Otto-von-Guericke-Universität MagdeburgMagdeburg
Faculty of Computer Science
Faculty of Business Informatics

Master Thesis

Analysis of Current Usability and User Experience Questionnaires and Creating an Optimized Usability Questionnaire

Shokoufeh Valadi

Winter semester 2020

Examiner and Supervisor:

Prof. Dr. rer. pol. habil. Hans-Knud Arndt

Examiner:

Dr.-Ing. David Broneske

Supervisor:

M.Sc. Hannes Feuersenger

Declaration

I herewith assure that I wrote the present thesis independently, that the thesis has not been partially or fully submitted as graded academic work and that I have used no other means than the ones indicated. I have indicated all parts of the work in which sources are used according to their wording or to their meaning.

I am aware of the fact that violations of copyright can lead to injunctive relief and claims for damages of the author as well as a penalty by the law enforcement agency.

Magdeburg, 9. November 2020

. . . . . .

Shokoufeh Valadi

Abstract

The improvement of user satisfaction with a product is the goal of all product-oriented businesses. Examining the user experience is essential to this process, which evaluates the hedonistic, aesthetic, and pragmatic properties of a product. Also, it studies the products' usability, the user's feelings, emotions, and desires before, after, and while using the product.

Usability and user experience questionnaires are a popular method to collect user experience data. In this thesis, these have been reviewed regarding their general characteristics, constituent factors, and their items' content. The analysis of the items was based on the Component model of User Experience (CUE) by Thüring and Mahlke [1]. It reveals shortcomings in usability questionnaires, such as a lack of consideration of hedonistic and aesthetic properties. Based on these findings, this thesis proposes a new questionnaire design, an extended model of the Post-Study System Usability Questionnaire (PSSUQ). Twelve new items and two new factors have increased to this questionnaire, examining Experiential Hedonic and Experiential Aesthetics qualities.

The evaluation of the Extended-PSSUQ has been done by surveying the Mattermost software users in Otto-von-Guericke-University. All factors' reliability results are above 0.8, indicating good results as standard usability and user experience questionnaire. Also, the concurrent validity of the overall Extended-PSSUQ correlated positively with the sum of the After Scenario Questionnaire (ASQ) ratings ($r(21) = 0.624$, $p = 0.003$).

List of Abbreviations

UX User Experience

ISO International Organization for Standardization

SUMI Software Usability Measurement Inventory

QUIS Questionnaire for User Interaction Satisfaction

PSSUQ Post-Study System Usability Questionnaire

SUS Software Usability Scale

CUSI Computer Usability Satisfaction Inventory

ASQ After Scenario Questionnaire

AttrakDiff2 AttrakDiff2

UEQ User Experience Questionnaire

meCUE Modular Evaluation of Key Components of User Experience

VisAWI Visual Aesthetics of Websites Inventory

SUPR-Q Standardized User Experience Percentile Rank Questionnaire

CUE Component model of User Experience

INQ Instrumental qualities

NIQ Non-instrumental qualities

PCA Principal Components Analysis

List of Figures

List of Tables

Contents

1 Introduction

In recent times, businesses shifted their marketing strategy from a product orientation to one on sales and marketing. It caused them to attach more importance to product usability, which is crucial to consumer experience and behavior [22]. This development impacted all types of businesses, such as agricultural, pharmaceutical, hardware, and software products.

The advancement of artificial intelligence and data analysis methods goes hand-in-hand with this paradigm shift because it helps to improve marketing tactics by processing consumer information. User data can contain activities, interests, desires, tendencies, and product user-friendliness. Knowing customer satisfaction level enhances businesses' performance and decision-making. Also, it facilitates the anticipation of the companies' future plans and customizes business goals.

For the reasons mentioned, collecting consumer information, evaluating, and analyzing that has a great significance. This information is sorted into three categories, depending on who acquired them: first, second, and third-party data. First-party data is the most useful and beneficial type of consumer information because it captures the consumer's opinion and data about the product to be marketed. This personalized collected information is obtained directly from the product's user and reveals product weaknesses from a consumer perspective. By using this data, the product can be adjusted to his demands. This information is usually compiled by web analytic operating systems, customer relationship management systems, and business analysis tools. Once this data is available, it can be sold as second-party data to other businesses that might have a similar product lineup. The third-party data is purchased from the organizations that have prepared the second and first-party data and are categorized according to the product, consumers, their specifications, behavior, and interests. The second and third-party data expand the consumer information pool and facilitate the evaluation of business product goals.

As explained in the previous paragraph, the first-party data plays an important role. And it is divided into quantitative and qualitative categories. The qualitative category is descriptive information that expresses users' views and opinions as users comment or observe a product. Examining this category is not easy because this information is not measurable. Instead its analysis needs text and sentiment analysis tools. The quantitative category of information is numerical, such as the number of consumers of a product, their specifications, and product ranking by users. Data mining, measurement, and numerical data analysis help understand the user better and enhance qualitative feedback.

The most important data collection issue is to examine the purpose, timing, and strategy of data gathering. Depending on the project's needs, data could be collected in

different ways such as surveying, investigating user interaction and behavior, monitoring financial transactions and purchased products, using social media information, and even buying information from other similar businesses. Surveying is one of the most straightforward and trustworthy ways to gather quantitative and qualitative information. Surveys consist of different types of questions and answers. Questionnaires, which are a tool of inquiry through a set of questions, are a subset of surveys. Questionnaires are the least expensive, most practical, fastest ways to collect quantitative and qualitative data. Among the advantages of questionnaires are their comparability and scalability. Also, questionnaires are easy to study and analyze. Consequently, questionnaires are one of the most useful methods of providing information, which will be discussed in this thesis and the next section, as stated in the subsequent section, which will give the thesis's problem statement and structure.

1.1 Problem statement

The previous section introduced the importance of data collection by questionnaires. Even if the questionnaires' concept appears straightforward, selecting the right one regarding their different types, usages, and target users group is complex. Questionnaires collect customer information about usability and user experience. The user experience covers various information, such as aesthetic, hedonistic, and pragmatic properties. It monitors the user's satisfaction and skill of using the product, the user's feeling, and the desire to use it again. Usability criteria are considered a subset of user experience, which mostly elaborates on its pragmatic aspect. The usability questionnaire considers the product's applicability, its impact on the user's activity, and the user performed tasks done by the product under review. The first standard usability questionnaires were designed in late 1980 by Chin et al., 1988 [11], Kirakowski, and Dillon, 1988 [23]. Lewis,1990 [14]. And then, gradually in the mid-1990s by Donald Norman [24], user experience terms based on the user feelings and emotions, which described more user interests and experiences, entered the market. Subsequently, user experience questionnaires were designed to gather more distinct user information.
Each of these questionnaires has strengths and weaknesses. At the same time, they have common factors and differences. Each of them is also suitable for measuring a specific product usage feature and user's opinion about it. Therefore, in this work, reviewing the following research questions and points have priority.

1. What are the popular questionnaires in terms of usability and user experience?
2. What are these questionnaires' use cases?
3. Examining the content of selected questionnaires, whether these questionnaires are used in usability or/and user experience?

4. Study the questionnaires in terms of the item number, item scale, and item style.

5. Examine the length or shortness of the chosen questionnaires regarding the number of items and their effect on the user's response.

6. What are the characteristics and factors of the selected questionnaire?

7. Examine all items of each selected questionnaire and specify which item is in the field of instrumental qualities, non-instrumental qualities, short-term affective Response, and long-term evaluation response?

8. What are the disadvantages of the chosen questionnaires?

In this thesis, the selected questionnaires' analysis is done according to the mentioned points and questions. Also, by studying each of these questionnaires, their strengths and shortcomings are collected. Also, the possibility of making changes in these questionnaires to improve their performance will be examined. The main goals and motivations of these studies include:

1. Based on observed shortcomings in these questionnaires, a new questionnaire will be designed to compensate for these leakages and examine a product in terms of usability and user experience.

2. Studying the performance of the designed questionnaire will be evaluated in terms of objectivity, reliability, and validity compared to other standard questionnaire.

3. Distributing the designed questionnaire among the students of Otto-von-Guericke-University to answer the Mattermost software survey and, as a result, check the performance of the designed questionnaire for data collection.

1.2 Structure of the Thesis

The thesis is organized as follows: Chapter 2 introduces the reader to relevant theory and presents previous work. In chapter 3, the theory and analyzing methods are used to reveal the weaknesses of selected questionnaires which lead to an extended-questionnaire design. The evaluation theories and results of the designed questionnaire are given in chapter 4. Conclusion and future work finalize the thesis with a discussion on the performance of the model. Limitations and ideas for future work are discussed in chapter 5.

2 Background

According to pipeline in figure 1, the background chapter consists of four fundamental parts. Section 2.1 describes the history of the user experience terms. Then, section 2.2 explains the terms *usability* and *user experience* according to ISO9241 and describes the relationship between these two terms. Then, the standard questionnaire and its assessment will be discussed in section 2.3. Finally, section 2.4 outlines the most popular questionnaires comprehensively in usability and user experience terms.

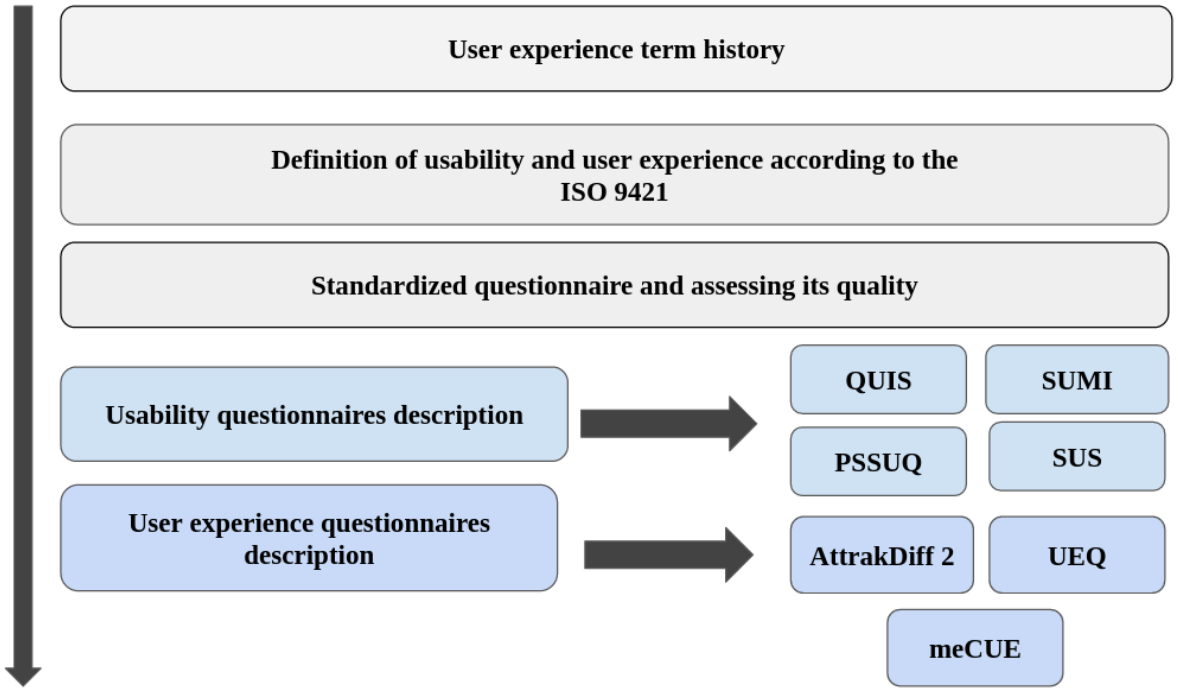


Figure 1: Background chapter pipeline

2.1 User Experience term history

Donald Norman introduced the term user experience in the mid-1990s. He used this term to differentiate between products which might be technically very similar but offer a different emotional experience [24].

One of the influential factors in user experience development was the growth of mobile and software development technologies, which shifted engineers from focusing on human-computer interaction to study human activities. It means that in addition to product usability, user needs and user experience also became important. In general, user experience describes the emotion, interest, and enthusiasm of the user [25] while using the product. Since the year 2000, user experience has received more attention in website design, which consists of usability, branding, and design activities based on aesthetic features. It was considered essential to provide an efficient and easy-to-use user interface to websites. Furthermore, it urged developers to design a website that

works abstractly and looks attractive [26]. Hence, the user experience exceeds the usability criteria by taking the user's feelings and motivation into account in addition to the product's performance.

2.2 Definition of usability and user experience according to the ISO 9421

So far, various definitions of usability have been provided. Each of these definitions has different descriptive parameters. In any case, usability is a fuzzy concept, and the process of defining it is still ongoing, which indicates the ambiguity and incompleteness of the usability description. The International Organization for Standardization (ISO) definition of usability in recent years is as follows:

Usability according to the ISO 9241-11: "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [27]. This definition refers to products and services provided to the customer. It also provides about product for usage research and development, evaluating its differences and similarities, marketing, and sales market information. Moreover, its focus is on the three characteristics effectiveness, efficiency, and satisfaction.

- Effectiveness: "Accuracy and completeness with which users achieve specified goals."
- Efficiency: "The effort invested in relation to accuracy and completeness with which the user achieves a certain goal."
- Satisfaction: "Extent to which the user's physical, cognitive and emotional responses that result from the use of a system, product or service meet the user's needs and expectations." [27]

User Experience according to the ISO 9241-210: The market is being saturated with a huge number of products and services and the user experience is the key to stick out the mass. Despite the great importance of the user experience, little fundamental research has been done on evaluation and measurement, which is still being explored and developed. Notwithstanding the lack of research, ISO has a comprehensive definition of user experience, which is generally true in all areas of services, products, and systems, which is cited as follows:
"A person's perceptions and responses that result from the use or anticipated use of a product, system or service" [28]. This definition refers to all "the feelings, thoughts, beliefs, ideas, perceptions, observations, responses, interactions, preferences, and patterns of physical and psychological behavior of the user before, after, and during usage

of the product or service[1]."

To give some details about subjective perception, all user feedback on the use of the product or service is influenced by various factors, including the brand, how it has been introduced and presented, and the system's performance and efficiency. This feedback also varies with the user's skill, expertise and his experience with previous or similar systems. Also, opinions about a product might change over time. The user may have a negative user experience before or during the product's initial use but may have a better user experience after a while and become more proficient. To obtain information and analyze the user experience, a dynamic, frequent, and variable review of the user himself is required.

Relation between usability and user experience: While usability evaluates the product's performance only during usage, user experience additionally looks into the time period before and after usage. For example, when the user does not have the product at hand but feels like using it. This would be called usage anticipation, which is one aspect of user experience. On the other hand, using the product might affect the user's emotion until after the usage, e.g., when the user builds an emotional connection with the product. The relation between usability and user experience is shown in figure 2. Usability criteria can be used to evaluate some aspects of user experience, but that does not mean that user experience is a good replacement for usability. Both are essential terms and should be considered together because an improper product design with low usability will always spoil the user experience. In the end, usability is what defines the user experience.

[1]https://www.iso.org/obp/ui/iso:std:iso:9241:-210:ed-1:v1:en

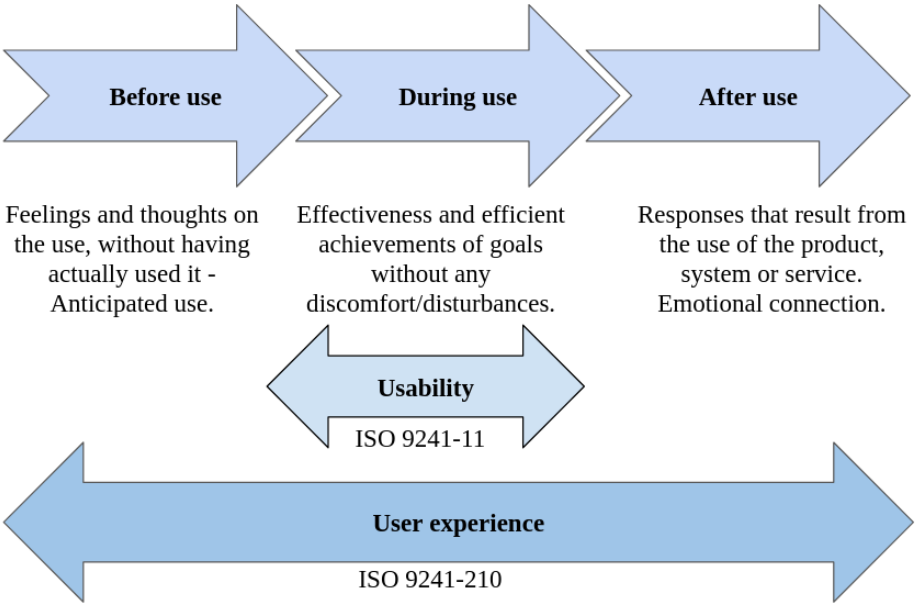


Figure 2: Relationship between usability and user experience [2]

2.3 Standardized questionnaire and assessing its quality

Various methods, including the questionnaire, can measure the user experience and usability of a product. A questionnaire is a designed form to obtain information from the consumer of a product or system. Also, a standard questionnaire is constructed for repeated use. It has a special order and consists of three important parts:

- Rating Scale:
  The set of response points that the respondent gives according to his preference.
- Item:
  The question, contrary adjectives or statement which is followed by a rating scale is defined as an item or element.
- Factor:
  When a set of items measures the same variable reliably and validly, they are named as a factor or scale.

The labeled example of the standard questionnaire is depicted in figure 3.

[2]https://www.uxbooth.com/articles/designing-usability-standards/

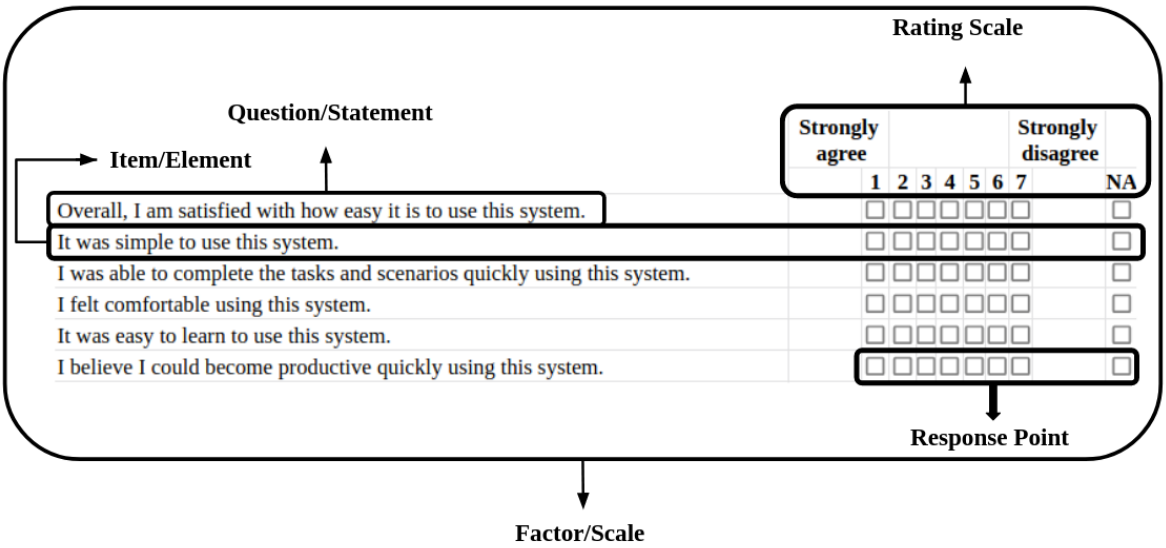


Figure 3: Labeled example of the standard questionnaire

Assessing the standardized questionnaire quality: There are three essential evaluation criteria for questionnaires, which build on each other.

- Objectivity: It has an essential role in evaluating the quality of the questionnaire. The questionnaire reliability and validity evaluation without considering the objectivity criteria is not possible. The test taker's opinion and prejudice should not affect the questionnaire data collection procedure and the interpretation of its results. In this case, the result of the questionnaire will be unbiased and true.
- Reliability: The first condition for checking reliability, or consistency of measurement, is that the questionnaire has objectivity. A reliable test always has the same result, even if it is repeated many times. There are several ways to calculate reliability, such as test-retest and split-half reliability. But the most common evaluation method is the coefficient alpha or Cronbach's alpha. Coefficient alpha can range from 0 (no reliability) to 1 (highest reliability).
- Validity: This test assesses that if the questionnaire measures what it claims. It checks if the measurements are consistent with those of other studies in that field. The first condition for measuring validity is that the questionnaire has objectivity and reliability quality. The most common method of measuring the validity is the Pearson correlation coefficient.

2.4 Usability and User Experience questionnaires

As mentioned, questionnaires are very important because they obtain information directly and without intermediaries from the user. In this part, question 1 of the problem statement section 1.1 will be answered.

Usability questionnaires: Initially, usability questionnaires were designed by industry companies to increase the use of products. One of the pioneers in developing this type of questionnaire is the company IBM, which started its activity in this field in 1979. With the advancement of information technology, research institutes and companies have designed other usability questionnaires. Sauro and Lewis [4] cited the most common ones according to ANSI [29] and ISO1998 [27].

- Questionnaire for User Interface Satisfaction (QUIS) [11]
- Software Usability Measurement Inventory(SUMI) [12] [13]
- Post-Study System Usability Questionnaire (PSSUQ) [14] [15] [16]
- System Usability Scale(SUS) [5]

As a side note, ANSI serves as a common indicator for institutions and companies that study the usability field. It provides its clients with methods, study results, tests, and experience in the field of usability, especially in effectiveness, efficiency, and satisfaction scopes.
Table 1 summarizes the general information of the selected usability questionnaires, including the year of design, their designers, and the translated languages of these questionnaires.

User Experience questionnaires: The user experience field is populated with a wealth of questionnaires, but [30], [31], [32] and [33] name three useful and popular questionnaires, which are:

- AttrakDiff [17]
- User Experience Questionnaire(UEQ) [18]
- Modular Evaluation of Key Components of User Experience(meCUE) [19]

In papers [32] and [31], these three questionnaires are among the most widely used in the field of standardized user experience questionnaires. Source [30] also describes these questionnaires' characteristics: the number of items, their factors, their application, and evaluation as a common standard questionnaire. However, none of them has explained how they achieved their popularity unlike [34], which applied for a systematic literature review in the field of user experience of these three questionnaires in 2018. According to this article, 946 articles from digital databases in the field of user experience were reviewed. In the meantime, 553 articles, which were reviewed in detail, used at least one of these three questionnaires. Among them, 341 articles used AttrakDiff (61.6%), and UEQ and meCUE were used in 200 (36.2%) and 12 (2.2%) articles, respectively. It should be noted that AttrakDiff, UEQ, and meCUE were each introduced in 2003, 2008, and 2013, respectively, and one factor that makes AttrakDiff more useful is that this questionnaire was presented earlier than the other two

Questionnaire	Author(s)	Available languages
Usability		
Questionnaire for User Interface Satisfaction (QUIS)	Chin et al. 1988	English, German, Italian, Portuguese(Brazilian), Spanish
Software Usability Measurement Inventory (SUMI)	Kirakowski and Corbett 1993	Dutch, English, Finnish, French, German, Greek, Italian, Norwegian, Polish, Portuguese, Swedish, Spanish
Post Study System Usability Questionnaire (PSSUQ)	Lewis 1995	English,Portuguese, Turkish
System Usability Scale (SUS)	Brooke 1986	Spanish, French, Dutch, Portuguese,Persian, Slovenian, German, Indonesian
User Experience		
AttrakDiff(AttrakDiff)	Hassenzahl et al. 2001	English, German
User Experience Questionnaire (UEQ)	Laugwitz et al. 2008	German, English, Spanish, Portuguese, Turkish, Indonesian, Chinese, French, Italian, Japanese, Dutch, Russian, Estonian, Slovenian, Swedish, Polish, Greek, Hindi, Persian, Marathi, Tamil, Arabic, Bosnian, Croatian, Finnish, Hungarian ,Norwegian, Slovak
Modular Evaluation of Key Components of User Experience (meCUE)	Minge and Riedel 2013	English, German

Table 1: Overview of usability and user experience questionnaires

questionnaires. Refer to Table 1 for general information of selected user experience questionnaires.
Visual Aesthetics of Websites Inventory (VisAWI), and Standardized User Experience Percentile Rank Questionnaire (SUPR-Q) are other popular example of user experience questionnaires [9], which examine the user experience quality of websites. Although these questionnaires are widely used, they were not reviewed in this thesis because they have been personalized only to analyze websites' user experience and do not survey other products. Considering them in this thesis does not lead to a comprehensive and usable result for surveying various products. And, it is beyond the main goals of this thesis.
The next sections give a brief overview of selected questionnaires.

2.4.1 Questionnaire for User Interface Satisfaction (QUIS)

QUIS was designed in 1988 by a group of university researchers at the University of Maryland. This questionnaire has been created to evaluate the satisfaction of users in the field of human-computer interface. The first version of QUIS was handed out as a short and a long version. The first long version has 90 items, consisting of 5 items assigned to the system's overall evaluation, and the remaining 85 items were assigned to 20 different factors. But the short version has 20 items, which includes five overall items. Each question has a nine rating scale from left to right, starting at one and ending at nine. This rating scale goes from the most negative to the most positive for each attribute (e.g., difficult - easy), and each item also has a "not applicable option" [11]. The overall reliability of this questionnaire is 0.94.
QUIS 7.0 is the latest version of this questionnaire, which examines demographic characteristics and collects user background information. It also measures and evaluates "11 specific interface factors, including screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, online tutorials, multimedia, voice recognition, virtual environments, internet access, and software installation." [3]. QUIS 7.0 assesses the overall level of user satisfaction from the interface perspective. It also analyzes each interface by a specific section according to the user's interest with a 9 rating scale. The QUIS 7.0 also includes short and long versions, a short version with 41 items and a long version with 122 items. The example of the QUIS 7.0 items is illustrated in figure 4.

[3]https://isr.umd.edu/news/story/quis-questionnaire-for-user-interaction-satisfaction-70-isr-ip

Screen		1	2	3	4	5	6	7	8	9		NA
Characters on the computer screen	hard to read				x						easy to read	
Highlighting on the screen	unhelpful			x							helpful	
Screen layouts were helpful	never					x					always	
Sequence of screens	confusing									x	clear	
Terminology and System Information		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
Use of terminology throughout system	inconsistent			x							consistent	
Terminology relates well to the work you are doing?	always					x					never	
Position of messages on screen	inconsistent				x						consistent	
Messages which appear on screen	confusing				x						clear	
Computer keeps you informed about what it is doing	never							x			always	
Error messages	unhelpful									x	helpful	
Learning		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**

Figure 4: Sample QUIS 7.0 items [2]

QUIS version 7.0 is accessible in English, German, Italian, Portuguese (Brazilian), and Spanish. The license of this questionnaire must be purchased from Maryland university. Prices vary for students, researchers, and business use.

2.4.2 Software Usability Measurement Inventory (SUMI)

The SUMI questionnaire was founded by Jurek Kirakowski at Human Factors Research Group in the University of Cork in Ireland based on the Computer Usability Satisfaction Inventory (CUSI) questionnaire. The CUSI was replaced by the SUMI questionnaire in 1990. This questionnaire is designed to examine the perceived quality of the final user experience. The SUMI questionnaire consists of 50 items whose global scale is based on 25 elements and five sub-scales, including efficiency, affect, helpfulness, control, and learnability. Each of these five sub-scales contains ten items, and each item has 3 rating scales. These rating scales are agree, undecided, and disagree. This questionnaire consists of a set of conceptually positive and negative statements. And it has an overall reliability of 0.94. The SUMI questionnaire sample is shown in figure 5.

Statements 1 - 10 of 50.	Agree	Undecided	Disagree
This software responds too slowly to inputs.	○	○	○
I would recommend this software to my colleagues.	○	○	○
The instructions and prompts are helpful.	○	○	○
This software has at some time stopped unexpectedly.	○	○	○
Learning to operate this software initially is full of problems.	○	○	○

Figure 5: Sample QUIS items [3]

The SUMI is translated into 12 languages, including Dutch, English, Finnish, French, German, Greek, Italian, Norwegian, Polish, Portuguese, Swedish, and Spanish. It is

also available for online, offline, and student use. The licensed version of this questionnaire must be purchased from Cork University.

2.4.3 Post Study System Usability Questionnaire (PSSUQ)

The PSSUQ questionnaire is designed to measure and understand the user's satisfaction with the computer system, product, and application. The questionnaire was based on an internal project from IBM called $SUMS$ (System Usability MetricS) in 1988. $SUMS$ was originally comprised of a set of items. IBM human factors group reviewed these items and applied some content changes, which led to the design of the first version of the PSSUQ questionnaire with 18 items [14].
Version 3 of the PSSUQ is the latest updated version. This questionnaire has 16 items, and each of these items is examined from two strongly agree and strongly disagree points on a 7 rating scale, in addition to the "not applicable option." PSSUQ version 3 is depicted in figure 6. The PSSUQ items consist of four scales. These four scales and their reliability are presented as following [16]:

- Overall: Average the responses for Items 1 through 16 (all the items), with the reliability of 0.94.
- System Quality (SysQual): Average Items 1 through 6, with the reliability of 0.9.
- Information Quality (InfoQual): Average Items 7 through 12, with a reliability of 0.91.
- Interface Quality (IntQual): Average Items 13 through 15, with the reliability of 0.83.

	The Post-Study Usability Questionnaire Version 3		Strongly agree						Strongly disagree		NA
			1	2	3	4	5	6	7		
1	Overall, I am satisfied with how easy it is to use this system.		O	O	O	O	O	O	O		O
2	It was simple to use this system.		O	O	O	O	O	O	O		O
3	I was able to complete the tasks and scenarios quickly using this system.		O	O	O	O	O	O	O		O
4	I felt comfortable using this system.		O	O	O	O	O	O	O		O
5	It was easy to learn to use this system.		O	O	O	O	O	O	O		O
6	I believe I could become productive quickly using this system.		O	O	O	O	O	O	O		O
7	The system gave error messages that clearly told me how to fix problems.		O	O	O	O	O	O	O		O
8	Whenever I made a mistake using the system, I could recover easily and quickly.		O	O	O	O	O	O	O		O
9	The information (such as online help, on-screen messages and other documentation) provided with this system was clear.		O	O	O	O	O	O	O		O
10	It was easy to find the information I needed.		O	O	O	O	O	O	O		O
11	The information was effective in helping me complete the tasks and scenarios.		O	O	O	O	O	O	O		O
12	The organization of information on the system screens was clear.		O	O	O	O	O	O	O		O
13	The interface* of this system was pleasant.		O	O	O	O	O	O	O		O
14	I liked using the interface of this system.		O	O	O	O	O	O	O		O
15	This system has all the functions and capabilities I expect it to have.		O	O	O	O	O	O	O		O
16	Overall, I am satisfied with this system.		O	O	O	O	O	O	O		O

**The "interface" includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens (including their graphics and language).*

Figure 6: Sample PSSUQ version 3 items [4]

Using the PSSUQ questionnaire is free. Only the reference source should be mentioned.

2.4.4 System Usability Scale (SUS)

The SUS questionnaire was designed in 1986 by Brooke [5] to assess the user's satisfaction with the usability of a product or service. This questionnaire has ten items with a 5 rating scale of measurement between the two options strongly disagree and strongly agree. The strongly disagree option is marked on the left with domain one and the strongly agree on the right with domain five. The odd items in this questionnaire have a positive semantic text, and even items are statements with negative content.
One of the most important points in answering this questionnaire is the speed of responding to the items of this questionnaire by the respondent. After reading each element, the respondents should give their immediate answer and not think deeply about each item. On the other hand, the respondent must answer all the items and, if he does not have a definite answer for each item, pick 3 in the middle of the item scale. The SUS questionnaire sample is displayed in figure 7.

	The System Usability Scale Standard Version		Strongly Disagree				Strongly Agree
			1	2	3	4	5
1	I think that I would like to use this system frequently.		O	O	O	O	O
2	I found the system unnecessarily complex.		O	O	O	O	O
3	I thought the system was easy to use.		O	O	O	O	O
4	I think that I would need the support of a technical person to be able to use this system.		O	O	O	O	O
5	I found the various functions in this system were well integrated.		O	O	O	O	O
6	I thought there was too much inconsistency in this system.		O	O	O	O	O
7	I would imagine that most people would learn to use this system very quickly.		O	O	O	O	O
8	I found the system very awkward to use.		O	O	O	O	O
9	I felt very confident using the system.		O	O	O	O	O
10	I needed to learn a lot of things before I could get going with this system.		O	O	O	O	O

Figure 7: Sample SUS items [5]

The SUS questionnaire is free to use. Only its reference source must be indicated.

2.4.5 AttrakDiff (AttrakDiff)

The AttrakDiff1 was designed in 2001 by Marc Hassenzahl. This questionnaire is based on a combination of both hedonic and pragmatic qualities. According to AttrakDiff1, the product should not only satisfy the consumer but also makes him happy. For each of the questionnaire's 23 items, the rating scale ranges between a pair of contrary attributes, such as complicated - simple or tacky - stylish. These items examine the product from three points of view: pragmatic quality, hedonic quality, and attractiveness. And it estimates the outcome as an overall attractiveness that results from the evaluation of the three mentioned qualities [35]. AttrakDiff1 questionnaire's hedonic factor includes two sub-factors: hedonic quality - identity and hedonic quality - stimulation. During evaluation, these two sub-factors were treated as one. However, based on the tests and studies performed on these two sub-factors, it was found that evaluating these sub-factors separately will lead to more reliable results [17]. Therefore, AttrakDiff2 was designed with 28 items and by separating the items of hedonic quality - identity and hedonic quality - stimulation sub-factors. These qualities and their subsets are described as follow:

1. Pragmatic quality

   When a product or service has a pragmatic quality, it serves it's purpose ef-

fectively and efficiently. In other words, the product's pragmatism means that the product has reached its ultimate goals to be useful, usable, and gives predictable results. Other pragmatic qualities are to be clear, supportive, and controllable [35].

2. Hedonic quality
   In terms of hedonic quality, the product should be useful and bring joy and entertainment. As a result, hedonic quality is divided into two aspects:
   - Hedonic quality - identity
     The design of the product establishes a special visual connection with the user. Features of this quality include a sense of connection, expertise, and closeness [35].
   - Hedonic quality - stimulation
     The product must be exciting, significant and motivate the user. It should also be tailored to the user's skills and knowledge to evoke a sense of satisfaction. Other features of this quality are creative, original and challenging [35].

3. Attractiveness
   The attractiveness of the product is based on the positive and negative statements collected. In this context, good and pleasant attributes can be mentioned.

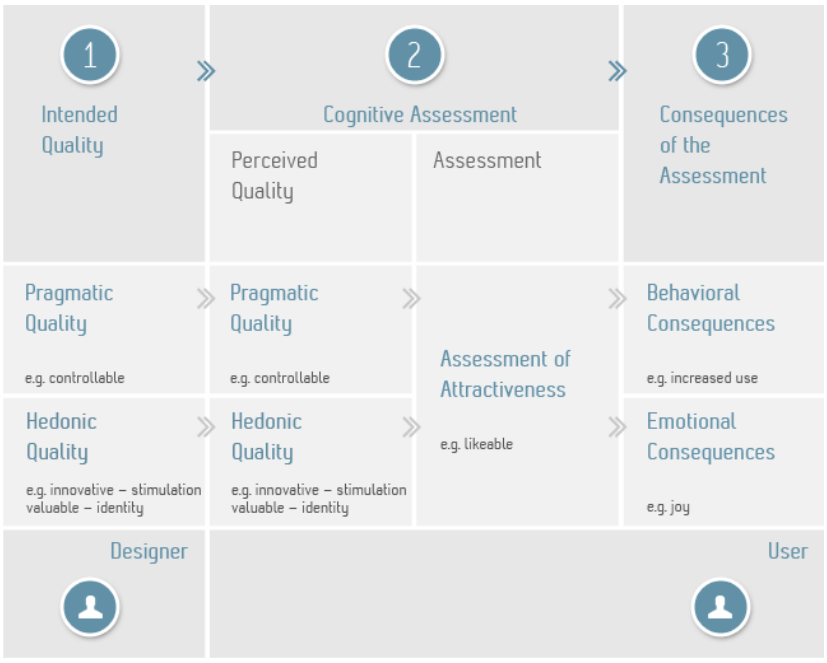


Figure 8: The pragmatic and hedonistic effect on the inner mental understanding of the product's attractiveness to the user and ultimately increases the use of the product and the user's happiness. [4]

[4] http://www.attrakdiff.de

The AttrakDiff2 questionnaire uses these four factors to evaluate the user experience of a product. The relationship and the result of their correlation are shown in figure 8. It is also possible to use AttrakDiff2 for free and without paying any fee, although it is necessary to mention the source. The AttrakDiff2 questionnaire sample is depicted in figure 9.

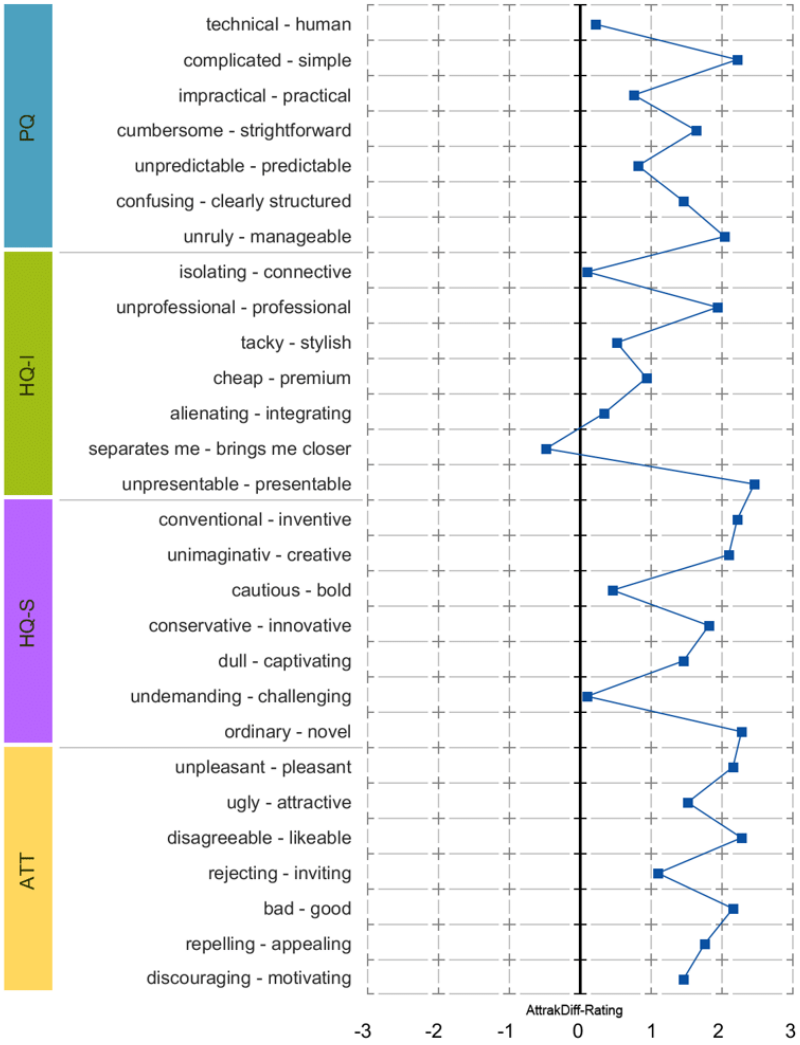


Figure 9: Sample AttrakDiff2 of items [6]

2.4.6 User Experience Questionnaire (UEQ)

The UEQ questionnaire was designed in 2005 to fast and quickly assess the user experience. This questionnaire has 26 items; each item includes a pair of contrary adjectives. Each item has a rating scale between -3 and 3 and assesses pragmatic and hedonic qualities. The most negative choice for an item is on the left, and the most positive is on the right. These two qualities (pragmatic and hedonic) were explained in the previous section 2.4.5. The difference is that the UEQ's pragmatic quality examines three sub-factors, including perspicuity, efficiency, and dependability. On the other hand, the hedonic quality is evaluated two sub-factors, including stimulation and novelty.

This section will give a brief explanation of the hedonic and pragmatic factors studied in the UEQ. Figure 10 shows the pragmatic and hedonic quality characteristics.

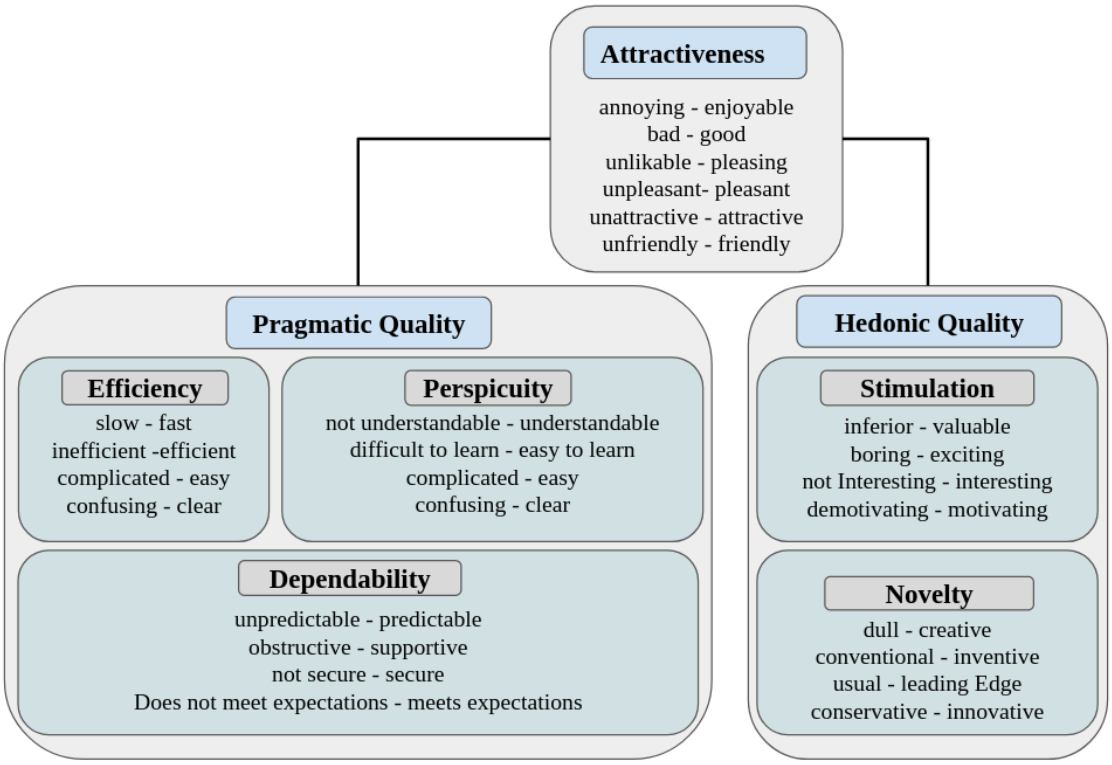


Figure 10: UEQ items with attributes

1. Attractiveness: "Overall impression of the product. Do users like or dislike it?"
2. Perspicuity: "Is it easy to get familiar with the product and to learn how to use it?"
3. Efficiency: "Can users solve their tasks without unnecessary effort? Does it react fast?"
4. Dependability: "Does the user feel in control of the interaction? Is it secure and predictable?"
5. Stimulation: "Is it exciting and motivating to use the product? Is it fun to use?"
6. Novelty: "Is the design of the product creative? Does it catch the interest of users?" [5]

[5]https://www.ueq-online.org/

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

Figure 11: Sample UEQ items [7]

UEQ has been translated into 20 languages and is free to use, and should only be cited. The sample English version of the UEQ has been shown in figure 11.

2.4.7 Modular Evaluation of Key Components of User Experience(meCUE)

The meCUE questionnaire was designed by Minge and Riedel in 2013 [19] to explore a holistic view of the user experience. Compared to other questionnaires, it examines the user's feelings, product utility features, and visual appeal of the product. The meCUE is based on the analytical CUE model by Thüring and Mahlke in 2007 [1]. The CUE model was developed by examining and understanding the instrumental and non-instrumental quality of the product and its relationship with the user's emotions. It acts as an interface between product evaluation and examining the user's perceptual implications for product utility, such as overall judgment, acceptance, and intention to use [8]. Figure 12 shows the components of the CUE model.

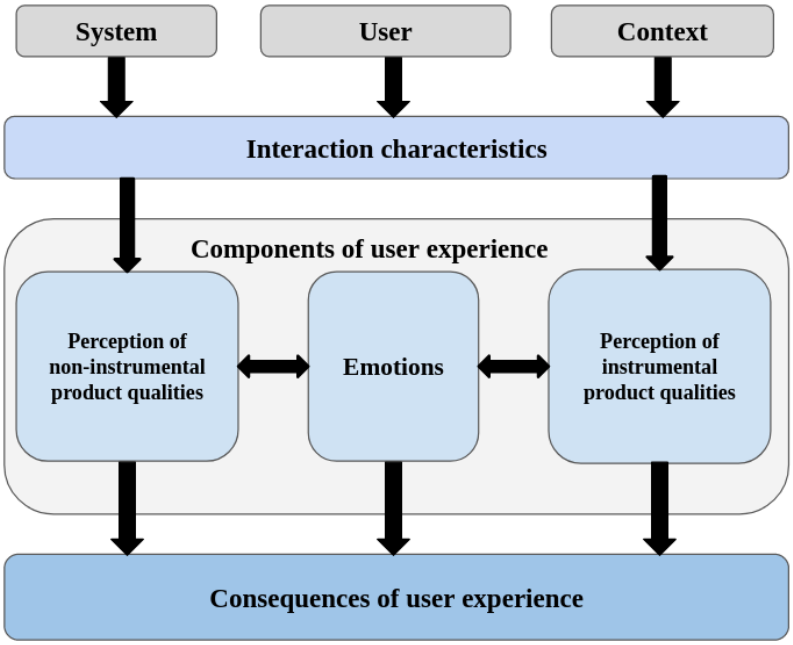


Figure 12: Components of User Experience (CUE model) [1]

The meCUE questionnaire follows the structure and components of the CUE model. This questionnaire is comprehensive compared to other questionnaires, and it is designed as three modules with separate evaluations. The structure of meCUE derived from the CUE model is shown in figure 13. The meCUE adopted these science-based modules to meet its research purpose. Module one examines product perception with both instrumental quality and non-instrumental quality. Instrumental quality includes perceived usefulness and perceived usability factors, and non-instrumental quality includes three factors: visual aesthetics, status, and commitment. The second module, called Emotions, checks the user's positive and negative emotions. Finally, the third module is labeled consequences and evaluates the results of using the product according to future use goals.

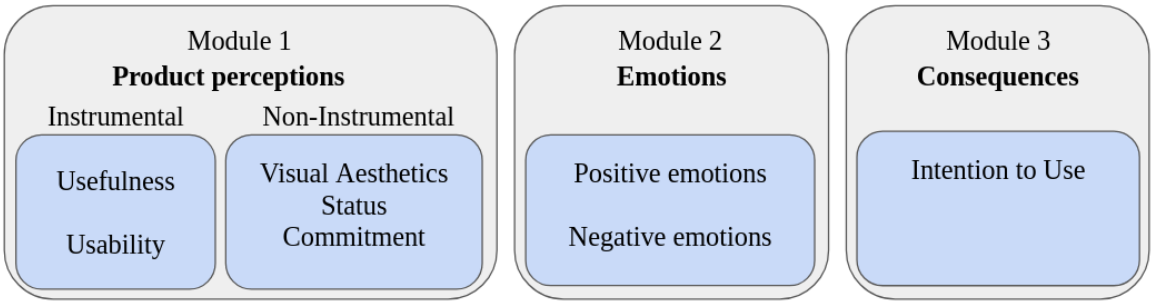


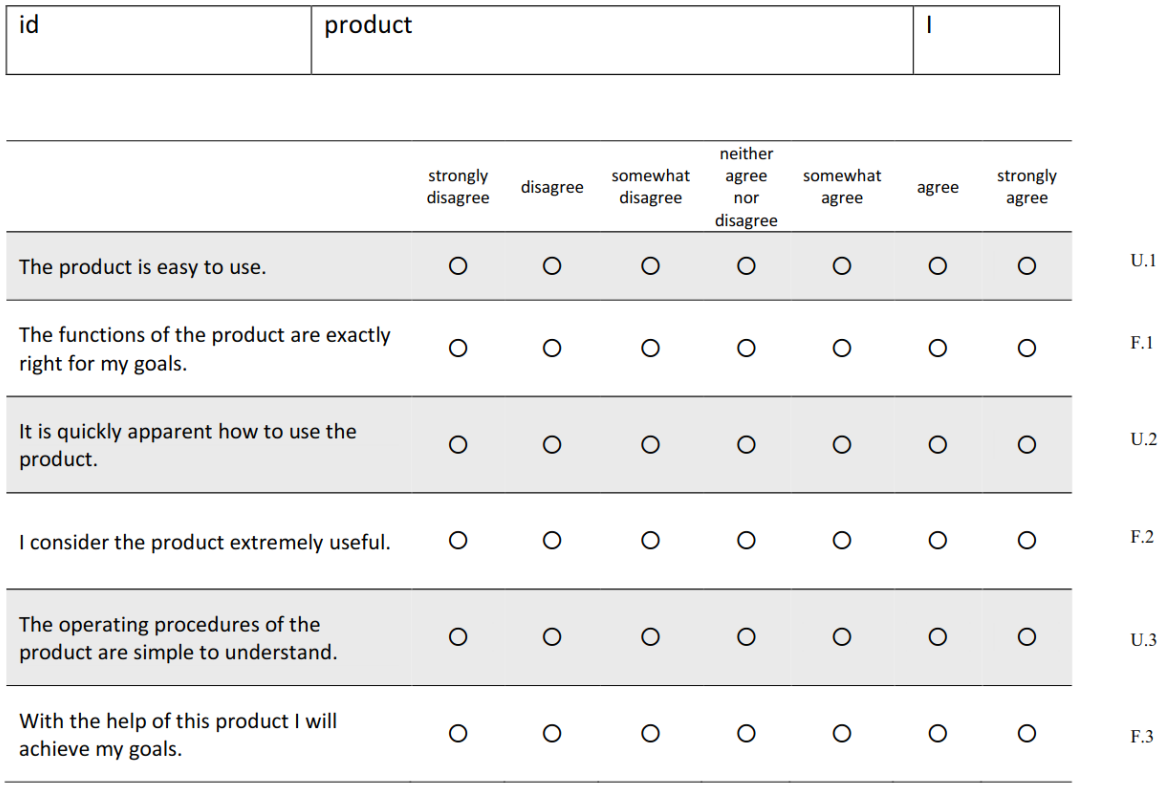
Figure 13: Structure of meCUE derived from the CUE model [8].

The meCUE has 34 items, and each item is measured on 7 rating scales, which starts with strongly agree and ends with strongly disagree. Finally, the items are sorted into

modular dimensions or factors, namely:

- Module I: Usefulness (F), Usability (U)
- Module II: Visual aesthetics (A), Status (S), Commitment (C)
- Module III: Positive Emotions (PA, PD), Negative Emotions (NA, ND)
- Module IV: Intention to use (IN), Product loyalty (L)
- Module V: Overall evaluation

Figure 14 shows the meCUE items according to its modular dimensions.



id	product	I

	strongly disagree	disagree	somewhat disagree	neither agree nor disagree	somewhat agree	agree	strongly agree	
The product is easy to use.	○	○	○	○	○	○	○	U.1
The functions of the product are exactly right for my goals.	○	○	○	○	○	○	○	F.1
It is quickly apparent how to use the product.	○	○	○	○	○	○	○	U.2
I consider the product extremely useful.	○	○	○	○	○	○	○	F.2
The operating procedures of the product are simple to understand.	○	○	○	○	○	○	○	U.3
With the help of this product I will achieve my goals.	○	○	○	○	○	○	○	F.3

Figure 14: Modules of meCUE questionnaire [8]

3 Methods and model development

The last chapter gave a review of popular questionnaires. These will be extensively analyzed in first part of this chapter and a new extended questionnaire will be proposed in the second one. The chart in figure 15 gives a more detailed overview of this chapter. Section 3.1 describes selected questionnaires from three different perspectives. First, these questionnaires are reviewed from general aspects, including item numbers, item styles, rating scales, and questionnaire use-cases. Then, in section 3.1.2, these questionnaires' factors and their functions are examined based on the constituent factors. Then, in part 3.1.3, the questionnaires' items are analyzed based on the CUE model, and the weaknesses and disadvantages of each questionnaire are examined. In section 3.2, which includes three sections, discusses the extended questionnaire and explains the purpose of extending the PSSUQ questionnaire, as shown in section 3.2.1. Finally, section 3.2.2 and 3.2.3 discuss the items and content added to the PSSUQ questionnaire.

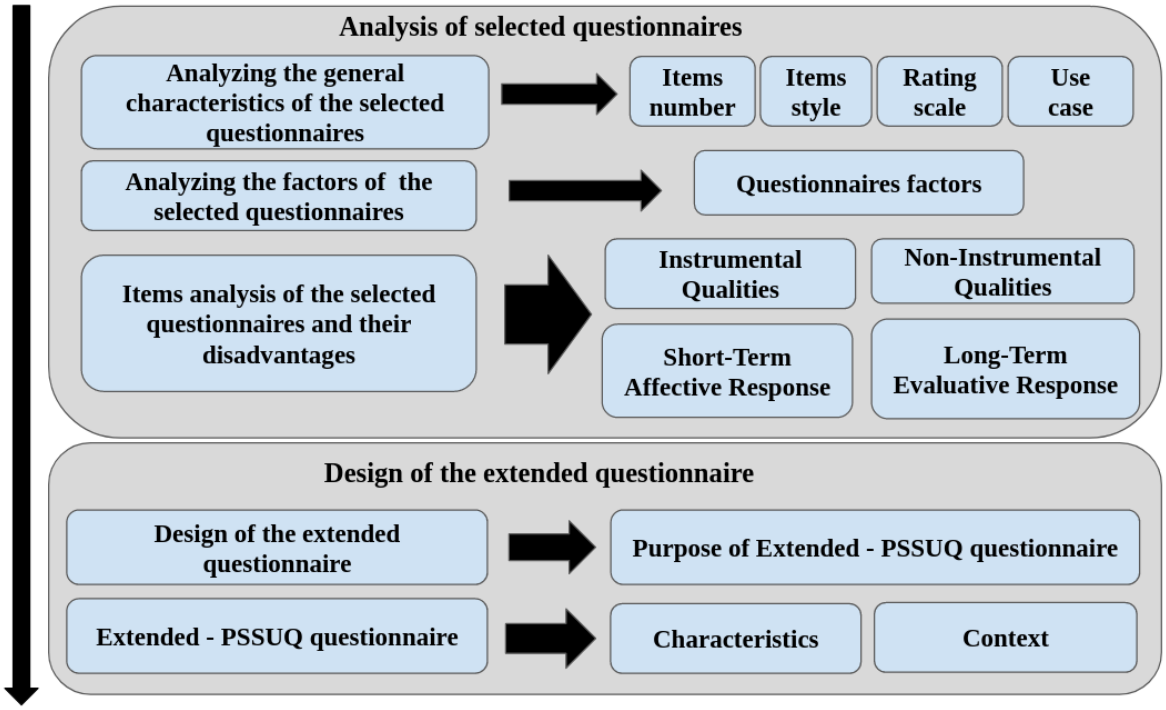


Figure 15: Method and model development pipeline

3.1 Analysis of selected questionnaires

In section 3.1.1, selected questionnaires' characteristics or their general structure are examined. In this section, based on problem statement 1.1, questions 2, 4, 5 are answered. Question 6 is then discussed in the factor analysis section 3.1.2. Finally, questions 3, 7, and 8 are answered in the item analysis section 3.1.3.

3.1.1 Analyzing the general characteristic of the selected questionnaires

In the previous chapter, the ISO definitions for the product properties usability and user experience have been introduced. In addition, popular questionnaires examining each of these were discussed. In this section, as shown in table 2, the general characteristics of these selected questionnaires are reviewed, including:

- Number of items
- The item styles (attributes and/or sentence)
- Rating scales
- Questionnaire's use cases

Questionnaire	Number of items	Item style	Rating scale	Use case
QUIS 7.0	Short version 41 Long version 121	Sentence	10 point semantic scale	Product, computer software
SUMI	50	Sentence	3 point dichotomous scale	Software product
PSSUQ 3.0	16	Sentence	7 point likert scale	Website, software applications, system, product
SUS	10	Sentence	5 point likert scale	System, computer software, product
AttrakDiff2	28	Attributes	7 point semantic scale	Interactive product
UEQ	26	Attributes	7 point semantic scale	Interactive product
meCUE	34	Sentence	7 point likert scale	Interactive product

Table 2: General characteristics of seven selected questionnaires

According to table 2, QUIS short/long version and SUMI respectively are the longest questionnaires. The SUS questionnaire has the lowest item number with ten items. The rest of the selected questionnaires range between 16 and 34. Generally, large numbers of items violate the guidelines of the BRUSO model [36], which is an acronym for "brief," "relevant," "unambiguous," "specific," and "objective". It says that an applicable and effective questionnaire should be concise and should refrain from any additional

words, sentences and comprehensively omit the unnecessary elements. The conciseness of the questionnaire facilitates its understanding and accelerates its response by the respondent. It will increase the satisfaction and motivation of the respondent to complete the questionnaire.
The item style of the questionnaires in table 2 is either a sentence or an attribute, which depends on the type of rating scales. Since the respondent must choose the appropriate option on the rating scale, all these questionnaires belong to the close-ended items. In the close-ended items, unlike the open-ended items, the respondent is given a list of questions or sentences with a set of possible answers. He has to pick the answer which comes as close as possible to his opinion. Close-ended items collect quantitative data. Open-ended items request the respondent to answer the questions by freely writing it down, and no pre-defined answers are given. The open-ended items are difficult to analyze and require a lot of time because of the respondents' diverse responses. Closed-ended items are used when researchers have accurate information about the respondents' answers. They want to examine defined and specific answers and the degree of agreement or disagreement of their audience about the product or service. Also, according to the *BRUSO* model, closed-ended items make respondents answer the questionnaire faster and easier. Also, it facilitates the interpretation of answers for researchers. The questionnaires, which are studied in this thesis, make use of three different types of rating scales, where each has its specific features. The types are:

- Semantic scale
  Each item of this rating scale consists of two contrary attributes, and the respondent must answer on a scale which ranges between these. Attributes are graded from the most negative semantic concept to the most positive concept or vice versa. An adjective with a negative semantic meaning has a smaller or even a negative numerical value, while the positive attribute has a larger and positive numerical value. This rating varies from questionnaire to questionnaire, but a neutral option is common, such as "Not Applicable". This is so that even respondents who do not agree with these two contrary attributes ratings can reply. The QUIS questionnaire has a rating scale with 9 numbers plus the "Not Applicable" option. The AttrakDiff2 and UEQ questionnaires also have 7 numbers on their scales. The semantic rating scale has many contrasting attributes to express the respondent's feelings and opinions, and respondents can clearly and accurately give their feedback and assessments.
- Dichotomous scale
  This type of rating scale usually consists of two explicit answers: Yes/No or Agree/Disagree. And sometimes, a third option is added to the dichotomous item scale so that respondents can respond to all the items when they do not

have an explicit answer, such as “Do not know” or “Undecided.” The advantages of a dichotomous rating scale are simplicity, clarity, and transparency in response. The SUMI questionnaire with three Agree, Undecided, and Disagree option has three points dichotomous rating scale.

- Likert scale

  This rating scale ranks the respondent’s response to the two strongly agree and strongly disagree options and determines the extent to which the respondent agrees or disagrees with the statement and conveys his general opinion about the item. The likert item scale has 5 or 7 numbers. Usually, this rating scale’s middle option is the neutral choice, or the item has the “Not Applicable” option. The likert rating scale has a simple and understandable analysis. It also does not force the participant to give an explicit and limited answer on the subject like dichotomous scale. But on the other hand, since it always examines the degree of satisfaction and agreement of the respondent about a particular issue, the respondent acts in a one-dimensional way and cannot fully express his feelings. By comparing the likert and the semantic rating scales, the semantic rating scale comprises a wide range of contrary adjectives. It can better express the respondent’s feelings and opinions. The PSSUQ uses the likert scale with 7 numbers, and so does the meCUE. However, while the former has an extra option named “Not Applicable” to express neutrality, the latter provides a middle scale point to serve that purpose. Similarly, the SUS questionnaire has a rating scale with five numbers and a neutral middle point.

All questionnaires in table 2 have been designed to analyze specific services or products. Three user experience questionnaires, AttrakDiff2, UEQ, and meCUE, have been made to analyze interactive products that are interactive in terms of audio and video. For example, home or computer gaming systems or products, computer systems, and multimedia systems that can be connected to a TV or computer. On the other hand, QUIS, SUS, and PSSUQ questionnaires can analyze various products, systems, and software. Additionally, PSSUQ is also applicable to website analysis. Lastly, the SUMI questionnaire’s purpose is to analyze software applications only.

3.1.2 Analyzing the factors of the selected questionnaires

Factors are measured by a set of semantically related items. It depends on the designer and purpose of the questionnaire which factors are to be chosen. For instance, usability and the user experience require different factors to be surveyed. A typical factor to describe usability could be *controllability*, as seen in QUIS or SUMI. For user experience, a quite common factor is *attractiveness*. Generally, factors collect information such as opinions, feelings, and the level of consumer satisfaction with the product or service. Each questionnaire measures the degree of reliability by measuring each of these factors. According to table 3, the factors of each questionnaire are introduced. To better understand and clarify each of these factors' meaning, they are briefly described below.

Appearance/Attractiveness: The attractive and tempting appearance of a product or service for the user is defined as attractiveness. Design, color, and shape, the product's layout should be pleasant to the user and encourage him to use it.

Controllability/Dependability: It evaluates if a user is able to quickly react in case an error occurs, and bring the system back to a stable state. It means that he is aware of the system status and performance and can predict its behavior. Every questionnaire gives this factor a different name and attaches specific meanings to it.
The equivalents to controllability in QUIS and PSSUQ are "Terminology and System Information" and "Information Quality", respectively. These terms describe the system's degree of compatibility with the user's task, system error messages, and system alert to the user to inform the existing error. In general, they evaluate how well the system's information guide the user so he can easily keep control of it. SUMI shortens the name of the controllability factor to "control". It refers to how the user feels about his control over the software he is using. On user experience side, only the UEQ questionnaire employs this factor, which is characterized by "dependability". This dependability feature includes three factors: user control over the system, system security, and predictability of the system.

Efficiency: It means optimum use of resources and removal of unnecessary process steps. The system or product has the highest efficiency, which solves the task in the shortest time and without wasting resources and energy. In the SUMI questionnaire, this factor refers to how the software helps the user to complete his task and evaluates the software transparency. However, in user experience questionnaires, efficiency refers to how the system makes the user most efficient in the shortest possible time. Also, efficiency in this domain includes the lack of complexity, efficiency, and transparency of the system.

Questionnaire	Factors
QUIS 7.0	Overall Reactions to the Software (Overall) Screen (Simplicity) Terminology and System Information(Controllability / Dependability) Learning (Learnability / Perspicuity) System Capabilities (Learnability / Perspicuity)
SUMI	Global(Overall) Learnability (Learnability / Perspicuity) Helpfulness (Helpfulness) Controllability (Controllability / Dependability) Affect (Emotion / Affect) Efficiency (Efficiency)
PSSUQ	Global(Overall) System Usefulness (Helpfulness) Information Quality (Controllability / Dependability) Interface Quality (Learnability / Perspicuity)
SUS	SUS-Scale (Overall)
AttrakDiff2	Attractiveness (Appearance / Attractiveness) Pragmatic Quality (Pragmatic Quality) Identity (Identity) Stimulation(Stimulation)
UEQ	Attractiveness (Appearance / Attractiveness) Perspicuity (Learnability / Perspicuity) Efficiency (Efficiency) Dependability (Controllability / Dependability) Stimulation (Stimulation) Novelty (Novelty)
meCUE	Usability (Effectiveness / Helpfulness) Usefulness (Efficiency) Visual Aesthetics (Appearance / Attractiveness) Status (Identity) Commitment (Loyalty) Positive / Negative Emotions (Emotion / Affect) Product Loyalty (Loyalty) Intention to use (Immersion) Overall Judgment (Overall)

Table 3: Selected usability and user experience questionnaires factors

Emotion/Affect: It evaluates the user's positive and negative emotional reactions to the product. The user's feelings before, during, and after using the product or service vary.

Helpfulness: It refers to the product or service's ability to serve and help the user solve a problem or facilitate his performance. It overlaps with the product's usefulness and profitability. That's why the questionnaires' interpretation of helpfulness differs greatly. According to the SUMI questionnaire, helpfulness refers to the degree of how self-explanatory the software is. Also, the software must have an appropriate help option and sufficient documentation to support the user. The PSSUQ equates helpfulness with the system's usefulness. And the meCUE questionnaire refers to helpfulness as the system's effectiveness.

Identity: The product's visual characteristics, such as color, symbol, and product design, make the product specific and prominent in the user's mind.

Immersion: The product so absorbed the user in terms of interaction and connection that he does not feel the time passing. The product or service has a profound and memorable effect on the user.

Learnability/Perspicuity: The product is designed to easily and quickly get acquainted with the product's instructions and features. And the user can easily communicate with the system and accomplish his tasks through the system.
The QUIS questionnaire's learnability includes easy-to-remember names of components and commands, system alerting messages for the user, and simple additional resources for easier system learning. Also, in the SUMI questionnaire, this factor depends on how quickly the user learns the system or its new features. In the PSSUQ, learning ability also refers to the quality characteristics of the interface. Also, UEQ, as a user experience questionnaire, deals with the simplicity and speed of getting to know the product and learning how it works.

Loyalty: The product and its features have made the user so happy and satisfied that he introduces the product to other people and prefers the product to other similar ones. The user is a regular customer of the product or service, and he consumes and purchases it repeatedly.

Novelty: This quality refers to whether the product is new or innovative. A new product comes with a new idea or concept. Innovative is a product that improves the current product or defines a new generation.

Overall: It is an overall quality of the system, which assesses the product's impact on the user's life and determines whether the whole product has a positive or negative effect. In other words, it is a measurement scale of the total performance of an entire

system or product from the user's perspective.
The QUIS questionnaire's overall factor is the sum of the ratings for the properties wonderful, easy, satisfied, stimulating, and flexible, which evaluate the system from the user point of view. In contrast to that, there are 25 item in the SUMI questionnaire that generally measures the user's overall satisfaction with the software. The SUS questionnaire does not even bother with any factors other than "overall". The PSSUQ considers average of all items as overall factor. Ultimately, the meCUE user experience questionnaire employs an item which directly asks for the user's opinion about the product's general experience.

Pragmatic Quality: Usability and usefulness of the product in achieving the goal of action and performing the task.

Simplicity: A low-cost, fast-acting product that efficiently meets the needs of the user is defined as product simplicity. And this product can be used intuitively and has a low complexity.

Stimulation: The use of the product satisfies the user's and he feels a need to use it. The product completely fulfills his needs, which makes the user convince and encourages him to use the product again.

Table 4 presents the occurrences of factors in selected questionnaires. It shows that controllability/dependability, learnability/perspicuity, and overall are the most commonly used factors. These typically belong to the usability aspect of a product. On the user experience side, which has its own exclusive factors, appearance/attractiveness is the only common one.
Each questionnaire examines a unique factor which distinguishes it from the others. In all of these questionnaires, only QUIS examines the simplicity factor. AttrakDiff2 also studies pragmatic quality directly. Moreover, the two characteristics of loyalty and immersion are used only by the meCUE. Finally, the novelty factor is just studied by the UEQ questionnaire.
According to table 4, the meCUE, a user experience questionnaire, and the SUMI, which is a usability questionnaire, examine eight and six quality factors, respectively. They have the highest number of quality factor reviews compared to other mentioned questionnaires. The SUS questionnaire has only one overall factor. According to Brooke [5], this questionnaire is designed to give a single score, which represents a product's overall usability. As a result, surveyors who work with the SUS questionnaire should not try to evaluate its items separately. Only the overall score of all items is acceptable. Ultimately, PSSUQ examines the lowest number of four usability factors.

Quality Factor	QUIS	SUMI	PSSUQ	SUS	AttrakDiff2	UEQ	meCUE	Occurrences
Controllability/Dependability	×	×	×			×		4
Learnability/Perspicuity	×	×	×			×		4
Helpfulness		×	×				×	3
Appearance/Attractiveness					×	×	×	3
Efficiency		×				×	×	3
Overall	×	×	×	×			×	4
Emotion/Affect		×					×	2
Identity					×		×	2
Loyalty							×	1
Pragmatic Quality					×			1
Simplicity	×							1
Stimulation					×	×		2
Immersion							×	1
Novelty						×		1
Number of Factors	4	6	4	1	4	6	8	

Table 4: Standard usability and user experience questionnaires with their quality factors and number of factor occurrences [9]

Same applies to the short version of QUIS, which otherwise consists of 11 factors in the long version.

3.1.3 Item Analysis of the selected questionnaires and their disadvantages

In the previous sections, selected questionnaires' characteristics and their quality factors have been examined. It has been pointed out that factors generally describe the semantic context of a set of items. For instance, the two items "the product exhilarates me" and "the product relaxes me" both are semantically related because they describe positive emotions, which are a factor in meCUE. Questionnaires define items to investigate the user and his relationship with the product. These could measure aesthetic features and their impact on the user's emotion and motivations, but also the product's instrumental features and usability. This chapter examine the items used in each questionnaire.
Each item consists of a set of words and attributes that convey a specific meaning to the user to evaluate his opinion about the product. Law, van Schaik, and Roto [10] researched the questionnaires' items' concept to find out if the user experience was measurable or not. This research included a questionnaire and an interview that lead to a categorization of these attributes. This classification is based on the the CUE model presented by Thüring and Mahlke [1]. The model describes a bridge between instrumental and non-instrumental qualities, which is the user emotion. Law, van Schaik, and Roto introduced the following four groups according to the CUE model to classify attributes.

- Instrumental qualities (INQ) (Utilitarian)
  The system's capabilities and how much it helps the user satisfy his needs and the easiness of the user interactions with the system.
- Non-instrumental qualities (NIQ) (Experiential-Aesthetics)
  The appearance and aesthetics features of the system. And how attractive and tempting the system is for the user.
- Short-term affective response (Experiential-Hedonic)
  The emotions and reactions of the user while encountering the system or immediately after using it.
- Long-term evaluative response (Experiential-Hedonic)
  The user's views, feelings, and perceptions about the system are affected by long-term communication and interaction.

Table 5 presents the characteristics and attributes that belong to these four groups. It determines the extent to which each of the standard questionnaires belongs to these groups, as well as the degree of difference between them. Throughout this chapter, table 5 will be used to classify the items of each questionnaire. For each questionnaire under consideration, a table will list all its items in their given categories. For comparison, all tables will contain a separate column where these items are categorized in accordance with the CUE model. The differences between the item categorizations will be the basis

Qualities	Attributes
Instrumental Qualities	
Measurable	benefit, reliability, clarity, response time, comfort, smoothness, control, speed, efficiency, time on task, learnability, responsiveness
Non measurable	-
Measurable and non measurable	Ease of use, intuitiveness, usability, usefulness
Non Instrumental Qualities	
Measurable	Attractive, cool, desirability, meaning
Non measurable	Identification
Measurable and non measurable	Aesthetic appeal, beauty, challenge, creative, stimulation
Short-Term Affective Response	
Measurable	Affect, arousal, delight, disgust, excitement, frustration, physical pain, stress
Non measurable	Enchantment
Measurable and non measurable	Annoyance, fear, anxiety, flow, attachment, immersion, emotion, joy, engagement, pleasure, enjoyment, surprise
Long-Term Evaluation Response	
Measurable	Competence, expectation, motivation, need fulfillment, relatedness
Non measurable	Self-fulfillment
Measurable and non measurable	Happiness, love, preference, satisfaction, trust

Table 5: Instrumental Qualities, Non-Instrumental Qualities, Short-Term Affective Response and Long-Term Evaluation Response attributes [10]

on which to improve performance measures of given questionnaires. It should be noted that both short-term affective responses and long-term evaluation response attributes were considered to be both Experiential-Hedonic, because both examine emotion of the user in short and long term. Table 23 in the appendix, which has been compiled by Microsoft [1] [20], lists attributes and their synonyms which are used by the model by Law, van Schaik, and Roto [10].
Also, in order to facilitate the analysis of the questionnaires' items, in the domain column of each questionnaire table, the abbreviations U, EA, and EH are used, which refer to INQ (Utilitarian), NIQ (Experiential-Aesthetics), and Experiential-Hedonic, respectively.

QUIS: Table 6 presents the features and attributes of QUIS version 7's items. For the analysis, the short version of the questionnaire was considered, while leaving out

product specific and optional items, so that 32 items were used in total. The INQ or utilitarian group makes up 78.12% or 25 items of the total set. Five items (15.62%) belong to the Experiential-Hedonic group, and two items (6.25%) are Experiential-Aesthetics (NIQ).

Factor and Items	Scales	Domain
Overall		
	terrible - wonderful	EH
	difficult - easy	U
	frustrating - satisfying	EH
	inadequate power - adequate power	U
	dull - stimulating	EA
	rigid - flexible	U
Screen		
Characters on the computer screen	hard to read - easy to read	U
Highlighting on the screen simplifies task	not at all - very much	U
Organization of information on screen	confusing - very clear	U
Sequence of screens	confusing - very clear	U
Terminology and system information		
Use of terms throughout system	inconsistent - consistent	U
Computer terminology is related to the task you are doing	never - always	EH
Position of messages on screen	inconsistent - consistent	U
Messages on screen which prompt user for input	confusing - clear	U
Computer keeps you informed about what it is doing	never - always	U
Error messages	unhelpful - helpful	U
Learning		
Learning to operate the system	difficult - easy	U
Exploring new features by trial and error	difficult - easy	U
Remembering names and use of commands	difficult - easy	U
Tasks can be performed in a straightforward manner	never - always	U
Help messages on the screen	unhelpful - helpful	U
Supplemental reference materials	confusing - clear	U
System capabilities		
System speed	too slow - fast enough	U
System reliability	unreliable - reliable	U
System tends to be	noisy - quiet	U
Correcting your mistakes	difficult - easy	U
Experienced and inexperienced users' needs are taken into consideration	never - always	EH
Usability and user interface		
Use of colors and sounds	poor - good	EA
System feedback	poor - good	EH
System response to errors	awkward - gracious	U
System messages and reports	poor - good	U
System clutter and UI ?noise?	poor - good	U

Table 6: Item analysis of the QUIS 7.0 questionnaire [11]

SUMI: The analysis of the SUMI questionnaire with 50 items is shown in tables 7 and 8.These 50 items include 36 INQ items (72%), 13 Experiential-Hedonic items (26%), and one NIQ (2%). It has been observed that this questionnaire uses sentences with either positive or negative semantic. Examples for noticeable words which belong to Experiential-Hedonic items are feeling, preference, recommendation, satisfaction, and like. Lastly, in the field of Experiential-Aesthetic, only the attractiveness of the product has been studied.

Items	**Domain**
This software responds too slowly to inputs.	U
I would recommend this software to my colleagues.	EH
The instructions and prompts are helpful.	U
This software has at some time stopped unexpectedly.	U
Learning to operate this software initially is full of problems.	U
I sometimes don't know what to do next with this software.	U
I enjoy the time I spend using this software.	EH
I find that the help information given by this software is not very useful.	U
If this software stops it is not easy to restart it.	U
It takes too long to learn the software functions.	U
I sometimes wonder if I am using the right function.	U
Working with this software is satisfying.	EH
The way that system information is presented is clear and understandable.	U
I feel safer if I use only a few familiar functions.	EH
The software documentation is very informative.	U
This software seems to disrupt the way I normally like to arrange my work.	U
Working with this software is mentally stimulating.	EH
There is never enough information on the screen when it's needed.	U
I feel in command of this software when I am using it.	EH
I prefer to stick to the functions that I know best.	EH
I think this software is inconsistent.	U
I would not like to use this software every day.	EH
I can understand and act on the information provided by this software.	U
This software is awkward when I want to do something which is not standard.	U
There is too much to read before you can use the software.	U

Table 7: Item analysis of the SUMI questionnaire items 1 to 25 [12] [13]

Items	Domain
Tasks can be performed in a straight forward manner using this software.	U
Using this software is frustrating.	EH
The software has helped me overcome any problems I have had in using it.	U
The speed of this software is fast enough.	U
I keep having to go back to look at the guides.	U
It is obvious that user needs have been fully taken into consideration.	EH
There have been times in using this software when I have felt quite tense.	EH
The organisation of the menus seems quite logical.	U
The software allows the user to be economic of keystrokes.	U
Learning how to use new functions is difficult.	U
There are too many steps required to get something to work.	U
I think this software has sometimes given me a headache.	EH
Error messages are not adequate.	U
It is easy to make the software do exactly what you want.	U
I will never learn to use all that is offered in this software.	U
The software hasn't always done what I was expecting.	EH
The software presents itself in a very attractive way.	EA
Either the amount or quality of the help information varies across the system.	U
It is relatively easy to move from one part of a task to another.	U
It is easy to forget how to do things with this software.	U
This software occasionally behaves in a way which can't be understood.	U
This software is really very awkward.	U
It is easy to see at a glance what the options are at each stage.	U
Getting data files in and out of the system is not easy.	U
I have to look for assistance most times when I use this software.	U

Table 8: Item analysis of the SUMI questionnaire items 25 to 50 [12] [13]

PSSUQ: The analysis of the PSSUQ version 3.0 questionnaire with 16 items is shown in table 9. In this questionnaire, ten items belong to INQ features (62.5%), and six items refer to Experiential-Hedonic features (37.5%). As can be seen in the table, this questionnaire does not examine NIQ features. Due to PSSUQ being a usability questionnaire, it lacks aesthetic features like the product's appearance. Thus, the user's feelings about the product are not being investigated.

SUS: According to the table 10, the SUS is a short questionnaire with ten items. Eight items examine INQ (80%), and the other two items belong to the Experiential-Hedonic feature (20%). NIQ features are completely absent in this questionnaire. Nevertheless, it claims that its results express an overall user satisfaction level.

Scores	Items	Domain
System Quality	Overall, I am satisfied with how easy it is to use this system.	U
	It was simple to use this system.	U
	I was able to complete the tasks and scenarios quickly using this system.	U
	I felt comfortable using this system.	EH
	It was easy to learn to use this system.	U
	I believe I could become productive quickly using this system.	EH
Information Quality	The system gave error messages that clearly told me how to fix problems.	U
	Whenever I made a mistake using the system, I could recover easily and quickly.	U
	The information (such as online help, on-screen messages, and other documentation) provided with this system was clear.	U
	It was easy to find the information I needed.	U
	The information was effective in helping me complete the tasks and scenarios.	U
	The organization of information on the system screens was clear.	U
Interface Quality	The interface of this system was pleasant.	EH
	I liked using the interface of this system.	EH
	This system has all the functions and capabilities I expect it to have.	EH
	Overall, I am satisfied with this system.	EH

Table 9: Item analysis of the PSSUQ 3.0 questionnaire [14] [15] [16]

Items	**Domain**
I think that I would like to use this system frequently	U
I found the system unnecessarily complex	U
I thought the system was easy to use	U
I think that I would need the support of a technical person to be able to use this system	U
I found the various functions in this system were well integrated	U
I thought there was too much inconsistency in this system	U
I would imagine that most people would learn to use this system very quickly	EH
I found the system very cumbersome to use	U
I felt very confident using the system	EH
I needed to learn a lot of things before I could get going with this system	U

Table 10: Item analysis of the SUS questionnaire [5]

AttrakDiff2: Table 11 shows the analysis of AttrakDiff2 questionnaire items with four factors and 28 items. There are seven items which describe INQ (25%) and eight items which are Experiential-Hedonic qualities (28.57%), and 13 items with NIQ (46.42%). The way AttrakDiff2 sorts its items into categories differs greatly from the CUE model. For example, in AttrakDiff2, attractiveness is related to the product's positive and negative statements. However, in the CUE model, attractiveness is the subset of NIQ, and it refers to the aesthetic appearance of the product. Also, in the CUE model, the user's connection with the product is part of the Experiential-Hedonic category. However, for AttrakDiff2, the very same thing is part of the Hedonic-Identification quality. Same category applies to product specific qualities; Therefore, according to the CUE model, the adjectives stylish, premium, and presentable are among the NIQ category mentioned in the Hedonic-Identification section. There are also attributes in the AttrakDiff2's attractiveness factor, such as pleasant, likable, inviting, and motivating, which according to the CUE model belong to Experiential-Hedonic.

Factor and items	**Domain**	**Factor and items**	**Domain**
Pragmatic		**Hedonic-Stimulation**	
technical - human	U	conventional - inventive	EA
complicated - simple	U	unimaginative - creative	EA
impractical - practical	U	cautious - bold	EA
cumbersome - straightforward	U	conservative - innovative	EA
unpredictable - predictable	U	dull - captivating	EA
confusing - clearly structured	U	undemanding - challenging	EA
unruly - manageable	U	ordinary - novel	EA
Hedonic -Identification		**Attractiveness**	
isolating - connective	EH	unpleasant - pleasant	EH
unprofessional - professional	EH	ugly - attractive	EA
tacky - stylish	EA	disagreeable - likeable	EH
cheap - premium	EA	rejecting - inviting	EH
alienating - Integrating	EH	bad - good	EA
separates me - brings me closer	EH	repelling - appealing	EA
unpresentable - presentable	EA	discouraging - motivating	EH

Table 11: Item analysis of the AttrakDiff2 questionnaire [17]

UEQ: The UEQ questionnaire examines six factors and 26 items. According to table 12, this questionnaire examines the characteristics of the INQ with 11 items (42.30%), the Experiential-Hedonic characteristics with eight items (30.76%), and the NIQ with seven items (26.92%). In this questionnaire, the attributes of enjoyable, pleasing,

pleasant, and friendly belong to the Attractiveness category, while in the CUE model, these attributes belong to the Experiential-Hedonic category.

Factor and items	Domain	Factor and items	Domain
Pragmatic - Perspicuity		**Hedonic - Novelty**	
not understandable - understandable	U	creative - dull	EA
easy to learn - difficult to learn	U	inventive - conventional	EA
complicated - easy	U	usual - leading edge	EA
clear - confusing	U	conservative - innovative	EA
Pragmatic - Dependability		**Hedonic - Stimulation**	
unpredictable - predictable	U	valuable - inferior	EA
obstructive - supportive	U	boring - exciting	EH
secure - not secure	U	not interesting - interesting	EH
meets expectationvdoes not meet expectation	EH	motivating - demotivating	EH
Pragmatic - Efficiency		**Attractiveness**	
fast/slow	U	annoying - enjoyable	EH
inefficient - efficient	U	good - bad	EA
impractical - practical	U	unlikeable - pleasing	EH
organized - cluttered	U	unpleasant - pleasant	EH
		attractive - unattractive	EA
		friendly - unfriendly	EH

Table 12: Item analysis of the UEQ questionnaire [18]

meCUE: The items analysis of meCUE questionnaire with 34 items and ten factors is shown in table 13. According to this table, this questionnaire examines the INQ feature with six items (17.64%), the Experiential-Hedonic attribute with 25 items (73.52%), and the NIQ feature with three items (8.82%). The item analysis of the meCUE questionnaire determines that it mostly studies the Experiential-Hedonic features.

The results of the item analysis of all selected questionnaires in this thesis are given in table 14. It compares the number of items which belong to one CUE category and their percentage. Generally, the categorizations given by the questionnaires roughly meet the CUE model's expectations. As can be seen in usability questionnaires, a high percentage of questionnaire items are based on the INQ attribute, while Experiential-Hedonic and aesthetic features (NIQ) are mentioned very little or not at all. For example, in two questionnaires, SUS and PSSUQ, no aesthetic features were considered. In the QUIS and SUMI questionnaires, the aesthetic quality of 6.25% and 2% are discussed, respectively. Due to several items, QUIS and SUMI have paid less attention to features other than usability.

Factor and items	Domain	Factor and items	Domain
Usefulness		**Intention to use**	
The functions of the product are exactly right for my goals.	EH	If I could, I would use the product daily.	EH
It is quickly apparent how to use the product.	U	I can hardly wait to use the product again.	EH
The operating procedures of the product are simple to understand.	U	When using the product, I lose track of time.	EH
Usability		**Product Loyalty**	
The product is easy to use.	U	I would not swap this product for any other.	EH
I consider the product extremely useful.	U	In comparison to this product, no others come close.	EH
With the help of this product I will achieve my goals.	U	I would get exactly this product for myself (again) at any time.	EH
Positive Emotions		**Negative Emotions**	
The product exhilarates me.	EH	The product makes me tired.	EH
The product relaxes me.	EH	The product annoys me.	EH
The product makes me feel happy.	EH	When using this product I feel exhausted.	EH
The product makes me feel euphoric.	EH	The product frustrates me.	EH
The product calms me.	EH	The product makes me feel passive.	EH
When using this product, I feel cheerful.	EH	The product angers me.	EH
Status		**Commitment**	
The product would enhance my standing among my peers.	U	I could not live without this product.	EH
By using this product, I would be perceived differently.	EH	This product is like a friend to me.	EH
I would not mind if my friends envied me for this product.	EH	If I ever lost the product, I would be devastated.	EH
Visual Aesthetics		**Overall Evaluation**	
The product is creatively designed.	EA	How do you experience the product as a whole?	EH
The design looks attractive.	EA		
The product is stylish.	EA		

Table 13: Item analysis of the meCUE questionnaire [19]

After analyzing the items of user experience questionnaires, it was found that these questionnaires have addressed all the features of INQ, Experiential-Hedonic, and NIQ to some extent. And they are a combination of all three features. Meanwhile, the meCUE questionnaire with 34 items deals with 73% of aesthetic features than other features. In AttrakDiff2 and meCUE, aesthetic features clearly dominate. UEQ reserves 42.30% of its items for INQ, which is more than any other questionnaire does.

Questionnaire	**Instrumental Qualities**	**Experiential Aesthetics**	**Experiential Hedonic**	**SUMS**
QUIS	25 items 78.12%	2 items 6.25%	5 items 15.62%	32 items
SUMI	36 items 72%	1 item 2%	13 items 26%	50 items
PSSUQ	10 items 62.5%	0 0	6 items 37.5%	16 items
SUS	8 items 80%	0 0	2 items 20%	10 items
AttrakDiff2	7 items 25%	13 items 46.42%	8 items 28.57%	28 items
UEQ	11 items 42.30%	7 items 26.92%	8 items 30.76%	26 items
meCUE	6 items 17.64%	25 items 73.52%	3 items 8.82%	34 items

Table 14: The item analysis of 7 selected usability and user experience questionnaires and with the percentage of Instrumental qualities, Non-instrumental qualities and Experiential-Hedonic [20]

3.2 Design of the extended Questionnaire

As shown in the previous section, all questionnaires focus their attention on single quality characteristics. As shown in table 14, that means that user experience questionnaires do not balance all qualities out, while usability questionnaires even naturally fail to investigate hedonistic and aesthetic qualities. They examine more INQ features than other quality features. Today, to be more profitable and sell more products, businesses should not only focus on instrumental features. It is essential to know the customer, his interests, preferences, and feelings about the product before and after and during its usage. Therefore, there is a need to establish a balance between different quality characteristics, such as hedonistic and aesthetic features in addition to the utilitarian features. Therefore, it seems reasonable to apply changes to the usability questionnaires for better efficiency and improve their performance in both usability and user experience.

Extending the questionnaire is one way to examine more quality features without changing the original model. In this way, a set of items is added at the end of the questionnaire. If a questionnaire falls short of a certain quality, new items or factors can be added for better balance. By applying this method, the questionnaire's previous structure is not being altered while no complexity is added to its evaluation. Another advantage is that the results from before and after the extension are still comparable.

One of the questionnaires reviewed in the previous section has to be chosen for these modifications. Since the usability questionnaires lack non-instrumental and hedonistic items, they leave more room for improvements. Hence, extending them with new items should have a greater effect on these, yielding clearly measurable results. By examining QUIS and SUMI questionnaires, firstly, it was found that they are not free of charge to use. And there is a fee for using them and applying changes. Secondly, these two questionnaires are both long. The QUIS 7.0 short version includes 32 items, regardless of the optional items. The SUMI also has 50 items. Therefore, expanding these two questionnaires makes them too long, and the user needs more time to answer them. This long response time is a negative point that may cause the user to become bored and leads to a lack of interest, attention, and even cause the user to be inaccurate when responding to the questionnaire.

Among the remaining questionnaires in usability, SUS is a questionnaire that consists of 80% utilitarian items, with the rest (20%) being Experiential-Hedonic items and zero aesthetic features. However, the SUS questionnaire with ten items did not serve as a suitable option for extensions, because its results describes an overall satisfaction level of the user, which is one single factor. Adding new factors would break the concept and is therefore infeasible.

Finally, the PSSUQ version 3.0 with 16 items examines the both, usability and hedonic

features, which is a good basis for extensions. First of all, the PSSUQ questionnaire is free and can be used or modified at the only cost of mentioning the source. Secondly, it does not include many items, and it is possible to add new items to it. Finally, the PSSUQ questionnaire does not examine aesthetic features at all. Adding items to evaluate this feature results in an analysis of a wider range of product quality features and a better understanding of the user and his opinions.

3.2.1 Purpose of Extended-PSSUQ Questionnaire

The PSSUQ version 3.0 has 16 items with four factors introduced in the background 2.4.3 and item analysis section 3.1.3. Adding new factors in the fields of hedonic and aesthetic experiences makes it possible to make changes to this questionnaire from item 16 on wards. In this way, the reliability of 3 factors of System Quality, Information Quality, and Interface Quality remains constant and does not change, with the overall factor being the only exception, because it has to account for new items in the average calculation. Besides the overall factor, the extended questionnaire can be evaluated by measuring each factor's reliability before and after the extension. This assessment is done separately for each factor, because they do not interfere with each other. For this reason, adding new factors and items is useful and efficient. This is expected to collect more information about the product's quality and its user experience.

3.2.2 Characteristics of Extended-PSSUQ Questionnaire

Since the PSSUQ deals with ten utilitarian items, and due to the lack of the hedonistic and aesthetic factors, two factors of the Experiential Hedonic and Experiential Aesthetics are added to this questionnaire as independent factors. For both of these two factors reliability can be measured separately, same applies to the native four factors in the PSSUQ questionnaire. Also, the overall reliability is determined by calculating the average response of all items in the questionnaire. Extended-PSSUQ questionnaire scores are as follow:

- Overall: Average the responses for of items 1 through 28 (all the items)
- System Quality (SysQual): Average of items 1 through 6
- Information Quality (InfoQual): Average of items 7 through 12
- Interface Quality (IntQual): Average of items 13 through 15
- Experiential Hedonic: Average of items 17 through 20
- Experiential Aesthetics: Average of items 21 through 28

3.2.3 Context of Extended-PSSUQ Questionnaire

The Extended-PSSUQ questionnaire has 28 items, which are 12 items more than PSSUQ 3.0. Four of these additional items measure the Experiential Hedonic qual-

ity and eight items the Experiential Aesthetics quality. The Extension of these items stem from the CUE model 3.1.3. In order to add a missing factor, table 5 presents related attributes, which will be used to create new items. For example, it is stated that aesthetics qualities are absent in PSSUQ. Then, table 5 provides a section for non instrumental qualities, like the aesthetic qualities we are looking for. Following this section, there is a list of measurable attributes, like attractive or cool. Ultimately, said attributes are used to form new items, e.g. "The audio notification of the system is cool". Item by item, the factor for aesthetic comes into being. Finally, table 15 shows the final Extended-PSSUQ items and domains. For the Experiential-Aesthetics factor for the system's attractiveness, the attributes creativity, novelty, innovative, stylish, and appealing have been picked. The new items deal with color and audio notifications of the system. These features make a connection with the user in terms of acoustic and visual attractiveness. The selected questionnaires did not have an item to check the system's sound feature, and this feature was not addressed. The sound attraction of the system affects the user's comprehensive perception of product quality and his emotions. Examining the system's acoustic signal also determines what the system's sound means to the user. Also, in Experiential Hedonic quality, the characteristics of confidence, motivation, stimulation, user expectations, and preferences are discussed.

Scores	Items	Domain
System Quality	Overall, I am satisfied with how easy it is to use this system.	U
	It was simple to use this system.	U
	I was able to complete the tasks and scenarios quickly using this system.	U
	I felt comfortable using this system.	EH
	It was easy to learn to use this system.	U
	I believe I could become productive quickly using this system.	EH
Information Quality	The system gave error messages that clearly told me how to fix problems.	U
	Whenever I made a mistake using the system, I could recover easily and quickly.	U
	The information (such as online help, on-screen messages, and other documentation) provided with this system was clear.	U
	It was easy to find the information I needed.	U
	The information was effective in helping me complete the tasks and scenarios.	U
	The organization of information on the system screens was clear.	U
Interface Quality	The interface of this system was pleasant.	EH
	I liked using the interface of this system.	EH
	This system has all the functions and capabilities I expect it to have.	EH
	Overall, I am satisfied with this system.	EH
Experiential Hedonic	I could intuitively use the system.	EH
	Using the system motivates me to solve my tasks.	EH
	I had spontaneous idea for other use cases of the system.	EH
	I prefer the system to other similar systems.	EH
Experiential Aesthetics	The color scheme of this system is beautiful.	EA
	The audio notification of the system is cool.	EA
	The system resolves the problems creatively.	EA
	The system integrated with other system inventively.	EA
	The system layout is stylish.	EA
	The system solves the problem in a novel way.	EA
	The system is represented in appealing way.	EA
	I can name an innovative feature of the system.	EA

Table 15: The Extended-PSSUQ questionnaire item analysis

The full version of the extended Extended-PSSUQ is shown in figure 16. To facilitate answering Extended-PSSUQ for German speakers, it was translated into the German language . The German sample of this questionnaire is attached in the appendix in figure 20.

	The Extended Post Study System Usability Questionnaire	Strongly agree							Strongly disagree	
			1	2	3	4	5	6	7	NA
1	Overall, I am satisfied with how easy it is to use this system.		☐	☐	☐	☐	☐	☐	☐	☐
2	It was simple to use this system.		☐	☐	☐	☐	☐	☐	☐	☐
3	I was able to complete the tasks and scenarios quickly using this system.		☐	☐	☐	☐	☐	☐	☐	☐
4	I felt comfortable using this system.		☐	☐	☐	☐	☐	☐	☐	☐
5	It was easy to learn to use this system.		☐	☐	☐	☐	☐	☐	☐	☐
6	I believe I could become productive quickly using this system.		☐	☐	☐	☐	☐	☐	☐	☐
7	The system gave error messages that clearly told me how to fix problems.		☐	☐	☐	☐	☐	☐	☐	☐
8	Whenever I made a mistake using the system, I could recover easily and quickly.		☐	☐	☐	☐	☐	☐	☐	☐
9	The information (such as online help, on-screen messages, and other documentation) provided with this system was clear.		☐	☐	☐	☐	☐	☐	☐	☐
10	It was easy to find the information I needed.		☐	☐	☐	☐	☐	☐	☐	☐
11	The information was effective in helping me complete the tasks and scenarios.		☐	☐	☐	☐	☐	☐	☐	☐
12	The organization of information on the system screens was clear.		☐	☐	☐	☐	☐	☐	☐	☐
13	The interface of this system was pleasant.		☐	☐	☐	☐	☐	☐	☐	☐
14	I liked using the interface of this system.		☐	☐	☐	☐	☐	☐	☐	☐
15	This system has all the functions and capabilities I expect it to have.		☐	☐	☐	☐	☐	☐	☐	☐
16	Overall, I am satisfied with this system.		☐	☐	☐	☐	☐	☐	☐	☐
17	I could intuitively use the system.		☐	☐	☐	☐	☐	☐	☐	☐
18	Using the system motivates me to solve my tasks.		☐	☐	☐	☐	☐	☐	☐	☐
19	I had spontaneous idea for other use cases of the system.		☐	☐	☐	☐	☐	☐	☐	☐
20	I prefer the system to other similar systems.		☐	☐	☐	☐	☐	☐	☐	☐
21	The color scheme of this system is beautiful.		☐	☐	☐	☐	☐	☐	☐	☐
22	The audio notification of the system is cool.		☐	☐	☐	☐	☐	☐	☐	☐
23	The system resolves the problems creatively.		☐	☐	☐	☐	☐	☐	☐	☐
24	The system integrated with other system inventively.		☐	☐	☐	☐	☐	☐	☐	☐
25	The system layout is stylish.		☐	☐	☐	☐	☐	☐	☐	☐
26	The system solves the problem in a novel way.		☐	☐	☐	☐	☐	☐	☐	☐
27	The system is represented in appealing way.		☐	☐	☐	☐	☐	☐	☐	☐
28	I can name an innovative feature of the system.		☐	☐	☐	☐	☐	☐	☐	☐

Figure 16: The Extended-PSSUQ questionnaire

4 Evaluation

As depicted in figure 17, this chapter consists of four sections. In order to test the Extended-PSSUQ questionnaire, a survey about Mattermost has been conducted. This software is part of the E-Learning system of the Otto-von-Guericke-University of Magdeburg. Chapter 4.1 gives more details. Also, to evaluate and compare the results of the PSSUQ questionnaire and Extended-PSSUQ, a short questionnaire with three items called ASQ has been used, which is briefly explained in part 4.1. Three techniques have been used to evaluate the Extended-PSSUQ. First, in section 4.2, the results of the Extended-PSSUQ are reviewed by using basic statistical methods, which are the mean and 95% confidence interval measurements. This method examines the central tendency of a set of data and finds the relationship between the population of interest and the mean. In the next section 4.3, the factor analysis technique, which is a reduction model, is examined. The input variables of this model are the user ratings of the items. The algorithm condenses a number of these variables to a few more meaningful variables called factors. This reveals the relationship between items and factors. Finally, the 4.4 subsection evaluates the performance quality of the Extended-PSSUQ questionnaire. The accuracy of the collected data will be measured with three metrics: the objectivity 4.4.1, reliability 4.4.2, and validity 4.4.3.

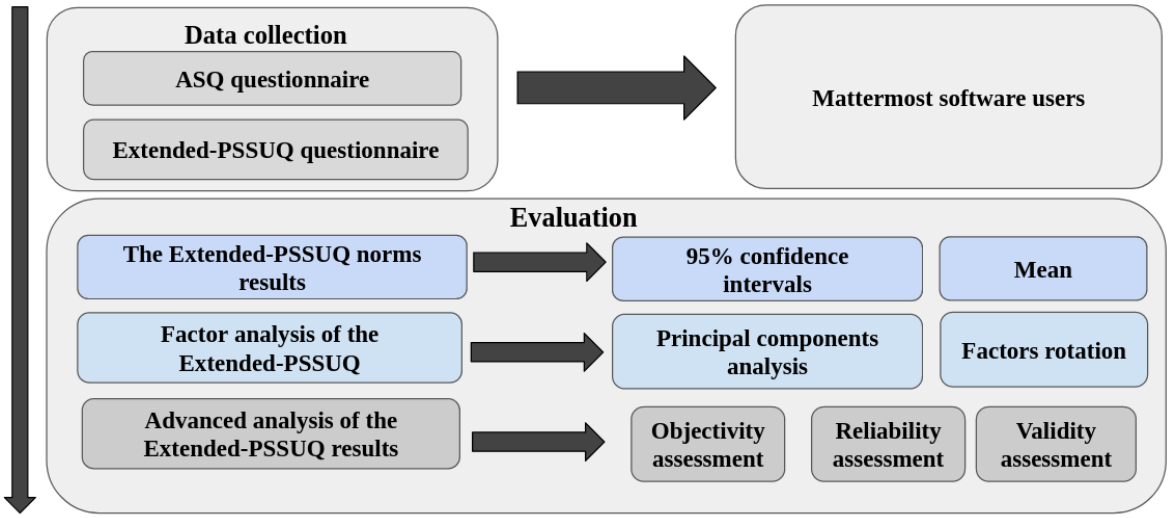


Figure 17: Evaluation chapter pipeline

4.1 Survey of the Mattermost software users

The Mattermost software establishes the online seminars and communication between students and professors in Otto-von-Guericke-University. It can be integrated into many softwares and platforms and personalized according to the users' wishes. With the help of other software, this software enables video and voice conferencing between a large number of users. It is also possible to install it on mobiles and personal computers

and send notifications. Its GIT integration also allows students to submit their homework. This software's versatility makes it a worthy candidate for the Extended-PSSUQ to assess users satisfaction.

Survey of the Mattermost software user by ASQ questionnaire: The ASQ is a usability questionnaire designed in 1995 by Lewis, J. R. [15]. This questionnaire has three items and a "Likert scale" with seven numbers ranging between "Strongly agree" and "Strongly disagree." The evaluation of this questionnaire is straightforward, and only the average of the three answered items should be calculated. The items of this questionnaire can be seen in figure 18.

	The After-Scenario Questionnaire Version 1	Strongly agree						Strongly disagree		
			1	2	3	4	5	6	7	NA
1	Overall, I am satisfied with the ease of completing the tasks in this scenario.		O	O	O	O	O	O	O	O
2	Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario.		O	O	O	O	O	O	O	O
3	Overall, I am satisfied with the support information (online help, messages, documentation) when completing the tasks.		O	O	O	O	O	O	O	O

Figure 18: The ASQ questionnaire

The ASQ questionnaire has been answered by Mattermost users to assess the validity of the Extended-PSSUQ questionnaire. In the validity section, the results of the PSSUQ and Extended-PSSUQ questionnaires are compared with ASQ results.
Also, both Extended-PSSUQ and ASQ were mailed to students of the Otto-von-Guericke-University in English and German. The raw data of the surveys can be requested from the author.

4.2 The Extended-PSSUQ Norms (Means and 95% Confidence Intervals)

Statistical parameter calculation is a method to find the relationship between the population of interest and the mean value. The mean estimation of the population is always practical and analyzable. Since the estimation of the parameters is not accurate and is affected by the measurement variation, it is better to present estimations as a confidence interval. The confidence interval determines the accuracy of population parameter estimation, and it gives a range for each parameter. A small confidence interval means that the estimation is precise and there is less room for errors. This is generally the more desirable option. A great interval means that a variable is so error prone that is hard to pin-point it to a single value rather than a range where

it is most likely located. The confidence interval depends on the variation between the population, sampling size, and desired confidence level. The greater the variation between populations, the greater the confidence interval, and vice versa. On the other hand, more samples amount for a smaller confidence interval, and vice versa. A larger sample count reduces sampling errors because more information is collected about the population of interest, and it leads to estimate the parameters more accurately.
This part introduces the math for the mean and confidence interval which will be calculated for the Extended-PSSUQ's 28 items. The math steps of the confidence interval calculation are as follows:

1. Compute Mean
$$\bar{x} = \frac{\sum_{i=1}^{N}}{N} \tag{1}$$
N = sample number

2. Compute standard deviation
$$S_x = \frac{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}{N-1} \tag{2}$$

3. Compute standard error
$$SE_{\bar{x}} = \frac{S_x}{\sqrt{N}} \tag{3}$$

4. Look up to t value in *T Distribution Table* for 95% confidence level.

5. Getting the critical difference(d) by multiplying t and *standard error*

6. Getting the upper and lower bounds of the confidence interval by adding and subtracting the critical difference from the *mean*

Table 16 shows items' means and confidence intervals (norms), where lower scores mean that the user agrees with a statement more. 21 users took part in this survey. The mean of all items' assessments falls below the scale midpoint of 4, and only item 19 ("I had spontaneous idea for other use cases of the system.") is an exception with a mean value of 4.04. Also, the upper limits of the 95% confidence intervals are often lower than the scale midpoint of 4, but only items 9, 18, 19, 20, 25, 26, and 28 have a higher upper limit.
Table 17 illustrates the factors norms. All the mean and upper limits of 95% confidence intervals are lower than the scale midpoint of 4, except for the Experiential Hedonic's upper limit, which is 4.43.
Since this is the first iteration where the Extended-PSSUQ have been used, it is not yet possible to conclude if these factors or items perform well or still have shortcomings.

Item	Item text	Lower limit	Mean	Upper limit
1	Overall, I am satisfied with how easy it is to use this system.	2.03	2.71	3.39
2	It was simple to use this system.	1.83	2.61	3.4
3	I was able to complete the tasks and scenarios quickly using this system.	1.14	2.19	3.23
4	I felt comfortable using this system.	1.79	2.42	3.065
5	It was easy to learn to use this system.	1.61	2.38	3.14
6	I believe I could become productive quickly using this system.	2.3	3.14	3.98
7	The system gave error messages that clearly told me how to fix problems.	1.39	2.14	2.89
8	Whenever I made a mistake using the system, I could recover easily and quickly.	1.08	1.76	2.43
9	The information (such as online help, on-screen messages, and other documentation) provided with this system was clear.	2.69	3.47	**4.25**
10	It was easy to find the information I needed.	1.51	2.33	3.15
11	The information was effective in helping me complete the tasks and scenarios.	1.65	2.28	2.91
12	The organization of information on the system screens was clear.	1.9	2.57	3.23
13	The interface of this system was pleasant.	1.72	2.19	2.65
14	I liked using the interface of this system.	1.91	2.42	2.93
15	This system has all the functions and capabilities I expect it to have.	2.15	2.95	3.74
16	Overall, I am satisfied with this system.	2.05	2.61	3.18
17	I could intuitively use the system.	1.81	2.9	3.99
18	Using the system motivates me to solve my tasks.	1.61	2.8	4
19	I had spontaneous idea for other use cases of the system.	2.8	**4.04**	**5.2**
20	I prefer the system to other similar systems.	2.37	3.71	**5.5**
21	The color scheme of this system is beautiful.	2.29	3.28	**4.2**
22	The audio notification of the system is cool.	0.36	1.61	2.87
23	The system resolves the problems creatively.	2.16	3	3.83
24	The system integrated with other system inventively.	1.17	2.14	3.11
25	The system layout is stylish.	2.45	3.38	**4.3**
26	The system solves the problem in a novel way.	2.52	3.38	**4.23**
27	The system is represented in appealing way.	2.1	3	3.89
28	I can name an innovative feature of the system.	1.71	3.04	**4.37**

Table 16: The Extended-PSSUQ Item Norms (Means and 95% Confidence Intervals)

Factors	Factor item rule	Lower limit	Mean	Upper limit
System Quality	Average Item 1-6	1.9	2.57	3.2
Information Quality	Average Item 7-12	1.89	2.42	2.95
Interface Quality	Average Item 13-15	2	2.52	3.03
Overall PSSUQ V3.0	Average Item 1-16	2.05	2.51	2.97
Experiential Hedonic	Average Item 17-20	2.3	3.36	**4.43**
Experiential Aesthetics	Average Item 20-28	2.01	2.85	3.69
Overall Extended-PSSUQ	Average Item 1-28	2.26	2.73	3.2

Table 17: The Extended-PSSUQ Factor Norms (Means and 95% Confidence Intervals)

First of all, the same questionnaire should be filled by the same users at different times so that temporal effects like low skills can be excluded. The respondent's opinion might change over time because he could learn new features or workflows of this product or one of market competitors. Then, the extended-PSSUQ should be applied to different users and products. This will yield more means and confidence intervals per item to analyze. In a case of non-stable outcomes, items or factors can be redesigned and checked for improvement. Any significant reduction of the confidence interval in subsequent questionnaire iterations indicates the success of the changes. For the results at hand (table 17), the Experiential Hedonic factor sticks out with a high mean value and upper limit. This stems from item 19, which has a mean value of 4.04. Since this is the first iteration of the Extended-PSSUQ, there are ways to deal with this anomaly or to interpret it:

1. If the anomaly persists through the iterations, the items need rework.
2. Finding other system applications is not easy for the user, and he should think more about it.
3. Try to introduce different system applications to the users and check if this reduces the item's score and consequently the Experiential Hedonic factor.

4.3 Factor Analysis of the Extended-PSSUQ

Factor analysis is a reduction model that reduces the number of main variables and calculates new variables named factors. It is used to find a relationship model for a set of measured variables. It combines variables to relate them to each other or to one dimension. Factor analysis is used in two stages of the questionnaire's construction. The first is exploratory factor analysis, which is usually used by questionnaire designers in the early stages of design to identify the relationships between variables/items to assign a factor to them. It also reduces the number of items and determines which items are appropriate for the questionnaire. The second is the confirmatory factor

analysis, which is applied in the last stage of the questionnaire design. It examines the consistency of the questionnaire factors with the concept for which the questionnaire is designed. This method confirms or denies the presence/assumption of a factor. In factor analysis, both models are calculated similarly but in different time stages.
The most popular factor analysis methods are Principal Components Analysis (PCA) and Maximum Likelihood. PCA is not considered a factor analysis method on its own, but it is a method of variable reduction, based on Maximum Likelihood. The PCA calculates the common features by using the total variance of the individual items. It transforms the main variables into a smaller set with a linear composition named principal components. In contrast, the Maximum Likelihood method is used to find the structure between variables in addition to the variable reduction. Also, the Maximum Likelihood considers the common covariances of the variables. This thesis, since the used model is just an extension of PSSUQ V3.0, both exploratory and confirmatory factor analyses have already applied by the authors of the original model. In the Extended-PSSUQ, items have been added for two aesthetic and hedonistic factors while preserving its main structure. Therefore, the purpose of factor analysis in this step is to compare and match only the new items with the two new added factors and eliminate incompatible items. In the following, a brief explanation of the PCA calculation is given:

1. Standardization of the data
$$z = \frac{x - \mu}{\sigma} \tag{4}$$

2. Computing the covariance matrix, where $\mathbf{Z} = [z_1, z_2, \dots]$
$$\mathbf{C} = COV(\mathbf{Z}) = \frac{1}{n-1}\mathbf{Z}^T\mathbf{Z} \tag{5}$$

3. Calculate the eigenvectors $\mathbf{V} = [v_1, v_2, \dots]$ of the covariance matrix, where v_1 corresponds with the greatest eigenvalue λ_1 and so on

4. Computing the Principal Components
   Eigenvectors and eigenvalues should be arranged in descending order. The eigenvector with the highest eigenvalue is the most significant and the first principle component, and so on. Finally, a feature matrix $\mathbf{Y}$ is formed with $\mathbf{Y} = \mathbf{VZ}$, which includes all significant variables. The variables should then be sorted according to the final principal components, which display the most significant information.

5. Decide the number of factors and factor loading

6. Rotation of factor loading
   A number of factor models can be extracted after factor analysis. The n-dimensional

space in which the factor analysis was performed can be rotated to extract these models. Before rotation, factors are ambiguous, and its interpretation is difficult. Rotation leads to a simpler and more efficient structure; also, it simplifies the interpretation of the factors. There are two main rotation methods, which include: orthogonal rotation and oblique rotation.

In orthogonal rotation, the factors rotate 90 degrees, and the factors are considered to be uncorrelated to each other. Since the factors are usually correlated, this method is not entirely realistic. One of the common and old techniques of the orthogonal rotation is Varimax. It reduces the number of variables that have high loadings on each factor. It also makes low loadings even smaller.

The factors are considered correlated in the oblique rotation method, so the factors are not rotated 90 degrees. This method rotates the factors in the direction of the vector clusters of the variables and gets closer to them. And as a result, variables interpretation becomes easier. One of the most common techniques for calculating oblique rotation is the Promax. This technique is preferred because it is based on the correlation of factors and performs oblique rotation faster than other techniques. Oblique rotation calculation is more complicated than orthogonal rotation, and it produces a more accurate answer if there is a correlation between variables. Also, it gives the same orthogonal rotation response if there is no correlation between the factors. The orthogonal rotation has been used frequently in the past because its mathematics calculations have been easier. Now, complex calculations have been made easier by computers advancement, so using oblique rotation is more feasible. Because if there is no correlation between the factors, oblique rotation creates the same orthogonal rotation results. The oblique rotation has been applied in this thesis.

7. Provide a number of factors

All calculations related to the factor analysis have been done with the software *SPSS* and are based on oblique rotation and the Promax technique. Also, factor analysis has been done only on items 17 to 28 of the Extended-PSSUQ questionnaire. Since new items have been added to the PSSUQ Version 3.0, and factor analysis has already been done for items 1 to 16. Changing the structure of the original version of the PSSUQ Version 3.0 is not the goal of this thesis.

Table 18 and the scree plot in figure 19 show the eigenvalues of the analysis. They suggest a two factors solution, because the first two eigenvalues are much greater than 1.0. The eigenvalue demonstrates the total amount of variance, and the given principal component describes it. The eigenvalues start from the first component, and the next component is taken from the partialization of the previous component. Therefore, the first component always shows the greatest variance score, and the last component

specifies the smallest variance score. Table 18 shows each component variance score. As mentioned earlier, the purpose of the PCA is to reduce the number of variables. Therefore, selecting the component in which its eigenvalues are greater than one helps find the optimal components. The optimal rotated components according to table 18 are the first two components with the eigenvalues 5.753 and 3.553, which are greater than one.

	Initial eigenvalues			**Rotation sums of squared loadings**
Component	**Total**	**% of variance**	**Cumulative %**	**Total**
1	5.923	49.359	49.359	**5.753**
2	2.923	24.362	73.721	**3.553**
3	1.00	8.351	82.072	
4	.857	7.145	89.217	
5	.507	4.224	93.441	
6	.322	2.680	96.120	
7	.164	1.368	97.489	
8	.113	.943	98.431	
9	.069	.574	99.005	
10	.056	.471	99.476	
11	.052	.436	99.912	
12	.011	.088	100.000	

Table 18: Total variance explained

According to table 18, the Total column lists the eigenvalues, which represent the amount of variance in the original variables calculated for each component. The percentage of the variance column represents the ratio that describes the percentage of the variance calculated for each component according of the total variance of all variables. The "Cumulative %" column indicates the percentage of variance calculated for by first n components.

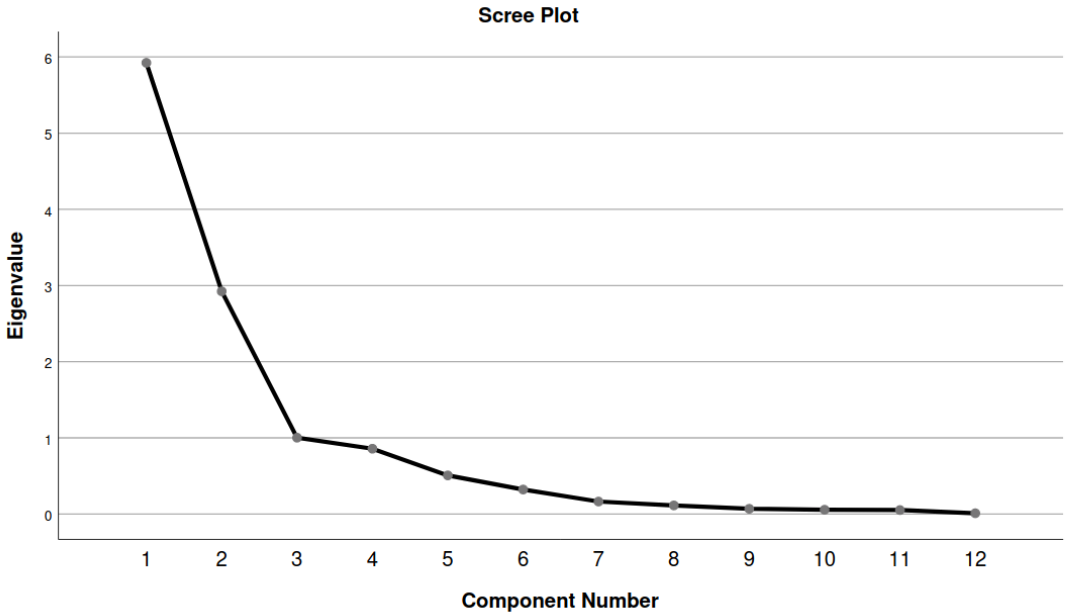


Figure 19: Scree plot from factor analysis

The pattern matrix shown in table 19 contains the linear composition coefficients of the variables. The table shows that items 21 to 28 have a common factor, and if any changes happen, they will variate together. On the other hand, items 17 to 20 also hang together.

	Component	
	1	2
Item23	**.918**	-.056
Item27	**.910**	-.022
Item22	**.896**	-.103
Item25	**.848**	.068
Item26	**.843**	.198
Item28	**.807**	-.114
Item24	**.761**	-.159
Item21	**.667**	.265
Item17	-.072	**.897**
Item20	.020	**.888**
Item19	-.104	**.871**
Item18	.071	**.868**

Table 19: Pattern matrix

The correlation between the items and each factor is depicted by the structure matrix shown in table 20. According to the table items, 21 to 28 correlated with factor 1. Moreover, items 23 and 27 have the highest correlation of 0.905 with a factor of one. On the other hand, items 17 to 20 correlated with factor 2, and item 20 has the highest

correlation with a value of 0.893 with factor 2.

	Component	
	1	2
Item23	**.905**	.159
Item27	**.905**	.190
Item26	.889	.394
Item22	.872	.107
Item25	.864	.266
Item28	.781	.074
Item21	.729	.420
Item24	.724	.019
Item20	.227	**.893**
Item18	.274	.885
Item17	.137	.880
Item19	.099	.847

Table 20: Factor structure matrix

4.4 Advanced Analysis of the Extended-PSSUQ

The data collected by the questionnaire and its interpretation are of great importance. Therefore, these results must be consistent with reality in order to lead to beneficial changes. Therefore, qualitative criteria should be used to evaluate the questionnaire. Studying the results of these qualitative criteria leads to a better understanding of the user and his behavior. Also, more accurate information will cause better company decisions. The most important quality criteria are objectivity, reliability, and validity, which are described below. The presence of these three metrics is interdependent, and each is a prerequisite for the other. Respectively, objectivity is a prerequisite for reliability, and also reliability is a prerequisite for validity.

4.4.1 Objectivity

In data collection and questionnaire evaluation, the results must be obtained without any dependence or bias. Only when the results are independent of any conditions, they can be compared with each other, which is called objectivity. There are different types of objectivity, which are discussed below.

- Data collection objectivity
  The conditions of those who answer the questionnaire must be the equal. The respondents should also not be selected by the researcher, and he should not have any interference or influence on the obtained results.

- Evaluation objectivity
  The questionnaire should be designed so that it does not discriminate against the person who evaluates it. The evaluation method for the results should be determined before the survey, because afterwards, the collected data could influence the decision of the surveyor. If the questionnaire is open-ended, a decision must be made in advance on how to evaluate it. If the questionnaire is closed-ended, it is usually easier to achieve objectivity than for an open-ended one, because the answers are determined beforehand and so is the evaluation.
- Interpretation objectivity
  All those who participate in the evaluation of the questionnaire must achieve the same result. Evaluators must assess the results in accordance with the same criteria and rules. They must refrain from freely interpreting and explaining according to their interests and opinions. In other words, evaluators' interpretations should be free of bias and prejudice.

Objectivity in this thesis is ensured in all three categories data collection, evaluation, and interpretation. Based on the data collection objectivity, all respondents had similar answering conditions, and they are all students of Otto-von-Guericke-University. This questionnaire was sent to the participants' emails online. There was no interference or influence from the researcher on the respondents and their answers. All the results have been calculated with high accuracy with $SPSS$ software for evaluating and interpreting the results. There was not any researcher preconception in interpreting results.

4.4.2 Reliability

Reliability determines if the items amount for consistent measurements. The reliability of questionnaires is defined by the degree of correlation or relationship of items of a factor with each other. The prerequisite for a reliable questionnaire is objectivity, which holds only if the questionnaire is independent of who made it. There are several ways to measure reliability, and Cronbach's alpha is the most important and widely used technique for estimating or measuring it.
Cronbach's alpha is used to assess the reliability of a set of scales or items. It specifies the measurement range of internal consistency. "Cronbach's alpha is calculated by correlating the score for each scale item with the total score for each observation and then comparing that to the variance for all individual item scores." [6] . Equation 6 shows its measurement:

$$\alpha = (\frac{k}{k-1})(1 - \frac{\sum_{i=1}^{k} \sigma_{y_i}^2}{\sigma_x^2}) \tag{6}$$

k is the number of factor items

[6] https://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/

$\sigma_{y_i}^2$ is the variance of item i
σ_x^2 is the variance of the observed total scores. (First, the sum of all the answered items per respondent is computed. Then the total variance of them will be calculated.)
For example, the Experiential Aesthetic factor has 8 items, so k would take the value 8 here.

The Cronbach's Alpha obtained with Equation 6 shows the degree of reliability. Usually, the range of this number is between 0 and 1. If Cronbach's alpha is zero, it means that the item scales are independent and have no covariance or any correlation with each other. And if this Cronbach's alpha is equal to one, it means the highest degree of correlation between existing item scales. The interpretation and relationship between Cronbach's alpha and internal consistency are shown in table 21.

Cronbach's alpha	**Internal consistency**
$0.9 \leq \alpha$	Excellent
$0.8 \leq \alpha \leq 0.9$	Good
$0.7 \leq \alpha \leq 0.8$	Acceptable
$0.6 \leq \alpha \leq 0.5$	Questionable
$0.5 \leq \alpha \leq 0.6$	Poor
$\alpha < 0.5$	Unacceptable

Table 21: Cronbach's Alpha and internal consistency [21]

The Extended-PSSUQ showed high reliability on each factor. The reliabilities are:

- System Quality: 0.878
- Information Quality: 0.826
- Interface Quality: 0.803
- Overall PSSUQ V3.0: 0.899
- Experiential Hedonic (extension): 0.9
- Experiential Aesthetics (extension): 0.93
- Overall Extended-PSSUQ (extension): 0.898

Table 21 indicates the relation between Cronbach's alpha and internal consistency. As seen in the bullet list above, all reliabilities exceed 0.8 which indicates sufficient reliability for standardized usability and user experience measurements. The overall factor is almost equally reliable in both, PSSUQ V3.0 and the extended version. The two new factors which measure Experiential Hedonic and Experiential Aesthetics even have higher values for reliability.

4.4.3 Validity

The validity of a test determines to what extent that test measures what it claims to measure. The validity of a questionnaire is a criterion used to measure the degree of how well the items fit the product or service under investigation. The validity checks each item in the questionnaire is in line with the purpose of the research. Reliability is a prerequisite for assessing validity. Concurrent validity is a method used in order to calculate PSSUQ and Extended-PSSUQ validity. A brief explanation of the concurrent validity and its interpretation is described here.

Concurrent Validity: It is used to assess a set of interrelated variables, which is obtained by measuring the coefficient of correlation between factors and the overall score obtained. Researchers often use the Pearson correlation coefficient to measure concurrent validity. The correlation of coefficient is the measurement of linear association between two variables. It is a relationship between the measure of interest and another criterion with the same interest or different prediction. Pearson correlation coefficient is obtained according to equation 7.

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{7}$$

r_{XY} refers to the correlation between two variables X and Y
$\overline{X}$ refers to the average of all X variables
$\overline{Y}$ refers to the average of all Y variables
n refers to the number of respondents

Pearson correlation coefficient ranges between -1 and 1. If this value is zero, it means that there is no correlation between the variables. Of course, in terms of the questionnaire's validity, if the Pearson correlation coefficient is between $r(x) = 0.3$ and $r(x) = 0.4$, it is large enough to use it as a validity criterion. P-value which is mentioned in results is significant level of the correlation coefficient.
For a sample of 21 participants who answered both Extended-PSSUQ and ASQ questionnaires in a usability and user experience study, the Overall Extended-PSSUQ score correlated highly with the sum of the ASQ ratings (r(21)=0.624, p = 0.003). Additionally, the Overall PSSUQ score correlated highly with the sum of the ASQ ratings (r(21)=0.561, p = 0.01). Table 22 shows the correlated overall scores of PSSUQ and Extended-PSSUQ with other factors.
According to the table 22, the overall PSSUQ score correlated significantly with the System Quality (r(21)=0.947, p = 0.01) and Information Quality (r(21)=0.867, p = 0.01) factors in comparison to Overall Extended-PSSUQ results with these two factors.

Also, Overall Extended-PSSUQ score correlated considerably with interface Quality (r(21)=0.547, p = 0.010) factor in contrast to overall PSSUQ score(r(21)=0.449, p = 0.041). And, both Overall Extended-PSSUQ score and overall PSSUQ score correlated highly with each other. Also, the Overall Extended-PSSUQ has remarkable correlation with Experiential Hedonic (r(21)=0.51, p = 0.018) and Experiential Aesthetics (r(21)=0.771, p = 0.01).

Sub-scores	**Overall Extended-PSSUQ**		**Overall PSSUQ**	
	r(21)	**p**	**r(21)**	**p**
System Quality	0.654	0.001	**0.947**	0.01
Information Quality	0.631	0.002	**0.867**	0.01
Interface Quality	**0.547**	0.010	0.449	0.041
Experiential Hedonic	0.51	0.018	-	-
Experiential Aesthetics	0.771	0.01	-	-
Overall PSSUQ	0.777	0.01	-	-

Table 22: The correlated scores of the Overall PSSUQ V3.0 and Overall Extended-PSSUQ V3.0 with other factors

5 Conclusion and Future work

In this thesis, the importance of product usability and the consumers' user experience were discussed. These two aspects together affect the economic development of companies, their activities, and decisions. Then, the important role of data collection, especially the first-party data, according to product usability and consumer experience, was considered. And the questionnaires were studied as a tool for first-party data collection. Since this thesis's subject is the analysis of current usability and user experience questionnaires, eight research questions in this thesis were raised 1.1, which were answered in different parts of this master thesis. A brief answer to each of the questions is summarized in this section.

According to the question 1 and 2, in this thesis, seven popular questionnaires were examined. Four questionnaires QUIS, SUMI, PSSUQ and SUS in the usability field and UEQ, AttrakDiff2, and meCUE in the user experience field were studied. The application of all three user experience questionnaires is interactive products. In the field of usability questionnaires, the SUMI questionnaire is applied for the software application. And three other questionnaires are used to study computer software, various products, and systems usability. In addition to the last usability use cases, only the PSSUQ questionnaire also considers the website usability features. The number of items, item style, and rating scale of these questionnaires were examined by question 4, which can be seen briefly in table 2. According to question 5, the questionnaire was examined based on their number of items, length, and shortness. It was found that QUIS and SUMI questionnaires are the longest, and the SUS questionnaire with only ten items is the shortest questionnaire in this group.

Based on question 6, each questionnaire consisted of a set of factors. Each factor was studied separately. Among these, meCUE and SUMI questionnaires with 8 and 6 factors, respectively, contain the highest number of factors. Further information about the factors can be seen in tables 3 and 4.

According to questions 3, 7, and 8, the questionnaires were studied in terms of their items' content. It was examined based on instrumental qualities, non-instrumental qualities, short-term affective response, and long-term evaluative response. It was found that the user experience questionnaires consider all three characteristics. But the usability questionnaires examine only the instrumental quality feature, and the application of the other two features are very few. Table 14 shows the application of each feature in each questionnaire.

The objective 1 of this thesis examines the different characteristics of each questionnaire separately. The user experience questionnaires were more optimal than the usability questionnaires regarding items' content and various factors. They examined instrumental qualities, non-instrumental qualities, and short-term affective response, and

long-term evaluative response. Also, these questionnaires were long enough, and they do not make their respondent bored. But on the other hand, usability questionnaires, as mentioned, study the instrumental qualities more. Therefore, it was decided to maintain the usability questionnaire structure and only by adding more items related to non-instrumental qualities, short-term affective response, and long-term evaluative response. It leads to examine more quality features. Among the usability questionnaires, QUIS and SUMI had many items and adding new items discouraged and bored the respondent. But PSSUQ and SUS questionnaires were good options for increasing the number of items. But the SUS questionnaire, despite its short length, was intended only to examine the overall satisfaction factor, and the addition of new items does not lead to significant results. But the PSSUQ questionnaire with 16 items examined four factors, and it was long enough to add new items. Therefore, it was decided to add twelve items, including four items in short-term affective response and long-term evaluative response (Experiential-Hedonic) and eight non-instrumental qualities (Experiential-Aesthetics) items, to optimize the application of this questionnaire in both usability and user experience, and to cover both areas. Therefore, the Extended-PSSUQ questionnaire was designed with 28 items and six factors.

According to objective 3 of the thesis, the Extended-PSSUQ questionnaire was distributed among the students of Otto-von-Guericke-University to evaluate the Mattermost software. This software is used as a communication tool for E-learning system in this university.

Then, based on objective 2, the Extended-PSSUQ questionnaire was evaluated according to Means and 95% confidence Intervals. The mean of all items based on 21 questionnaire respondents fall below the scale midpoint of 4, except for item 19 which is has a higher mean value. Also, the factor's norms were studied; all the mean and upper limits of 95% confidence intervals are lower than the midpoint of 4. Then, the factor analysis of the added items was applied by Principal Components Analysis and oblique rotation. Factor analysis confirmed the presence of two additional factors. Also, items 17 to 20 were hanged together, and linear composition coefficients of items 21 to 28.

Then, the reliability and validity of the Extended-PSSUQ questionnaire were assessed. All six factors have acceptable reliability above 0.8. Although the overall factors of PSSUQ and Extended-PSSUQ are equal, the two factors of Experiential Hedonic and Experiential Aesthetics have a reliability score of 0.9 and 0.93, respectively. On the other hand, to check the validity of the Extended-PSSUQ questionnaire, the ASQ questionnaire was given to the respondents in parallel with the Extended-PSSUQ questionnaire for responding. According to the concurrent validity results, the Extended-PSSUQ factor is correlated highly with the sum of the ASQ ratings ($r(21)=0.624$, $p = 0.003$). Also, the overall PSSUQ is correlated highly with the sum of the ASQ ratings ($r(21)=0.561$, $p = 0.01$). This result indicates that the overall Extended-PSSUQ factor

is better than the overall PSSUQ factor and has a higher validity. According to table 22, the validity of the Experiential Hedonic and Experiential Aesthetics factors have a high correlation with the overall Extended-PSSUQ factor respectively ($r(21)=0.51$, $p = 0.018$) and ($r(21)=0.771$, $p = 0.01$). In general, the Extended-PSSUQ questionnaire has acceptable good results according to reliability and validity.

5.1 Limitations

There were limitations in this thesis that better results might have been achieved if these limitations were removed. Including:

- One of the limitations of this survey is the low response rate. The number of respondents to the Extended-PSSUQ questionnaire is 21 people. This number is suitable for the first evaluation. However, to evaluate the performance of this questionnaire more accurately, more respondents had to answer, which was not happened due to the lack of familiarity with the user of Mattermost software in the Otto-von-Guericke-University.
- Due to this questionnaire's online distribution, it was not possible to review the answers with the respondents. There are doubts that respondents understood each item's meaning and concept exactly and chose the appropriate answer. Incorrect and incomprehensible responses affect the quality of the data collection and also impact the evaluation result.
- Mattermost software was not used in most departments at the Otto-von-Guericke-University and was used only in the informatics department. It would be better to use a more popular and practical software or digital product with more users.
- For more precise questionnaire evaluation, it should be republished again and again for different respondents and different products. And, if shortcomings exist in its items and factors, the deficiencies are eliminated, and the questionnaire's performance will be improved. Due to lack of time, this process was not repeated for more people and other products.

5.2 Future works and directions

One of this project's future tasks is to republish the Extended-PSSUQ questionnaire for more people and other products to study and evaluate the strengths and weaknesses of this questionnaire. Also, because there is an item that considers audio notification quality in this questionnaire, one of this questionnaire's future applications is to assess computer games' usability and user experience.
Questionnaires are a powerful tool for collecting data about the product and its users. They are usually reviewed for factor analysis, and the quality of their results is assessed

based on reliability and validity scores. However, these results alone can not accurately measure the usability of a product and the user experience. Therefore, new methods for extracting information from the collected data should get the highest efficiency. Effective data analysis of questionnaires leads to designing a better questionnaire structure. Also, extracting knowledge from collected data leads to companies' better decisions about improving their product usability and obtaining a more satisfying user experience. Therefore, since the present age is the age of artificial intelligence, it is realistic to use artificial intelligence techniques in the design stage and optimizing the extraction of knowledge from data.

One of the proposed future tasks for extracting information from the questionnaire collected data is artificial neural networks. First, training the neural network with the questionnaire collected data and then using artificial neural networks as a predictor or classifier system to extract the trained neural network rules. Then, according to the weights and activated patterns of the neural network's hidden and output layers, a decision tree will be designed, which can give output using these extracted rules.

References

[1] M. Thüring and S. Mahlke, "Usability, aesthetics and emotions in human–technology interaction," *International journal of psychology*, vol. 42, no. 4, pp. 253–264, 2007.

[2] K. Moumane, A. Idri, and A. Abran, "Usability evaluation of mobile applications using iso 9241 and iso 25062 standards," *SpringerPlus*, vol. 5, 12 2016.

[3] J. Sauro and J. R. Lewis, "Correlations among prototypical usability metrics: evidence for the construct of usability," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1609–1618, 2009.

[4] J. Sauro and J. R. Lewis, *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016.

[5] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[6] R. Horst and R. Dörner, "Virtual reality forge: Pattern-oriented authoring of virtual reality nuggets," in *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–12, 2019.

[7] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Design and evaluation of a short version of the user experience questionnaire (ueq-s).," *IJIMAI*, vol. 4, no. 6, pp. 103–108, 2017.

[8] M. Minge, M. Thüring, I. Wagner, and C. V. Kuhr, "The mecue questionnaire: a modular tool for measuring user experience," in *Advances in Ergonomics Modeling, Usability & Special Populations*, pp. 115–128, Springer, 2017.

[9] A. Hinderks, D. Winter, M. Schrepp, and J. Thomaschewski, "Applicability of user experience and usability questionnaires.," *J. UCS*, vol. 25, no. 13, pp. 1717–1735, 2019.

[10] E. L.-C. Law, P. Van Schaik, and V. Roto, "Attitudes towards user experience (ux) measurement," *International Journal of Human-Computer Studies*, vol. 72, no. 6, pp. 526–541, 2014.

[11] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 213–218, 1988.

[12] J. Kirakowski and M. Corbett, "Sumi: The software usability measurement inventory," *British journal of educational technology*, vol. 24, no. 3, pp. 210–212, 1993.

[13] R. McSweeney, "Sumi–a psychometric approach to software evaluation," *Unpublished MA (Qual) thesis in Applied Psychology, University College Cork, Ireland*, 1992.

[14] J. R. Lewis, "Psychometric evaluation of the post-study system usability questionnaire: The pssuq," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 36, pp. 1259–1260, SAGE Publications Sage CA: Los Angeles, CA, 1992.

[15] J. R. Lewis, "Ibm computer usability satisfaction questionnaires: psychometric evaluation and instructions for use," *International Journal of Human-Computer Interaction*, vol. 7, no. 1, pp. 57–78, 1995.

[16] J. R. Lewis, "Psychometric evaluation of the pssuq using data from five years of usability studies," *International Journal of Human-Computer Interaction*, vol. 14, no. 3-4, pp. 463–488, 2002.

[17] M. Hassenzahl, M. Burmester, and F. Koller, "Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität," in *Mensch & computer 2003*, pp. 187–196, Springer, 2003.

[18] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Symposium of the Austrian HCI and Usability Engineering Group*, pp. 63–76, Springer, 2008.

[19] M. Minge and L. Riedel, "mecue-ein modularer fragebogen zur erfassung des nutzungserlebens," *Mensch & Computer 2013: Interaktive Vielfalt*, 2013.

[20] T. K. Chung and N. Sahari, "Utilitarian or experiential? an analysis of usability questionnaires," *International Journal of Computer Theory and Engineering*, vol. 7, no. 2, pp. 167–171, 2015.

[21] F. Petermann, "Bühner, m.(2006). einführung in die test-und fragebogenkonstruktion," *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, vol. 57, no. 3, pp. 227–228, 2009.

[22] I.-D. Anič, A.-C. Suleska, and E. Rajh, "Decision-making styles of young-adult consumers in the republic of macedonia," *Economic research-Ekonomska istraživanja*, vol. 23, no. 4, pp. 102–113, 2010.

[23] J. Kirakowski and A. Dillon, "The computer user satisfaction inventory (cusi): Manual and scoring key," *Cork, Ireland: Human Factors Research Group, University College of Cork*, 1988.

[24] D. Norman, J. Miller, and A. Henderson, "What you see, some of what's in the future, and how we go about doing it: Hi at apple computer," in *Conference companion on Human factors in computing systems*, p. 155, 1995.

[25] N. Bevan, J. Carter, and S. Harker, "Iso 9241-11 revised: What have we learnt about usability since 1998?," in *International Conference on Human-Computer Interaction*, pp. 143–151, Springer, 2015.

[26] D. Berry, "The user experience: The iceberg analogy of usability," *Retrieved August*, vol. 19, p. 2007, 2000.

[27] W. Iso, "9241-11. ergonomic requirements for office work with visual display terminals (vdts)," *The international organization for standardization*, vol. 45, no. 9, 1998.

[28] I. DIS, "9241-210: 2010. ergonomics of human system interaction-part 210: Human-centred design for interactive systems (formerly known as 13407)," *International Standardization Organization (ISO). Switzerland*, 2010.

[29] A. NCITS, "354: Common industry format for usability test reports. american national standards institute," 2001.

[30] C. Lallemand and V. Koenig, "How could an intranet be like a friend to me? why standardized ux scales don't always fit," in *Proceedings of the European Conference on Cognitive Ergonomics 2017*, pp. 9–16, 2017.

[31] J. Baumgartner, A. Sonderegger, and J. Sauer, "No need to read: Developing a pictorial single-item scale for measuring perceived usability," *International Journal of Human-Computer Studies*, vol. 122, pp. 78–89, 2019.

[32] Y. Forster, S. Hergeth, F. Naujoks, and J. F. Krems, "How usability can save the day-methodological considerations for making automated driving a success story," in *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 278–290, 2018.

[33] A. Marcus and W. Wang, *Design, User Experience, and Usability: Users, Contexts and Case Studies: 7th International Conference, DUXU 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings*, vol. 10920. Springer, 2018.

[34] I. Díaz-Oreiro, G. López, L. Quesada, and L. A. Guerrero, "Standardized questionnaires for user experience evaluation: A systematic literature review," in *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 31, p. 14, 2019.

[35] M. Hassenzahl, F. Koller, and M. Burmester, "Der user experience (ux) auf der spur: Zum einsatz von www. attrakdiff. de," *Tagungsband UP08*, 2008.

[36] S. H. Lee, "Constructing effective questionnaires," *Handbook of human performance technology, Hoboken, NJ: Pfeiffer Wiley*, pp. 760–779, 2006.

Appendix

Additional Words	Similar Words	Attribute
Clear	Well-Defined, Obvious, Tidy	U
Comfortable	Contended, Relaxed, Happy, Easy,Calm	U
Complex	Difficult, Complicated	U
Cumbersome	Awkward, Weighty, Burdensome	U
Confident	Self-assured, cool	EH
Control	Management, Power, Charge	U
Consistent	Reliable, Steady, Constant, Stable, Coherent, Even, Uniform	U
Difficult	Problematic, In comprehensive, Hard,Obstinate	U
Effective	Operational, Applicable, Successful, Useful	U
Exploring	Investigating,Searching	U
Feel	Experience, Think, Believe	EH
Flexible	Supple, Malleable, Malleable, Compliant, Docile	U
Improve	Increase, Enhance, Enrich	U
Like to use	Prefer, choose, approve	EH
Learning	Studying	U
Memorable	Unforgettable, Notable, Striking,Impressive, Remarkable	U
Quickly	Rapidly, Fast, Speedily, Swiftly	U
Pleasant	Enjoyable, Pleasing, Satisfying, Friendly, Nice	EH
Productive	Helpful, Beneficial, Useful	U
Recommend	Suggest, Propose	EH
Rigid	Unbending, Inflexible, Firm	U
Standard	Normal, Usual, Regular	U
Terrible	Awful, Horrible, Severe	EH

Table 23: The additional words generated from MICROSOFT word's thesaurus [20]

	The Extended Post Study System Usability Questionnaire	Trifft völlig zu							Trifft gar nicht zu	
			1	2	3	4	5	6	7	NA
1	Insgesamt bin ich damit zufrieden, wie einfach dieses System zu benutzen ist.		☐	☐	☐	☐	☐	☐	☐	☐
2	Es war einfach, dieses System zu benutzen.		☐	☐	☐	☐	☐	☐	☐	☐
3	Ich war in der Lage, die Aufgaben mit diesem System schnell zu lösen.		☐	☐	☐	☐	☐	☐	☐	☐
4	Ich fühlte mich beim Verwenden dieses System wohl.		☐	☐	☐	☐	☐	☐	☐	☐
5	Es war leicht, die Bedienung dieses Systems zu erlernen.		☐	☐	☐	☐	☐	☐	☐	☐
6	Ich glaube, dass ich durch die Verwendung dieses Systems schnell produktiv werden könnte.		☐	☐	☐	☐	☐	☐	☐	☐
7	Das System gab Fehlermeldungen aus, die klar erklärten, wie Probleme gelöst werden können.		☐	☐	☐	☐	☐	☐	☐	☐
8	Fehler konnte ich einfach und schnell korrigieren.		☐	☐	☐	☐	☐	☐	☐	☐
9	Die Informationen (wie Direkthilfe, Bildschirm-Meldungen und andere Dokumentation), die das System zur Verfügung gestellt hat, waren klar verständlich.		☐	☐	☐	☐	☐	☐	☐	☐
10	Es war leicht, benötigte Informationen zu finden.		☐	☐	☐	☐	☐	☐	☐	☐
11	Die Informationen halfen mir erfolgreich dabei, die Aufgaben zu lösen.		☐	☐	☐	☐	☐	☐	☐	☐
12	Die Anordnung der Informationen auf dem Bildschirm war logisch.		☐	☐	☐	☐	☐	☐	☐	☐
13	Das Nutzerinterface ist angenehm.		☐	☐	☐	☐	☐	☐	☐	☐
14	Ich benutze das Nutzerinterface gerne.		☐	☐	☐	☐	☐	☐	☐	☐
15	Das System bietet alle Funktionen und Möglichkeiten, die ich von ihm erwarte.		☐	☐	☐	☐	☐	☐	☐	☐
16	Insgesamt bin ich mit diesem System zufrieden.		☐	☐	☐	☐	☐	☐	☐	☐
17	Das System lässt sich intuitiv bedienen.		☐	☐	☐	☐	☐	☐	☐	☐
18	Das System macht Lust, meine Aufgaben zu lösen.		☐	☐	☐	☐	☐	☐	☐	☐
19	Mir kamen spontan Ideen in den Sinn, wie man das System noch verwenden könnte.		☐	☐	☐	☐	☐	☐	☐	☐
20	Ich bevorzuge dieses System gegenüber vergleichbaren.		☐	☐	☐	☐	☐	☐	☐	☐
21	Die farbliche Gestaltung des Systems sieht gut aus.		☐	☐	☐	☐	☐	☐	☐	☐
22	Die Audiomitteilungen des Systems klingen cool.		☐	☐	☐	☐	☐	☐	☐	☐
23	Das System löst das Problem vernünftig.		☐	☐	☐	☐	☐	☐	☐	☐
24	Das System fügt sich gut in andere Systeme ein.		☐	☐	☐	☐	☐	☐	☐	☐
25	Das Systemlayout ist stylish.		☐	☐	☐	☐	☐	☐	☐	☐
26	Das System kann das Problem geschickt lösen.		☐	☐	☐	☐	☐	☐	☐	☐
27	Das System wird ansprechend präsentiert.		☐	☐	☐	☐	☐	☐	☐	☐
28	Ich kann ein innovatives Feature des Systems benennen.		☐	☐	☐	☐	☐	☐	☐	☐

Figure 20: The Extended-PSSUQ questionnaire German version