deci & Ryan, 2000).

Autonomy (VUA 2011)

1. I had the impression that I was able to make many different events happen in the story
2. I noticed many opportunities to influence the story
3. I felt strong limitations to my decisions how the story should proceed
4. The system gave me precisely those options to influence the storyline that I had in mind

4.1.2 Part B: User Experience - Common and frequent experiential qualities

Curiosity

Users maintain interest in what will happen next and how they could affect the story

Only the Melbourne Curiosity Inventory (Naylor, 1981), and the State-Trait Personality Inventory (STPI; Spielberg et al, 1979) assess “curiosity as state”, which was needed. The curiosity scale below is adapted from the STPI. We changed the wording a bit, to make the scale applicable for measuring curiosity as a user experience. Again, agreement is given on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”. “N” denotes a negatively framed item.

Curiosity as a user experience (adapted from STPI, Spielberg et al, 1979)

During the experience, I felt...

1. ...like exploring my environment.
2. ...curious.
3. ...interested.
4. ...inquisitive.
5. ...eager.
6. ...in a questioning mood.
7. ...stimulated.
8. ...disinterested. (N)
9. ...mentally active.
10. ...bored. (N)

Suspense

Users develop hopes and expectations about the story progress, but also face uncertainty about that progress

Most scales used to measure suspense are context-specific. Based on the literature, and applying the suspense concept to IS environments, suspense has been conceptualized as “emotional involvement in a story’s outcomes”. The term “story outcomes” may apply to individual scenes, or to the story as a whole. Outcomes may be either positive or



negative, and therefore the emotions evoked may be positive or negative as well (e.g. hope, wistfulness, anticipation vs. fears and worries). Moreover, emotional responses may occur before the outcome (e.g. hope) or after the outcome (e.g. relief). A 10-item scale is used to assess suspense evoked by IS environments. Most statements explicitly relate emotional responses to story outcomes. Only Item 2 explicitly relates to the interactive nature of IS systems – although the notion of interaction may increase suspense, it is not necessarily conceptually related to it. Agreement is given on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”. “N” denotes a negatively framed item.

Suspense evoked by IS environments scale (VUA, 2010)

1. At some moments I was anxious to find out what would happen next
2. I was really hoping that the choices I made would work out well
3. I didn't care less how the story developed (N)
4. I found myself staring at the screen in anticipation
5. Sometimes I was worried about how the story would develop
6. Some moments were rather suspenseful
7. At some points I breathed a sigh of relief
8. I found myself wishing for a particular story outcome
9. The story did not affect me (N)
10. At some points I was afraid that things would go wrong

Flow

Users become absorbed in ongoing, continuous interaction with the story world

A 9-item flow scale is derived from the original Flow Short Scale (FSS-2) by Jackson, Martin and Eklund (2002) is readily available. The 9 items cover all nine dimensions of their original conceptualization of flow and apply neatly to the user experience of Interactive Storytelling applications. However some of the statements were slightly adapted for a better fit. Agreement is given on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”.

Short flow scale (VUA 2010, adapted from FSS-2; Jackson, Martin & Eklund, 2002)

During the experience...

1. ...I felt competent enough to meet the demands of the situation
2. ...I acted spontaneously and automatically without having to think
3. ...I had a strong sense of what I wanted to do
4. ...I had a good idea while I was performing about how well I was doing
5. ...I was completely focused on the task at hand
6. ...I had a feeling of total control over what I was doing
7. ...I was not concerned with how others may be evaluating me
8. ...the way time passed seemed to be different from normal
9. ... I found it extremely rewarding.



Aesthetic Pleasantness

Positive experiences of beauty or artistic impressiveness

A 5-item scale based on the work of Rowold (2008), Cupchik, G.C. & Laszlo, J. (1994), and Cupchik, G.C., Leonard, G., Axelrad, E., & Kalin, J. (1998) was developed to assess aesthetic pleasantness of Interactive Storytelling systems. Agreement is given on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”.

Aesthetic pleasantness (VUA, 2010, adapted from Rowold, 2008 and Cupchik et al, 1994)

The experience...

1. ... made me think (Rowold / VUA)
2. ... made me think about my personal situation (Rowold)
3. ... told me something about life (Rowold)
4. ... was inspiring (Cupchik / VUA)
5. ... moved me like a piece of art (VUA)

Enjoyment

An overall sense of positively valenced experiential quality

Here, enjoyment is conceptualized as a merely attitudinal measure, since the underlying dimensions of enjoyment are already covered by other scales. 13 items constitute this enjoyment in IS environments scale. The first 3 items refer to general enjoyment; the latter 10 refer to enjoyment with respect to five facets of enjoyment mentioned in Vorderer, Klimmt & Ritterfeld (2004): Amusement, Suspense, Melancholy, Aesthetic appreciation, and Sense of achievement. Agreement is given on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”.

Enjoyment in IS Environments (VUA, 2010)

The experience...

1. ...was pleasant (General)
2. ...was gratifying (General)
3. ...was rewarding (General)
4. ...was amusing (Amusement)
5. ...was exhilarating (Amusement)
6. ...was thrilling (Suspense)
7. ...was exiting (Suspense)
8. ...was melancholy (Melancholy)
9. ...was moving (Melancholy)
10. ...was appealing (Aesthetics)
11. ...was pleasing to the senses (Aesthetics)
12. ...made me feel proud (Achievement)
13. ...made me feel competent (Achievement)



4.1.3 Part C: System specific user reactions

Emotional state: positive / negative

Emotional reactions to IS application

To assess the emotional state directly after interacting with a storytelling system, the *Positive and Negative Affect Scale (PANAS)* by Watson et al. (1988) was used. This scale consists of a number of words that describe different feelings and emotions. Agreement is given on a 10-point Likert scale with numbered 1-10 without further labels. "N" denotes a negatively framed item.

PANAS (Watson et al., 1988)

How do you feel at this moment, after experiencing the story?

1. Interested
2. Sad (N)
3. Excited
4. Troubled (N)
5. Powerful
6. Guilty (N)
7. Scared (N)
8. Hostile (N)
9. Enthusiastic
10. Proud
11. Annoyed (N)
12. Alert
13. Ashamed (N)
14. Inspired
15. Nervous (N)
16. Determined
17. Careful (N)
18. Hysterical
19. Lively
20. Anxious (N)

Role adoption

Degree of identification with the story's protagonist

Three items to measure the role adoption and the identification with the protagonist were taken from the work within the European FUGA project (Hefner, Klimmt, & Vorderer, 2007). Agreement is given on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree".

Role adoption/identification (FUGA, 2010)

1. I felt like a was in the main character's skin
2. I sometimes forgot about myself because I was so focused on the actions of the main character
3. I felt more like the character than like myself



Comparison with other entertainment offers

The user is asked to compare the application with other story experiences

This scale was created to get an idea about the media perception of the user and how he compares it to already established story experiences found in different entertainment media and actions. Agreement is given on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”.

Comparison (VUA, 2011)

The experience was somewhat similar to...

1. Playing a video game
2. Watching a movie
3. Reading a story
4. Writing a story
5. Performing a play
6. Doing Improvisation Theater

Recommendation

Would you recommend this game to a friend? (yes/no)

Open Questions

“What did you like about the story?”, “What did you like or dislike?”

4.1.4 Part D: User information / Demographics

Gender (M/F)

Age (years)

Education

Computer Literacy

To assess the level of experience with video games a direct questions with three answer possibilities was created:

- How much experience do you have with video games? (beginner, moderate, experienced)
- Did you play this game before? (yes/no)

Open Questions

“Do you have any further questions or remarks?”