Exploiting the meCUE Questionnaire to Enhance an Existing UX Evaluation Method Based on Mental Models

Stefano Filippi(✉) and Daniela Barattin

DPIA Department, University of Udine, Udine, Italy
filippi@uniud.it

Abstract. Several definitions of User eXperience (UX) are present in the literature, representing the references for different UX evaluation methods and tools. Among these methods and tools, the irMMs-based method and the meCUE questionnaire show different peculiarities and seem good candidates to speculate on their integration, aiming at improving applicability, completeness and effectiveness of UX evaluation activities. The goal of this research is indeed to integrate the irMMs-based method with the meCUE questionnaire in order to enhance the former by exploiting the stronger points of the latter. To achieve this, we analyze the lacks of the irMMs-based method and verify the meCUE questionnaire capabilities in overcoming these lacks. Once verified, we proceed with the integration by modifying the evaluation activities of the irMMs-based method. Finally, we perform a first validation of the enhanced release by comparing it to the old one in the field. Researchers can exploit the same procedure to integrate other methods and tools in order to improve them; on the other hand, industrial practitioners can use the enhanced release for their UX evaluations since it is ready to be applicable even by non-developers.

Keywords: UX evaluation · irMMs-based method · meCUE questionnaire

1 Introduction

Nowadays, User eXperience (UX) is a key factor in designing, developing and managing interactive products and services [1]. Nevertheless, a clear and universal UX definition is still missing. UX fundamentals vary in terms of scopes and objects/elements considered; researchers and industrial practitioners seem to have different opinions about UX [2]. According to the ISO 9241-2010, the UX includes users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments that occur before, during and after the use of a product, system or service [3]. The hedonic/pragmatic model of Hassenzahl [4] assumes that people perceive interactive products along two dimensions, pragmatics and hedonics. Pragmatics refers to the perceived product ability to support the achievement of specific "do-goals" connected to users' behaviors like "making a call", "choosing a dress", etc. The assessment of pragmatics focuses on the usefulness and usability of a product relating to potential tasks. Hedonics refers to the perceived product ability to support the achievement of specific "be-goals" connected to the user

A. Marcus and W. Wang (Eds.): HCII 2019, LNCS 11586, pp. 117–133, 2019.
https://doi.org/10.1007/978-3-030-23535-2_8

himself/herself like “being expert”, “being special”, etc. The assessment of hedonics focuses on human needs considering three different facets: “stimulation” (novelty and change, personal growth), “identification” (communication of identity to relevant others, relatedness) and “evocation” (provoking memories, symbolizing). Users, products and contexts of use influence pragmatics and hedonics; specific emotions and behaviors are generated consequently. The UX framework of Kort, Vermeeren and Fokker [5] considers UX as based on three aspects influenced by the design elements of products: the compositional aspects focusing on usability and on pragmatic and behavioral aspects of the experience, the aspects of meaning representing the symbolic significance of the experience and the aesthetic aspects addressing the capability of the products to satisfy one or more senses. Emotions are the results of the interactions among these three aspects. Other theories consider emotions as a direct component of UX instead of as consequences of other things. The product experience framework of Desmet and Hekkert [6] considers UX as composed by the aesthetical experience - the product capability to delight one or more human sensory modalities, the experience of meaning - personal and/or symbolic significance assigned to a product by a user, and the emotional experience - appraisal of an event or a situation as potentially beneficial or harmful. In the CUE model [7], the UX components are the perception of instrumental product qualities (usability and usefulness), the perception of non-instrumental product qualities (aesthetic, symbolic and motivational aspects) and the emotional reactions. These components generate the overall judgment of the product, usage behavior, user preferences between alternatives and intention to use.

The literature offers different UX evaluation tools based on these definitions. The AttracDiff 2 [8], based on the hedonic/pragmatic model, is a questionnaire that measures the perceived pragmatic qualities and a part of the perceived hedonic qualities (stimulation and identification). Based on the same model, the User Experience Questionnaire (UEQ) measures the pragmatic (perspicuity, efficiency and dependability) and hedonic (stimulation and novelty) qualities through the evaluation of 26 items [9]. Among the methods and tools that consider emotions as a direct UX component, the irMMs-based method [10], based on the product experience framework, provides a qualitative UX evaluation by exploiting mental models and emotions. This evaluation occurs thanks to the comparison of the users’ expectations about the interaction with a product and the users’ evaluation after the real interaction. The meCUE questionnaire [8], referring directly to the CUE model, quantifies the UX of a product by exploiting question-like statements about instrumental and non-instrumental product qualities, emotional reactions and consequences of use of the product.

The irMMs-based method and the meCUE questionnaire show different peculiarities. For example, the former provides a qualitative UX evaluation while the latter a quantitative one; the latter considers product aspects not covered by the former like aesthetics, consequences of use, etc. This makes the two tools good candidates to speculate on their integration, aiming at improving applicability, completeness and effectiveness of UX evaluation activities. The goal of this research is indeed to integrate the irMMs-based method with the meCUE questionnaire in order to enhance the former by exploiting the stronger points of the latter. To achieve this, we analyze the lacks of the irMMs-based method and verify the meCUE questionnaire capabilities in overcoming these lacks. Once verified, we exploit the meCUE questionnaire to modify the

irMMs-based method by highlighting the phases to affect and the way to intervene. Finally, we perform a first validation of the enhanced release by comparing it to the old one. The outcomes of this research can be useful for both researchers and industrial practitioners. The former can exploit the same procedure adopted here to integrate other methods and tools in order to improve them; on the other hand, the latter can use the enhanced release for their UX evaluations since it is ready to be applicable even by non-developers.

2 Background

2.1 The irMMs-Based UX Evaluation Method

The irMMs-based UX evaluation method (irMMs method, hereafter) evaluates the experiences of users interacting with products by exploiting interaction-related mental models (irMMs). The irMMs are cognitive processes that users generate in their mind before to act in order to satisfy a specific need in a specific situation of interaction. They consist of lists of users' meanings and emotions, including users and products' behaviors determined by these meanings and emotions [10]. The generation of an irMM develops through five steps, based on the Norman's model of the seven stages of the action cycle [11]. In the first step, the user perceives and interprets the need. Thanks to this interpretation, in the second step the user recovers previous irMMs and selects those ones that could be suitable to satisfy the need. The presence of one or more real products could influence this selection. The selected irMMs allow the user to establish the goals in the third step. Goals represent intermediate and time-ordered results to achieve; they help in defining the path to follow to satisfy the need. In the fourth step, the user assigns expected meanings and emotions to the need. These come from the elaboration of meanings and emotions belonging to the irMMs selected before. Positive meanings and emotions tend to remain as they are; negative ones could change into their positive correspondences depending on the amount of previous experiences that report those negative meanings and emotions. In the fifth step, the user defines his/her behavior and the related product behavior in order to achieve those expected meanings and emotions. These behaviors come again from elaborating the behaviors present in the irMMs selected before.

The adoption of the irMMs method happens through user tests. Users generate their irMMs respect to a specific need to satisfy and compare these irMMs to the real interaction with the product. The adoption generates two lists of positive and negative UX aspects that describe the strong points and the criticalities of the experience, respectively. More specifically, the irMMs method consists of three sections that differ in the knowledge about the product of the users who undergo the tests. The first section considers users who do not know the product at all; they generate their irMMs based on previous experiences with different products only. This is the absolute beginners (AB) section. The second section considers again users who do not know the product; nevertheless, before the generation of the irMMs (and before to know the need to satisfy) they are allowed to interact freely with the product for some time. This is the relative beginners (RB) section. Finally, the third section considers users who already

know the product. This is the relative experts (RE) section. Figure 1 summarizes the adoption of the irMMs method. The tests for the three sections can run in parallel, providing that no influences among them happen in the meantime.

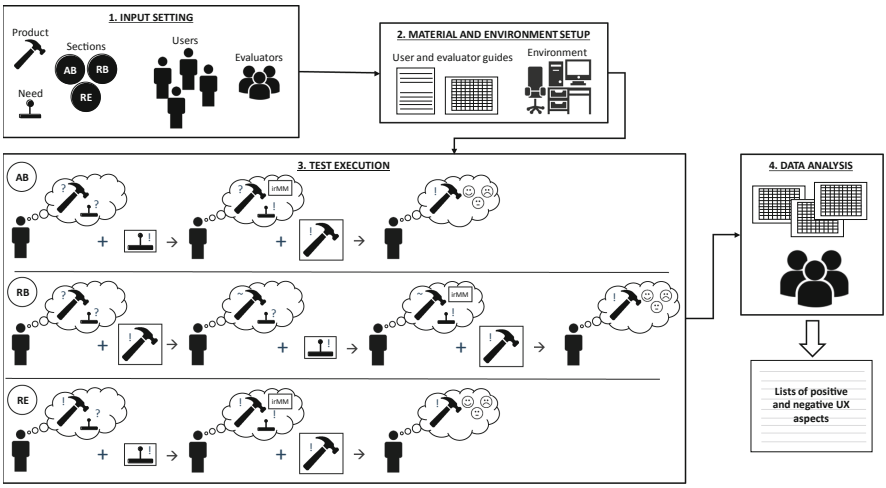

Fig. 1. The irMMs method adoption.

Phase 1. Input Setting. This phase defines five inputs. The features of the product to evaluate are the first input; the need to satisfy is the second one. Features can be functions, procedures, physical components, etc. and users will try to satisfy the need by exploiting these features. The third input is the sections of the irMMs method to adopt. Their selection depends on the expectations about the evaluation outcomes, on the resources available, etc. For example, aiming at investigating about the learnability of a feature, AB and RB sections would be the best candidates because RE users - those users involved in the tests of the RE section - already know the product and cannot give indications about learnability. The fourth input are the users who undergo the tests. Their selection comes by obeying to several requirements. For example, AB and RB users must not know the product before the evaluation; on the contrary, RE users must know the product and have used it at least for a given period (its duration varies time by time, depending on the product complexity, on the required reliability of the evaluation results, etc.). Another requirement, valid for every section, is "users must have similar knowledge about the field the product belongs to". More requirements can apply due to specific evaluation characteristics. Finally, the fifth input concerns the evaluators. Regarding AB and RB sections, evaluators can be almost any, from skilled and knowledgeable ones about the product to those barely aware of it; contrarily, for the RE section, evaluators must know the product very well because, in case of users running into very specific problems, they must be able to overcome these problems quickly and easily.

Phase 2. Material and Environment Setup. The second phase prepares material and environment to perform the tests. Material consists of two documents for each section, the user and the evaluator guides. The user guide helps users in generating the irMMs and in performing the interaction with the product. Its first part reports the need to satisfy and the instructions to generate the irMM, together with examples. The second part shows suggestions to perform the interaction with the real product and to compare it to the irMM. This document contains tables and empty spaces to describe the irMM and to comment the comparison. In the case of the AB section, the user guide is free of references to the product to evaluate in order to avoid bias. On the contrary, the user guides for the RB and RE sections contain precise information on the product under evaluation to make the users' attention focusing on it rather than on products that the users could know and/or have used in the past. The evaluator guide allows evaluators collecting data classified against their type (meanings, emotions, etc.) and tagged with the references to the users who generate them; moreover, this document contains instructions to analyze the data. The test execution phase also requires the setup of a suitable environment reflecting the common use of the product under evaluation. Depending on the characteristics of the specific evaluation, this phase selects a UX lab where the equipment to collect data is already present or identifies a real environment to perform the tests and moves the equipment there.

Phase 3. Test Execution. The third phase starts with the RB users freely interacting with the product. The evaluators invite them to focus on the product features selected in the input setting phase. Once this free interaction comes to the end, every user generates his/her irMM. Then, he/she tries to satisfy the need by using the product. The user must behave as described in his/her irMM; in the meantime, he/she checks if the product allows performing his/her behavior and if he/she gets the expected reactions/feedback (behavior) from the product. Of course, problems can occur anytime; these problems are addressed as gaps. If a gap shows up, the evaluators suggest the way to overcome the problem in terms of user and product behaviors. The user reports the reasons for the gap from his/her point of view and judges if the allowed behaviors are better or worse than those described in his/her irMM. Once finished the interaction, a debriefing where the user reasons about his/her experience and expresses further comments about it takes place. In particular, the evaluators invite the user to reconsider meanings and emotions expressed during the generation of the irMM in order to highlight and comment any change.

Phase 4. Data Analysis. Three rules lead the evaluators' analysis of the data collected during the tests. These rules allow generating the positive and negative UX aspects from the expected and real meanings and emotions as well as from the gaps. These rules are as follows.

- First rule. Every expected meaning and emotion expressed by the RB and RE users generates an UX aspect. This aspect will be positive or negative depending from the positivity/negativity of the meaning or emotion it comes from. For example, consider the need "washing delicate clothes with a washing machine". One meaning associated to this need can be "temperature", with a positive judgment because of the reason "since it is difficult to know the best temperature for washing any type of

clothes, the washing machine has predefined programs where all parameters are already set, temperature included". This meaning and the related reason would allow generating the positive UX aspect "the washing machine offers predefined programs where all parameters are already set; the user must only select the right program according to the type of clothes to wash". This rule does not consider the AB users because their expected meanings and emotions refer to products different from the one under evaluation.

- Second rule. Every gap generates an UX aspect. Even in this case, the UX aspect will be positive or negative depending from the positivity/negativity of the gap it comes from. For example, consider again the need about washing delicate clothes. A gap related to the expected product behavior "all parameters of the washing program shown in the display" could raise during the real interaction because the product shows only washing time and temperature. The judgment about the gap would be negative, with the reason "when a washing program is selected, all washing parameters should be shown on the display to be sure to wash the clothes in the right way and safely. Unfortunately, the spin dryer speed is not shown". This gap would allow generating the negative UX aspect "the information reported on the display about the selected washing program is incomplete because spin dryer speed is missing".
- Third rule. Every change between the expected meanings/emotions and the corresponding real ones generates one or more UX aspects. If the change shows a positive trend, the UX aspect(s) will be positive, negative otherwise. This generation occurs considering the reasons of the changes expressed by the users. These reasons can refer to the interaction in general or they can point at specific gaps. In the first case, only one UX aspect arises; in the second case, also the UX aspects generated starting from the pointed gaps are considered as responsible for the change. For example, consider always the need of washing delicate clothes. The expected emotion "happy" can become "very happy" once the real interaction has occurred. There is a positive trend between the expected emotion and its real mate. The reason for this change could be "programming the washing time to work at nighttime is very interesting because, in this way, the risk to forget switching on the washing machine highly decreases as well as money are saved (at nighttime, electricity is cheaper)". This reason would allow generating the positive UX aspect "the washing machine is smart in choosing the right time to launch the washing program". Nevertheless, also the positive gap "the display proposed to the user a washing time to launch the washing program, expecting just a confirmation" could be responsible for this change; therefore, also the UX aspect "the washing machine suggests the best time to launch the washing program automatically" derived from that gap would be associated to this change.

As soon as the application of the rules comes to the end, the comparison of all the UX aspects takes place in order to delete doubles. Then, the UX aspects are classified against the interaction topics they refer to (specific procedures, product components, etc.) and, for each topic, the UX aspects are split into positive and negative and ordered against the number of occurrences and the impact. If an UX aspect refers to product characteristics rarely involved in the interaction, its impact will be low; on the contrary,

if the UX aspect deals with core procedures determining the cognitive compatibility of the product with the users' problem solving processes, the impact will be higher.

2.2 The meCUE Questionnaire

The meCUE questionnaire 2.0 (meCUE, hereafter) provides a quantitative UX evaluation starting from the Components model of User Experience (CUE model) of Thuring and Mahlke [12]. As shown in Fig. 2, this model considers the perceptions of instrumental and non-instrumental product qualities and the emotional reactions as main components of the UX.

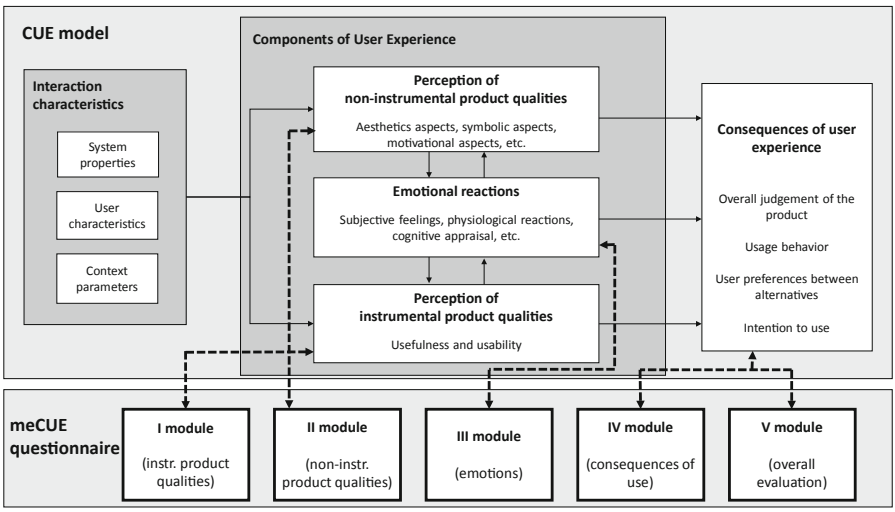

Fig. 2. CUE model and meCUE questionnaire.

The perceptions of non-instrumental product qualities consider aesthetics, symbolic and motivational aspects (visual and/or haptic qualities, associative or communicative symbolics, and personal growth, respectively). The emotional reactions consider aspects like subjective feelings, physiological reactions, cognitive appraisal, etc. Finally, the perceptions of instrumental product qualities include aspects like usefulness and usability. System properties, user characteristic and context parameters, all of them addressed in the figure as interaction characteristics, influence these three components. The consequences of the user experience, i.e., the overall judgement of the product, the usage behavior, the user preferences between alternatives and the intention to use, represent the outcomes of the main components. The meCUE questionnaire covers the CUE model thanks to five modules describing the perception of instrumental and non-instrumental product qualities, emotional reactions, consequences of use and the overall evaluation [8]. The first four modules contain question-like statements (thirty-four in all) that are evaluated using a 7-point Likert Scale reaching from 1 (strongly disagree) to 7 (strongly agree). The fifth module contains only the statement

"finally, how would you rate the product overall?" and users rate it using values between -5 (bad) and 5 (good). All users' answers are collected separately for each module in order to compute specific mean values. Any combinations of modules can be selected to match the goal of the evaluation at best.

3 Activities

In order to achieve the goal of the research, we analyze possible lacks of the irMMs method and verify the meCUE capabilities in overcoming these lacks. After this verification, we integrate the irMMs method with the meCUE. Then, a first adoption of the enhanced release takes place in the field to verify its applicability, completeness and effectiveness.

3.1 Highlighting of irMMs Method Lacks

The analysis of the irMMs method, together with the authors' experience about its development and adoptions, allow highlighting the following lacks.

(a) The irMMs method allows performing a qualitative evaluation of UX by highlighting positive and negative aspects; nevertheless, it does not quantify the UX.
(b) The irMMs method manages the cognitive compatibility and the meanings (symbolic aspects) of the product, but it does not focus on other non-instrumental product qualities like aesthetics, motivational aspects, etc. Moreover, regarding the instrumental product qualities, the irMMs method takes care about usability but deserves usefulness.
(c) In the irMMs method, the way to express expected and real emotions is left to the users; unfortunately, this allows users referring to the same emotions differently time to time and makes their comparison hard to do.
(d) The irMMs method does not take care about possible changes in the users' opinions after the real interaction regarding usage behavior, user preferences between alternative products and intention to use.

3.2 Verifying the meCUE Capabilities in Overcoming the irMMs Method Lacks

The strongest point of the meCUE is the quantitative evaluation of the UX. This assures the overcoming of the first lack of the irMMs method (labelled as 'a' in the previous paragraph). The first two meCUE modules regard instrumental and non-instrumental product qualities; this can help overcoming the lack 'b' of the irMMs method since they deal with all those UX components deserved by the irMMs method now. The meCUE provides a fixed set of generic statements to define emotions; this makes their quantification and comparison easier, overcoming the lack 'c' of the irMMs method. Finally, the fourth module of the meCUE focuses on the consequences of use by collecting users' opinions about usage behavior, user preferences between alternatives and intention to use while the fifth module performs an overall evaluation of the whole

experience by referring to the overall judgment of the product. Indeed, all of this can help in overcoming the lack 'd' of the irMMs method because it allows a deeper investigation of users' evaluation after the real interaction.

3.3 Integrating the irMMs Method with the meCUE

Once assessed the meCUE suitability to the research goal, the integration of the irMMs method takes place. Figure 3 shows the result of this. White boxes represent those activities of the irMMs method that remain as they are; four grayed boxes contain activities modified and seven grayed boxes with bold text indicate the new activities added by the integration.

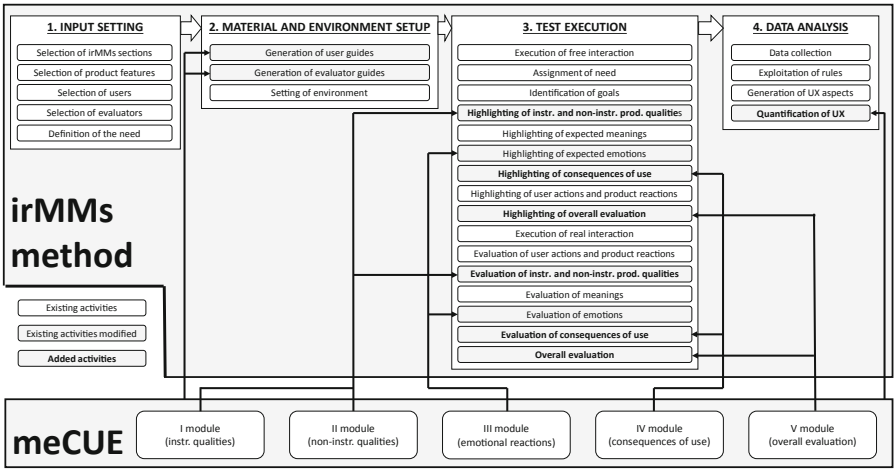

Fig. 3. Integration of the irMMs method with the meCUE.

The integration implies changes in phase two (material and environmental setup), changes and additions in phase three (test execution) and additions in phase four (data analysis).

Changes in phase two affect the existing activities "generation of user guides" and "generation of evaluator guides"; they occur as follows.

- The statements of the first meCUE module, dealing with the instrumental product qualities, are considered in all user and evaluator guides. Referring to the RB and RE user guides, these statements are introduced after the goal identification and before the highlighting of the meanings phases in order to collect the expectations about the product. The same statements are also introduced after the real interaction for all the sections to collect the users' final evaluation. The evaluator guides are modified to collect the users' answers classified by section and to compute the final mean values. In order to obtain this, empty tables are introduced after the free space devoted to the generation of the UX aspects.

- The second module, dealing with the non-instrumental product qualities, is considered in all user and evaluator guides as well. Its statements are inserted in the RB and RE user guides just after those of the first module. The same statements are also introduced again after the real interaction after the statements of the first module for all the sections. The presence of the meanings in the irMMs method suggests adding two statements not appearing in the meCUE. These statements are "the product evokes in my mind many different past experiences" and "I can identify the product in many situations of my life". The new statements are added to quantify the contribution of the meanings while reasoning about non-instrumental product qualities. In the evaluator guides, empty tables are added after those focused on the instrumental product qualities to collect and manage the users' answers.
- The statements of the third module, dealing with emotions, replace completely the free space devoted to emotions in all user guides. In the evaluator guides, the free space devoted to the expected and real emotions management is replaced by empty tables collecting the users' answers.
- Finally, the fourth and fifth modules, dealing with consequences of use and overall evaluation, are considered again in all user and evaluator guides. In the RB and RE user guides, the statements of these two modules are added just after the highlighting of the expected emotions phase. The same statements are also introduced after the evaluation of emotions phase for all the sections to collect the users' final evaluation. In the evaluator guides, empty tables are added after those referring to emotions to collect the users' answers.

As an example of changes due to the integration, Fig. 4 shows the new part of the user guides dealing with the highlighting of the consequences of use phase. The statements are visible in the lower part of the figure.

Considering phase three, the integration adds the new activities "highlighting of instrumental and non-instrumental product qualities", "highlighting of consequences of use", "highlighting of overall evaluation", "evaluation of instrumental and non-instrumental product qualities", "evaluation of consequences of use" and "overall evaluation". Moreover, the integration introduces changes to the existing activities "highlighting of expected emotions" and "evaluation of emotions". These additions and changes correspond to the answering to every statement just introduced in all the user guides.

Finally, considering phase four, the integration adds the new activity "quantification of UX". The new empty tables added by the integration allow the evaluators collecting the users' answers divided by section. Considering each section separately, evaluators compute the mean values two times. The first computation is on the answers of the RB and RE users about the highlighting of their expectations on the product; the second computation refers to the answers of all the users about the final evaluation of the UX. Finally, the evaluators compute the overall mean value considering the sections altogether. This new activity enriches the evaluation outcomes from two points of view: first, quantitative data appear along with the qualitative ones; second, the granularity of this computation allows comparing the results of the sections to each other.

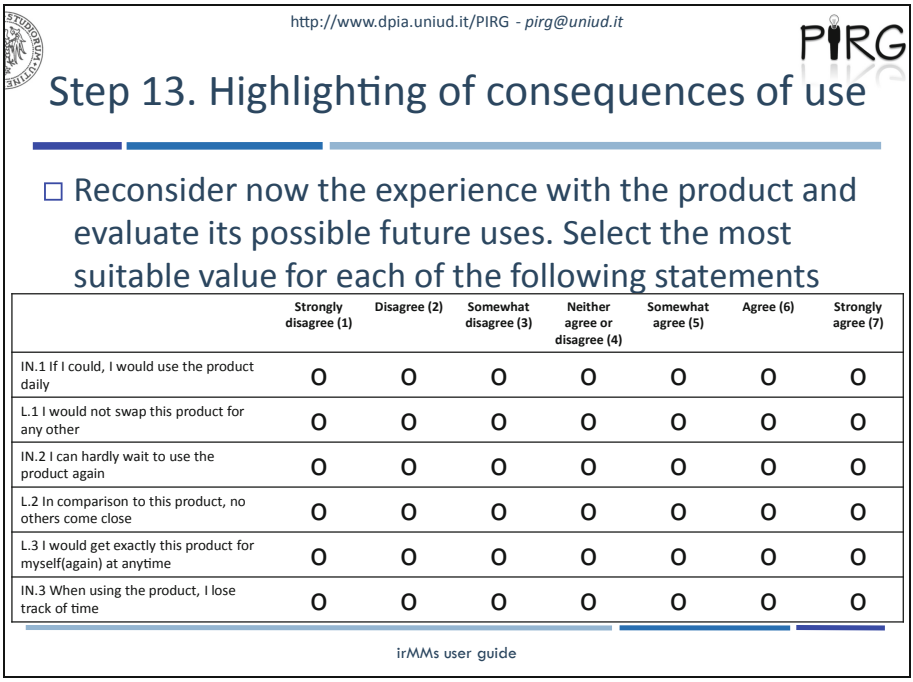

Fig. 4. The new part of the user guides related to the highlighting of the consequences of use phase.

4 First Validation of the Enhanced irMMs Method

What follows describes the adoptions of the old and of the enhanced releases of the irMMs method. These adoptions evaluate the UX of a state-of-the-art CAD software package developed by a well-known software house working in the engineering research and application field. This choice comes from the availability of past adoptions of the old release with the same product as well as of that of the users and evaluators to involve.

4.1 Adoption of the Old Release of the irMMs Method in the Field

The adoption of the old release occurred as described in Filippi and Barattin [10]. Only a summary of that experience is reported here.

The 3D modeling tool of the CAD software package was selected as product feature to evaluate and the need was “use the CAD software package available on the PC to generate the 3D model of the socket shown in the enclosed figure. This model will be for production. Please respect the assigned dimensions”. All the sections of the irMMs method were selected and thirty users (ten for each section) took part to the tests; they were students of mechanical engineering university courses with good skill and knowledge about 3D modeling since they used one or more CAD software

packages in the past. Three researchers very knowledgeable with the CAD field acted as evaluators. The user and evaluator guides were customized on the specific evaluation, while a university lab consisting of two separate rooms was selected as the environment to perform the tests.

Overall, this adoption generated seventeen positive and seventy-one negative UX aspects in thirty-eight hours.

4.2 Adoption of the Enhanced Release of the irMMs Method in the Field

The adoption of the enhanced release uses the same input and environment of the old release adoption except for the users. Of course, these are different to avoid bias due to fixation, etc.; nevertheless, they are always thirty, ten for each section, with comparable skill and knowledge. Due to all these similarities, the description of the adoption of the enhanced release highlights only the differences respect to the old one.

The main differences are in the test execution and data analysis phases. In the test execution phase, RB and RE users answer to the statements about instrumental and non-instrumental product qualities after the identification of the goals. The highlighting of expected emotions phase happened thanks to the statements coming from the third meCUE module; users express emotions easier and in a more structured way than before. After that, RB and RE users answer to the statements about consequences of use and overall evaluation. After the real interaction, users of all sections evaluated instrumental and non-instrumental product qualities that they had just experienced. The same statements to highlight expected emotions allow expressing the emotions after the real interaction. Then, users state the consequences of use and overall evaluation of the product. In the data analysis phase, the evaluators perform the quantitative analysis by exploiting the answers to the statements and compute the mean values for each meCUE module.

No problems arise all along the tests and this starts assessing the applicability of the enhanced release.

Overall, this adoption generates twenty-six positive and eighty-seven negative UX aspects in forty hours.

4.3 Results of the Two Adoptions

Table 1 reports the results of the two adoptions. AB, RB and RE columns contain the values corresponding to the three sections, separately; the ALL columns contain the results independently from the sections the data come from. The first row of the table reports the numbers of positive and negative UX aspects, without doubles. Next, gaps are classified as positive and negative as well, doubles excluded again. The Goals row contains the mean values of the goals the users think as necessary to satisfy the need. The Expected meanings and Expected emotions rows contain the numbers of them, doubles excluded. It must be noticed that the number of expected emotions highlighted by the enhanced release is considered differently than before. Now, the set of emotions is fixed and users just rate them. If a user does not feel one or more emotions, he/she rates them with the neutral value 4, “neither agree nor disagree”, because the user guides do not allow marking those emotions as “not applicable”. For this reason, the

enhanced release considers as expected all those emotions scoring differently from 4. The Changes of meanings and Changes of emotions rows contain the numbers of positive and negative changes, doubles excluded. The rows from Instrumental product qualities to Overall evaluation contain the mean values of the scores given by the users to the corresponding statements before and after the real interaction. Obviously, these rows apply only for the enhanced release; moreover, the mean values consider RB and RE sections only (except for the module devoted to the emotions that considers AB users too) because AB users cannot be asked for opinions about a specific product that they cannot know in advance. Finally, the last row reports the time to conduct each adoption.

Table 1. Results of the adoptions of the old and of the enhanced releases.

Results		Old release				Enhanced release			
		AB	RB	RE	ALL	AB	RB	RE	ALL
UX aspects (#; no doubles)	Positive	4	8	11	17	9	10	13	26
	Negative	36	24	22	71	40	31	22	87
Gaps (#; no doubles)	Positive	2	4	4	8	5	7	7	14
	Negative	20	11	9	36	20	14	15	39
Goals (mean)		6.0	4.7	3.3	4.7	5.2	5.0	4.7	4.9
Expected meanings (#; no doubles)		5	10	8	14	9	5	8	16
Expected emotions (#; no doubles)		8	5	5	8	10	12	11	12
Changes of meanings (#; no doubles)	Positive	0	0	2	2	2	0	1	3
	Negative	3	1	4	8	2	1	2	4
Changes of emotions (#; no doubles)	Positive	4	8	6	12	13	4	6	9
	Negative	17	10	5	22	4	2	4	6
Instrumental product qualities (means)	Before	N/A	N/A	N/A	N/A	N/A	5.89	5.11	5.50
	After	N/A	N/A	N/A	N/A	5.33	5.89	5.17	5.53
Non-instrumental product qualities (means)	Before	N/A	N/A	N/A	N/A	N/A	4.39	3.53	3.96
	After	N/A	N/A	N/A	N/A	3.89	4.36	3.26	3.81
Emotions (means)	Before	N/A	N/A	N/A	N/A	2.97	3.72	3.58	3.42
	After	N/A	N/A	N/A	N/A	3.39	3.86	3.58	3.61
Consequences of use (means)	Before	N/A	N/A	N/A	N/A	N/A	4.33	3.56	3.95
	After	N/A	N/A	N/A	N/A	3.06	4.78	3.22	4.00
Overall evaluation (means)	Before	N/A	N/A	N/A	N/A	N/A	2.33	2.67	2.50
	After	N/A	N/A	N/A	N/A	2.67	4.00	2.67	3.34
Time for adoption (hours)		16	13	9	38	17	15	10	42

The enhanced release generated the 153% of positive UX aspects and the 123% of negative ones of the old release. A similar situation occurred for the gaps, where the enhanced release found more positive and negative gaps than the old one, considering

both each section separately and in general. The mean values related to the goals as well as the numbers of expected meanings are similar for the two releases, in general. The numbers of expected emotions in the enhanced release are higher than in the old one, considering both each section separately and in general. Changes of meanings decreased from the old (10) to the enhanced (7) release and the same occurred for the changes of emotions (34 vs. 15). Among the general mean values of the data collected thanks to the statements, the instrumental product qualities show the highest mean values before and after the real interaction (5.50 and 5.53, respectively) while the overall evaluation shows the lowest mean values before and after the real interaction (2.50 and 3.34, respectively); the non-instrumental product qualities, emotions and consequences of use report quite low mean values, all of them in the range [3.42, 4.00]. Finally, the time to adopt the enhanced release was slightly higher (42 h) than that of the old release (38 h).

Qualitatively speaking, the most of the positive and negative UX aspects identified by the enhanced release were highlighted also by the old one. Examples of them are the positive UX aspects referring to the intuitiveness of some menu icons and to the possibility of modeling parts and assemblies in the same environment and the negative UX aspects about the functionality to create square holes, about the setting of the model view and about the dimensioning command. Nevertheless, the enhanced release reported thirty fresh UX aspects; some examples of them are the negative UX aspects referring to the low number of options in the main menu bar and to the missed command to generate construction lines and the positive UX aspect referring to the presence of the local menu in the extrusion functionality. On the other hand, the old release reported only five fresh UX aspects; one example of them, positive, consists of the availability of different file formats to export the model.

5 Discussion

The adoptions in the field of the two releases of the irMMs method and the comparison of the results start assessing the applicability, completeness and effectiveness of the enhanced release. Its adoption occurred without problems and this supports its applicability. Completeness is witnessed by the presence of both quantitative and qualitative evaluations; moreover, these evaluations consider all the UX components (instrumental and non-instrumental product qualities and emotional reactions), as well as the consequences of use and the overall evaluation of the product. Finally, two elements assess the effectiveness of the enhanced release. The first element is the presence of thirty fresh aspects against the five of the old release; the most of them refer to the usefulness, usability and appearance of the product and are mainly due the different number of gaps and to the richer final users' comments. The second element is the capability of the quantitative analysis to identify precisely the weak points of the product features to improve (lower-scoring modules than the others) and the strong points to strengthen (modules showing higher mean values). Clearly, all of this happens because of the integration with the meCUE. The following two reasons seem to be responsible for it. First, the statements seems to push the users towards paying attention to other aspects than those strictly connected to the cognitive compatibility. For example, an RB user

wrote as a final comment "to me, this CAD software package is not suitable for professionals; it is more for education because of the coarse way commands are shown (many commands are hidden), because of the unique way to save models (in the cloud only), etc." As the user reported, this aspect came to his mind thanks to the statements about aesthetics and consequences of use. The second reason supporting the effectiveness of the enhanced release seems to be the ability of the meCUE to encourage users focusing on the positive characteristics of the products under evaluation. For example, an RB user and two RE users reported that they appreciated the icons of the main menu as well as the labels of the local menus of different functionalities because they were clear and easy to interpret without further explanations. One of the RE users also wrote that his considerations were evoked by answering to the statements focused on usability (instrumental product qualities module). None of the users exploiting the old release reported similar reasons in their final comments, nor these comments were so focused on positive characteristics of the product under evaluation.

Currently, this research shows some drawbacks. Some statements appear as too generic against the way the irMMs method conducts the evaluation where all the elements (the need, the request for meanings, user actions and product reactions/feedback, etc.) are customized on the product. All of this makes the user guides uneven; sometimes users asked for changing the statements to make them referring more precisely to the CAD software package because they did not know how to interpret them. Unfortunately, this customization is not allowed by the enhanced release now. The enhanced release does not ask the users the reasons why they give specific scores to the statements. This is a problem because UX aspects can be generated only starting from the reasons that support those scores. Some users just wrote their reasons as final comments but the most of them did not report anything. Therefore, precious pieces of information got lost. The possible benefit of the UX quantification has only been mentioned (weak and strong points identification); there is not any discussion about this. The enhanced release requires more time to conduct the tests and to analyze the data because of the presence of many new activities. Sometimes users were annoyed or almost upset because of this. Finally, the enhanced release has not been validated with adoptions involving different products and users as well as against other UX evaluation methods and tools.

6 Conclusions

The research described in this paper aimed at enhancing the existing irMMs-based UX evaluation method. This was achieved by analyzing its lacks and verifying if the meCUE questionnaire capabilities could overcome these lacks. Once assessed this, the irMMs-based UX evaluation method has been integrated with the meCUE by modifying materials and activities. Finally, a first validation of the enhanced release took place to assess the improvement in terms of applicability, completeness and effectiveness.

Some research perspectives are as follows. The statements in the user guides should be customized against the product under evaluation. All of this would avoid users misunderstanding them during the tests. The reasons why users give specific scores to the statements should be asked and collected in suitable tables or free spaces, all of this

in order to allow generating new UX aspects from them. Possible exploitations of the UX quantification should be deeply addressed because they could give precious hints and suggestions for product redesign. For example, the strong point of the CAD software package concerning the instrumental product qualities module can be strengthened by turning those appreciated images that appear now on menu buttons into short videos showing the correct use of the functionalities. Because of the addition of the statements, the activities of all sections of the irMMs method should be optimized to avoid time-consuming tests. For example, the same statements should not appear in two separate tables because this forces the users to jump back and forth in the user guides to compare their expectations (before the real interaction) to what happens for real (after it). There should be unique tables only, containing both the spaces to fill before and after the real interaction. Users would be more concentrated and the results less affected by annoyance and/or tiredness. Finally, a deeper validation of the enhanced release should take place in order to demonstrate its completeness and effectiveness. This validation must involve different products, situations and users (not all recruited from university courses), as well as other UX evaluation methods and tools, like the Valence method [13] or the Tracking Real-Time User Experience [14].

References

1. Law, E.L.C., Roto, V., Hassenzahl, M., Vermeeren, A.P.O.S., Kort, J.: Understanding, scoping and defining user eXperience: a survey approach. In: Greenberg, S., Hudson, S.E., Hinckley, K., Morris, M.L., Olsen, D.R. (eds.) Conference on Human Factors in Computing Systems, CHI 2009, pp. 719–728. ACM, New York (2009)
2. Park, J., Han, S.H., Kim, H.K., Cho, Y., Park, W.: Developing elements of user experience for mobile phones and services: survey, interview, and observation approaches. Hum. Fact. Ergonom. Manuf. Serv. Ind. **23**(4), 279–293 (2013)
3. International Standard ISO 9241-210.: Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems (2010)
4. Hassenzahl, M.: The hedonic/pragmatic model of user experience. In: Law, E., Vermeeren, A., Hassenzahl, M., Blythe, M. (eds.) Towards a UX Manifesto. COST294-MAUSE Affiliated Workshop 2007, pp. 10–14 (2007)
5. Kort, J., Vermeeren, A.P.O.S., Fokker, J.E.: Conceptualizing and measuring user eXperience. In: Law, E., Vermeeren, A., Hassenzahl, M., Blythe, M. (eds.) Towards a UX Manifesto. COST294-MAUSE Affiliated Workshop 2007, pp. 57–64 (2007)
6. Desmet, P.M.A., Hekkert, P.: Framework of product experience. Int. J. Des. **1**(1), 57–66 (2007)
7. Minge, M., Thüring, M.: The meCUE questionnaire (2.0): meeting five basic requirements for lean and standardized UX assessment. In: Marcus, A., Wang, W. (eds.) DUXU 2018. LNCS, vol. 10918, pp. 451–469. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91797-9_33
8. Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. Hum. Comput. Interact. **19**, 319–349 (2004)
9. Rauschenberger, M., Schrepp, M., Cota, M.P., Olschner, S., Thomaschewski, J.: Efficient measurement of the user experience of interactive products. How to use the user experience questionnaire (UEQ). Example: Spanish Language Version. Int. J. Artif. Intell. Interact. Multimed. **2**(1), 39–45 (2013)

10. Filippi, S., Barattin, D.: Considering users' different knowledge about products to improve a UX evaluation method based on mental models. In: Marcus, A., Wang, W. (eds.) DUXU 2018. LNCS, vol. 10918, pp. 367–378. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91797-9_26
11. Norman, D.: The Design of Everyday Things. Revised and Expanded Edition. Basic Book, New York (2013)
12. Thuring, M., Mahlke, S.: Usability, aesthetics and emotions in human-computer technology interaction. Int. J. Psychol. **42**(4), 253–264 (2007)
13. Burmester, M., Mast, M., Kilian, J., Homans, H.: Valence method for formative evaluation of user experience. In: Halskov, K., Petersen, M.G. (eds.) Designing Interactive Systems Conference 2010, pp. 364–367. ACM, New York (2010)
14. Kim, J.H., Gunn, D.V., Schuh, E., Phillips, B.C., Pagulayan, R.J., Wixon, D.: Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. In: Burnett, M., et al. (eds.) International Conference of Human-Computer Interaction, CHI 2008, pp. 443–451. ACM, New York (2008)