EVALUATION METRICS

Protocol and instruction

Project acronym:	POSTHCARD
AAL JP project number:	AAL-CALL-2017-045
Deliverable Id:	D4.3a
Deliverable Name:	Evaluation metrics
Status:	Running document
Dissemination Level :	Public
Due date of deliverable:	M12, M20
Actual submission date:	
Author(s):	Lisa Abdel Alim, Tessa Dadema & Henk Herman Nap
Lead partner for this deliverable:	VILANS
Contributing partners:	HUG; UNIGE

Contents

The process of co-design is an iterative process, in which potential end-users (informal carers) are closely involved in discovering how the POSTHCARD concept and platform should be shaped, functioning and used. The co-design approach reflects a fundamental change in the traditional developer-user relationship, enabling a wide range of people to make a creative contribution in the formulation and solution of a problem. While initially aiming to generate ideas with an open approach, the assessment of end-user perspectives becomes increasingly specifically focused on a certain problem and/or solution as the co-design process progresses, which is reflected in the figures below. The co-design process follows an iterative loop-based design and evaluation path from ideation, concept development, product development to product optimization.
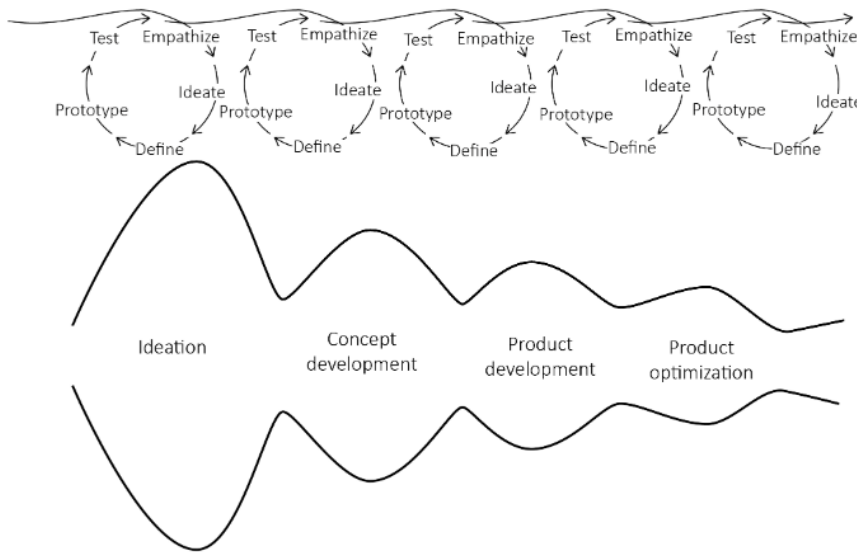
This deliverable desribes evaluation metrics for concept development and some phases in product development. The evaluation phases are described in the pilot test guidelines D4.2a deliverable (for alpha and beta prototype testing)
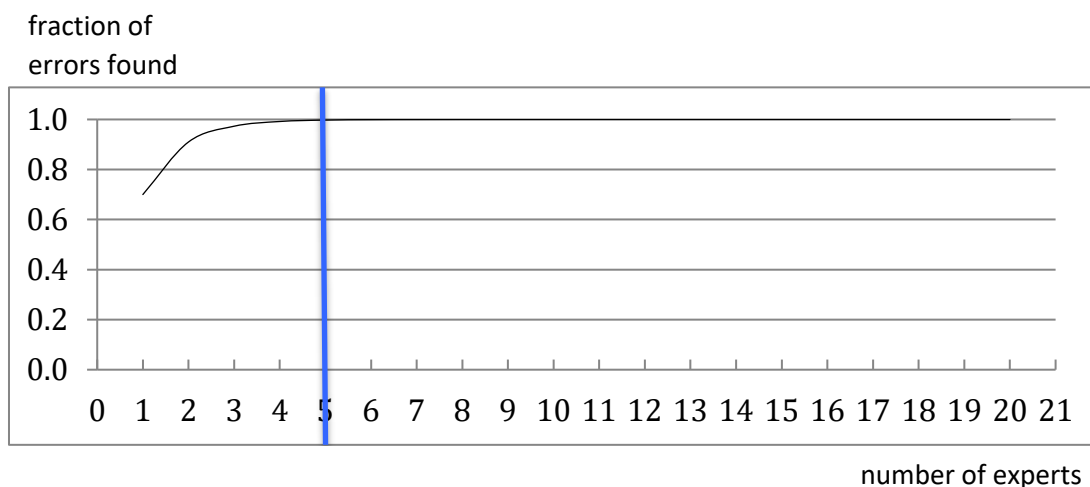
1. Heuristic Evaluation - Plan

1.1 Goal

For POSTHCARD, whenever the first mid-fidelity interfaces are ready for end-users to control settings and interact with the game elements, then a heuristic evaluation is planned to gain insight into the usability of the POSTHCARD interfaces.

During the heuristic evaluation the usability experts will analyse interaction with the product and detect usability problems/errors, so they can be solved before we will evaluate the game with the end users. Heuristic evaluation (Nielsen and Molich, 1990; Nielsen 1994) is a usability engineering method for finding the usability problems in a user interface design so that they can be attended to as part of an iterative design process. Heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognized usability principles (the "heuristics"). During the evaluation session, the evaluator goes through the interface several times and inspects the various dialogue elements and compares them with a list of recognized usability principles (the heuristics). These heuristics are general rules that seem to describe common properties of usable interfaces. In this blogpost the heuristics are very well described and examples are given: https://blog.prototypr.io/10-usability-heuristics-with-examples-4a81ada920c See Appendix for the heuristics that we will use in the Heuristic Evaluation.

The Heuristics evaluation will be conducted by means of an online evaluation form in Google Forms. This tool automatically collects and analyzes the replies of participants.

1.2 Participants

With 4-5 experts you find most of the errors found in an interface. See the Figure below. For the evaluation of POSTHCARD our goal is to evaluate the game with 6 experts in total. This means 2 experts per country, which should be persons outside of the POSTHCARD project group. The selected experts should have experience with usability of interfaces. Preferably ask experts in your close environment, who will take the time to seriously try out the POSTHCARD platform and provide feedback throughout the online Google Form.

1.3 Protocol

Instructions for the Heuristics evaluation:

1. Invite 3-4 usability experts outside of the project to evaluate the POSTHCARD platform (participants should reserve approximately one hour for this evaluation. Supervision is not required, the instructions that will be send to the participant provides sufficient information to complete the evaluation independently).

2. Send the link to the Google Form to the participants: https://goo.gl/forms/ors5jf4wkmvq4PAx2. Include the following instructions in the email to the participants:

   Dear *[participant],*

For a European AAL project, called POSTHCARD, we have developed a prototype of a serious game in which the interaction between an Alzheimer patient and an informal caregiver is simulated. The goal of this simulation is to teach informal caregivers suitable coping strategies for real-life situations they could encounter when caring for someone with Alzheimer's disease.

We would like to ask you to participate in a heuristic evaluation of the prototype of this simulation. For this heuristic evaluation we ask experts in the field of usability and interface design to test the prototype and provide feedback on the usability and understandability of the interfaces. Trying out the prototype and filling out the evaluation form will take up to one hour.

In this evaluation you will pay attention to different 'heuristics', about which you can read more when opening the link to the evaluation form. You will pay attention to the consistency and simplicity of information and visual elements, intuitiveness and perceived effort of performing actions, freedom in navigation, and the offered support throughout the simulation.

You can access the prototype, evaluation form and further instructions via the following link: https://goo.gl/forms/ors5jf4wkmvq4PAx2. Please complete all the questions and submit your feedback before *[date].*

We really appreciate your help and are looking forward to receive your feedback. If you have any questions, please don't hesitate to contact us.

Best regards,
[sender]

3. Wait until the replies of all participants are collected in Google Forms.
4. If replies are missing or incomplete, please remind/notify the participant. If any questions arise or incorrect replies are given, please assist the participants in completing the evaluation correctly.
5. When the results of all participants have been received, please inform Vilans. Vilans will then analyze the feedback and provide a sheet with points of improvement to be integrated in the simulation.

1.4 Rating & Ranking

After the heuristic evaluation, a rating & ranking study will be performed. As not all usability problems and functionalities might be as importance to be implemented directly in order for POSTHCARD to be successful, the level of importance of a singular requirement will be rating using the labels from MoSCoW. This divides requirements into Must, Could, Should, and Would:

- Must include requirements that are critical for the POSTHCARD system to become a success and be a useful product for its users.
- Should include requirements that are equally as important as must requirements but could be implemented in a different manner or at a later stage as well.
- Could include requirements that are desirable but not a necessity and could improve user experience or customer satisfaction for little development cost. These requirements will typically be included if time and resources permit.
- Would include requirements that are least critical, might have the lowest-payback items, or are less appropriate at this time. They are requirements that might be considered again in a later stage of the project.

The labels will be scored with a 4 for a Must, a 3 for a Should, a 2 for a Could, and a 1 for a Would. Means will be calculated and the problems found will be ranked and discussed with the project partners in respect to the final prioritization (also in respect to costs, feasibility, project goals).

2. Lab Testing (To be defined after prototype ready)

2.1 Goal

2.2 Participants

2.3 Protocol

3. Narrative Evaluation (To be defined after prototype ready)

3.1 Goal

3.2 Participants

3.3 Protocol

4. Credibility Evaluation (HUG)

4.1 Goal

The user interface design process consists of several steps. Some of them are evaluation steps. This is particularly the case with the credibility assessment technique. According to the eight criteria of interactive persuasion (Némery and Brangier; 2014), the credibility of an interface reflects its level of reliability, expertise, loyalty and legitimacy, its level of consistency with the reality of users' experiences. It is essential that this criterion be respected insofar as it refers to the ability of the interface to generate user confidence, which is a guarantee of their use and resilience (Némery and Brangier; 2014) and therefore satisfaction.

 In accordance with the recommendations of the design of the "user-centred" interface, the designers of the platform developed as part of the POSTACRD project must take into account all the characteristics and needs of its future users, by actively involving them at the various stages of the design process. Thus, as part of this credibility assessment, the interaction of formal caregivers with the platform will be analyzed in order to identify ergonomic problems/errors so that they can be resolved before an assessment is made with its future users.

During the evaluation session, the evaluator will be invited to test the platform, to inspect the different elements of dialogue, possibilities of action, game environment, and non-player characters's reactions/behaviours, either the reality of the system. They will then be invited to compare their observations with their reality as experts in this type of interaction simulated here.

More precisely, we will seek to determine the level of coherence between the reality of the system, the state of the platform in its current state and the real world. The objective will therefore be to find ways to increase it. To do this, we will call on experts in the care of people with Alzheimer's disease, formal caregivers.

4.2 Participants

For the POSTHCARD evaluation our goal is to evaluate the game with 10 healthcare professionals caring for people affected by Alzheimer disease.

Registered participants are professionals working and trained in the field of Alzheimer's disease care. They may be doctors, psychologists, health care assistants, nurses, home support workers, socio-educational assistants, community health care assistants, health care assistants working in medical and social institutions, hospitals or in the home as part of this type of care

4.3 Protocol

Instructions for assessing credibility:

Solicit 10 formal caregivers of people with Alzhiemer disease to evaluate the POSTHCARD platform (participants should reserve approximately one hour for this evaluation). These will be semi-directive interviews that will take place after the platform has been tested by participants.

1. Consent form

Dear Madam / Dear Sir,

As part of a European AAL project, called POSTHCARD, we developed a prototype of a serious game in which the interaction between an Alzheimer patient and a caregiver is simulated. The objective of this simulation is to teach caregivers coping strategies adapted to the real situations they may encounter when caring for someone with Alzheimer's disease.

We would like to ask you to participate in an evaluation of the prototype of this simulation, in which we are soliciting formal caregivers involved in the care and/or management of people with Alzheimer's disease.

Before agreeing to participate, it is important that you read this statement of consent and understand the procedure we will follow. If you have any questions or comments, please do not hesitate to let us know.

PURPOSE OF THE PROJECT

The objective of the POSTHCARD project is to develop an educational simulation for caregivers of people with dementia.

This simulation aims to improve the care of these people and thus reduce the overload and stress that caregivers may encounter during care.

PURPOSE OF THE CO-CREATION SESSION

The purpose of this co-creation session is to identify ways to improve the first prototype of the platform so that it is as adapted as possible to the situation of its future users, the caregivers, and can thus help them.

PROCEDURE

During this co-creation session, we will first ask you to play with the first prototype of the platform, then we will ask you questions about your gaming experience. We will use an audio recorder.

FREE PARTICIPATION

Participation in the co-creation session is completely free. You can decide to stop at any time without having to justify your departure.

RISKS

There is no risk associated with this study.

BENEFITS

Your personal experiences and opinions are very useful to us because we hope to offer an effective new tool for the relatives and for that we need you.

ANONYMITY

Information that could be linked in any way to any of your answers will automatically be deleted from the results collected and replaced by a code. This is done to ensure that the processing and presentation of data is done anonymously. Confidential information will be saved and password protected and accessible only by authorized personnel. The results of this co-creation session will be used for scientific research purposes.

Questionnaire content

General open question in link with Laaksolahti *et al. method (2001)*

*The purpose of this section is to collect information on the general level of credibility through an open-ended question, i*n order to know the feeling of formal caregiver about the platform.

The aim is to detect the part of emotionally aspect in their response to identify if the platform is credible or not.

If they describe the characters and their lives using emotionally rich terms, it means that users have felt empathy towards the characters and that they found them credible. Otherwise, if users are hesitant in their description and find that the characters are strange, so they didn't feel any empathy towards them and do not find them credible.

Question 1 and 2

Context credibility

The aim of this section is to assess the level of credibility of the game environment: contextualization, material/physical environment (objects)

Questions 3 to 5

Caregiver's credibility

> The purpose of this section is to gather information about the caregiver's credibility level (behavior, speech)
>
> Question 6 to 8

Sick character's credibility

> The purpose of this section is to gather information on non-player character's credibility level, the sick character. (reactions, expression of neuropsychological symptom).
>
> Questions 9 to 10

General questions to improve

> *Here, it will be a question of asking general questions about the strong points (question 12) and the weak points of the platform (question 13) identified by the healthcare professionals.*
>
> *We will also explicitly ask them if they have any ideas for improvements for the simulations, in particular to improve the pedagogical aspect of the platform (questions 14)*

Questionnaire

Question 1:

Can you please describe the characters. You can refer to their appearance, behavior and feelings.

Question 2:

Did you find that the characters interactions are coherent, credible and interesting?

Question 3:

Can you please tell me what do you think about the physical environment of the simulation?

Question 4:

Can you please tell me what do you think about the contextualization (text that you read at the beginning) of the simulation?

Question 5:

Have you got any idea to improve this aspect?

Question 6:

Could you or one of your colleagues say what the caregiver said?

Question 7:
Could you or your colleagues be brought to the same actions as the caregiver character did?

 Question 8:
Have you got any idea to improve caregiver dialogue?

Question 9:
Could one of your residents/patients react in this way?

Question 10:
Do you find the expression of neuropsychological symptoms of the disorder realistic?

Question 11:
Have you got any idea to improve this aspect?

Question 12:
What do you think are the strengths of the prototype you tested?

Question 13:
What do you think are the weak points?

Question 14:
Do you have any suggestions for us to improve the platform? In particular, in order to develop the pedagogical aspect?

5. Personalization evaluation (UNIGE)

5.1 Goal

5.2 Participants

5.3 Protocol

6. Appendix

HEURISTICS	Problems		
1. Consistency Icons, labels, buttons, and menus (i.e., elements) displayed on screen should be consistent in, location, terminology and meaning.	- - - -		
	Sub-Questions:	**Problem Description & Problem Severity**	
	- Do the elements follow platform conventions? (do as everyone else does) yes ☐ no ☐ n.a. ☐ *If **no**, describe problem & problem severity* ➔	*1* *not severe* *2* *hardly severe* *3* *severe* *4* *very severe*	
	- Are the elements directly understandable (i.e., yes ☐ no ☐ n.a. ☐		

	not ambiguous), in language and visuals?	*If **no**, describe problem & problem severity* ➜				
			1 *not severe*	*2* *hardly severe*	*3* *severe*	*4* *very severe*

HEURISTICS	Sub-Questions:	Problem Description & Problem Severity			
1b. Consistency *Continued* Icons, labels, buttons, and menus (i.e., elements) displayed on screen should be consistent in, location, terminology and meaning.	- Is a particular system action always displayed in the same manner and always achievable by one particular user action? ☐ yes ☐ no ☐ n.a. *If **no**, describe problem & problem severity* ➜				
		1 *not severe*	*2* *hardly severe*	*3* *severe*	*4* *very severe*

HEURISTICS	Problems
2. Simplicity Elements displayed on screen should not contain functionalities or information which is rarely needed or irrelevant.	- - - -
	Sub-Questions: Problem Description & Problem Severity

	- Do the rarely needed or irrelevant elements compete with and diminish the visibility of relevant units of information?	yes no n.a. *If **yes**, describe problem & problem severity* ➔				
			1 *not severe*	*2* *hardly severe*	*3* *severe*	*4* *very severe*

HEURISTICS	Problems					
3a. Feedback Elements displayed on screen should keep you informed about the past, current, and future system status.	- - - -					

	Sub-Questions:		Problem Description & Problem Severity			
	- Do these feedback elements keep you informed about what is going on within a reasonable time?	yes no n.a. *If **no**, describe problem & problem severity* ➔				
			1	*2*	*3*	*4*

		not severe	hardly severe	severe	very severe
- Do these feedback elements provide an answer to the questions: Where am I? Where have I been? & Where can I go?	yes no n.a. *If **no**, describe problem & problem severity →*				
		1 not severe	2 hardly severe	3 severe	4 very severe

HEURISTICS	Sub-Questions:	Problem Description & Problem Severity			
3b. Feedback *Continued* Elements displayed on screen should keep you informed about the past, current, and future system status.	- Do these feedback elements also provide information about how you've got here, how you can go back, and how you can go somewhere else? yes no n.a. *If **no**, describe problem & problem severity →*				
		1 not severe	2 hardly severe	3 severe	4 very severe
	- Are the responses of elements that provide feedback for minor and frequent actions modest? yes no n.a. *If **no**, describe problem & problem severity →*				
		1 not severe	2 hardly severe	3 severe	4 very severe

4

HEURISTICS	Problems
4a. Control Elements displayed on screen should provide you with control and freedom.	- - - -

HEURISTICS	Sub-Questions:	Problem Description & Problem Severity
4b. Control *Continued* Elements displayed on screen should provide you with control and freedom.	- Is there are a clearly marked 'emergency exit' to leave an unwanted state? ☐ ☐ ☐ yes no n.a. *If **no**, describe problem & problem severity* →	*1* *2* *3* *4* *not severe* *hardly severe* *severe* *very severe*
	- Are there undo and redo options? ☐ ☐ ☐ yes no n.a. *If **no**, describe problem & problem severity* →	*1* *2* *3* *4* *not severe* *hardly severe* *severe* *very severe*
	- Do the elements respond to your actions? ☐ ☐ ☐ yes no n.a. *If **no**, describe problem & problem severity* →	*1* *2* *3* *4* *not severe* *hardly severe* *severe* *very severe*

HEURISTICS	Problems
5. Error Elements displayed on screen should help you recognize, diagnose, and recover from an error.	- - - -

	Sub-Questions:	Problem Description & Problem Severity
	- Do these elements display the error in natural language, indicate the problem, and suggest a solution and what the effect of this will be?	☐ yes ☐ no ☐ n.a. *If **no**, describe problem & problem severity ➜*

	1 *not severe*	2 *hardly severe*	3 *severe*	4 *very severe*

- Are the displayed errors blaming the problem on user deficiencies? (the user is always right)	☐ yes ☐ no ☐ n.a.

	If **yes**, describe problem & problem severity ➔	1 not severe	2 hardly severe	3 severe	4 very severe

HEURISTICS	Problems				
6a. Overload The elements displayed on screen should minimize the memory load of the user.	- - - -				
	Sub-Questions:	**Problem Description & Problem Severity**			
	- Are there elements that provide instructions for use of the system and are these instructions simple and understandable?	☐ ☐ ☐ yes no n.a. *If **no**, describe problem & problem severity* ➔			
			1 not severe	2 hardly severe	3 severe	4 very severe
	- Are the elements on screen static or at	☐ ☐ ☐ yes no n.a.			

	least low in motion frequency? *If **no**, describe problem & problem severity* →	1 not severe	2 hardly severe	3 severe	4 very severe

HEURISTICS	Sub-Questions:	Problem Description & Problem Severity

6b. Overload *Continued*

The elements displayed on screen should minimize the memory load of the user.

- Are there more than 7 elements within (wide) an action sequence, and more than 3 action sequences necessary to perform a task?

yes   no   n.a.

*If **yes**, describe problem & problem severity* ➔

1	2	3	4
not severe	*hardly severe*	*severe*	*very severe*

- Do the elements and action sequences contain metaphors that are known by you?

yes   no   n.a.

*If **no**, describe problem & problem severity* ➔

1	2	3	4
not severe	*hardly severe*	*severe*	*very severe*

- Do you have to remember information from one part of the system to another?

yes   no   n.a.

*If **yes**, describe problem & problem severity* ➔

1	2	3	4
not severe	*hardly severe*	*severe*	*very severe*

HEURISTICS	Sub-Questions:	Problem Description & Problem Severity
6b. Overload *Continued* The elements displayed on screen should minimize the memory load of the user.	- Are there elements that provide shortcuts to frequently made actions? ☐ yes ☐ no ☐ n.a. *If **no**, describe problem & problem severity* ➤	

			1 *not severe*	2 *hardly severe*	3 *severe*	4 *very severe*

A1. Could you come up with additional **future** usability problems when the system has been frequently used, possibly by multiple users?

comments & suggestions:

Literature

Coats, R.B., & Vlaeminke, I. (1987). *Man-computer interfaces: an introduction to software design and implementation.* Cambridge, MA: Blackwell Scientific Publications, Inc.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

Molich, R., and Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM 33*, 3 (March), 338-348.

Nielsen, J., and Molich, R. (1990). Heuristic evaluation of user interfaces, *Proc. ACM CHI'90 Conf.* (Seattle, WA, 1-5 April), 249-256.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. *Proceedings ACM CHI'92 Conference* (Monterey, CA, May 3-7), 373-380.

Nielsen, J. (2005). How to Conduct a Heuristic Evaluation. Retrieved 20 November, 2008, from: http://www.useit.com/papers/heuristic/heuristic_evaluation.html

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3rd ed.) . Reading, MA: Addison-Wesley Publishing.