PET-USES: Privacy-Enhancing Technology - Users' Self-Estimation Scale

Erik Wästlund, Peter Wolkerstorfer, Christina Köffel

PET-USES: Privacy-Enhancing Technology – Users' Self-Estimation Scale

Erik Wästlund[1], Peter Wolkerstorfer[2] and Christina Köffel[2]

[1] Dep. of Psychology – Karlstad University, Sweden
erik.wastlund@kau.se
[2] CURE – Center for Usability Research and Engineering, Vienna, Austria
wolkerstorfer@cure.at, koeffel@cure.at

Abstract. This paper describes the "Privacy-Enhancing Technology Users' Self-Estimation Scale (PET-USES)", a questionnaire that enables users to evaluate PET user interfaces for their overall usability and to measure six different PET aspects. The PET-USES is intended to be used during usability testing and evaluation of PET user interfaces. The focus of the PET-USES is the subjective experience of the user rather than the intrinsic PET functionality of the application being tested. Although the test has been developed within the PrimeLife[1] project to test the usability of PETs developed therein, the test is constructed in such a fashion that it should be applicable to a wide variety of PETs. The objective of this paper is to outline the creation and the background of the PET-USES questionnaire and invite the usability community not only to use the test, but also to contribute to the further development of the PET-USES.

Keywords: PET-USES, HCI, Usability, PET

1 Introduction

PET-USES (Privacy-Enhancing Technology Users' Self-Estimation Scale) is a questionnaire that enables users to evaluate PET User Interfaces (UIs). The reason for developing and using PET-USES was to be able to measure the perceived usability of UIs, both during single user trails and during large group walkthroughs of screen recordings.

Today there are a number of questionnaires measuring user experience, usability and various HCI (Human-Computer Interaction) aspects such as the hedonic quality [1] of software, websites, and services [2, 3], to our knowledge none includes PET-related issues.

[1] The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 216483.

Although there is no single widely accepted definition of PETs, they can be described as:

> *Privacy-Enhancing Technologies is a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system.* [4]

This above definition is focused on the principle of data minimization whereas others focus more on privacy principles and legislation or how PETs give the user power over his/her own data [5]. There are also attempts to classify PETs into classes such as General PET Controls, Separation of Data, Privacy Management Systems, and Anonymisation Tools [6].

One PET solution currently being investigate within the PrimeLife[2] project is an identity management system that solves a number of the above mentions issues and the usability evaluation of this system is the stepping stone for the development of the PET-USES. In short the system is comprised of a number of subsystems which relate to the handling and release of an individual's personal data. The PrimeLife system also informs the user of the trustworthiness of data recipient and to what extent the data recipient's privacy policy matches the desired privacy policy of the data subject.

The PET-USES consists of two major parts of questions: one part measuring overall usability and one part measuring PET aspects. Thus, the PET usability scales have a dual purpose. They evaluate the software's general usability and the extent to which the software assists the user in learning and understanding privacy related issues. An important feature of the measurement of PET aspects is the modularity of the questionnaire, enabling the inclusion or exclusion of scales measuring specific aspects based on the tasks and features being evaluated.

This text is organized as follows: Section 2 "Related Work" depicts current views on usability testing. The following Section 3 "The PET-USES Approach" describes development criteria for the PET-USES questionnaire and sketches the main modules. This is followed by a discussion in Section 4 on when and how the PET-USES can be employed. Section 5 gives a conclusion and shows next steps in testing the perceived usability of some PETs with the PET-USES questionnaire which version 1.0 is published in the Appendix.

2 Related Work

The PET-USES questionnaire is based on the ISO 9241 general standard of usability [7] as well as the more PET specific HCI guidelines presented by Patrick et al. 2003 [8] and utilized in the work with the PRIME[3] integrated identity management prototype [9]. The former defines usability as the "extent to which a product can be

[2] "PrimeLife – bringing sustainable privacy and identity management to future networks and services" is an EC FP7 project: http://www.primelife.eu/

[3] "PRIME – Privacy and Identity Management for Europe" was an EC FP6 project: http://www.prime-project.eu/

used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction" whereas the latter promotes the four categories comprehension (to understand or know), consciousness (be aware or informed), control (to manipulate or be empowered) and consent (to agree). Although the two views might seem divergent at first they can readily be combined within the structure of usability testing proposed by Hornbæk [10]. Based on a review of 180 studies, published in core HCI journals and proceedings, he argues for a change in terminology from the ISO 9241, to better encompass what is actually being measured. The relationship between the concepts of Hornbæk, ISO 9241, HCI Guidelines and generally often used measures of usability can be seen below (Table 1).

Table 1. The table shows possible constructs of interest for the PET-USES, their relationship to other usability constructs and how they relate to the framework proposed by Hornbæk.

Hornbæk	ISO 9241	HCI Guidelines[4]	Other measures/concepts of usability
Outcomes	Effectiveness	Consent (agree) Comprehension (to understand or know) Consciousness (be aware of, be informed)	User Value Usefulness Functionality
Interaction-Process	Efficiency	Control (to manipulate or be empowered)	Efficiency Ease of Learning Ease of Use
Attitudes & Experiences	Satisfaction		Satisfaction Affect / Likeability Trust Helpfulness Awareness of PET-Related Issues

Thus, by using the terminology of Hornbæk, one can for instance investigate the outcomes of using a particular interface in terms of Effectiveness of Goal Completion but also in terms of User Value and what the user learns from the interaction. This framework makes it easy to integrate the above-mentioned constructs into one model as well as adding further constructs if that should be deemed necessary.

3 The PET-USES Approach

The PET-USES scale General Usability is measured as a composite of the sub-scales Ease of Learning, Ease of Use and User Value. The rationale for differentiating between the sub-scales Ease of Learning and Ease of Use is that intuitive interfaces are perceived to have a better learnability whereas a less intuitive interface can be used easily only once the user gets accustomed to it. It is also noteworthy that the

[4] As defined by Patrick et al. 2003 [8].

General Usability value will be less influenced by perceived User Value than Ease of Learning and Ease of Use. This reflects the fact that, although user value is an important driver for software adoption the focus of the PET-USES lies more on the usability than on the perceived benefits of a system.

The PET aspects modules currently developed are derived from the user-controlled identity management approach of the projects PRIME and PrimeLife: Data Management, Credential Management, PrivPrefs[5], Recipient Evaluation, Data Release, and History. They can all be used to evaluate specific PET-related functionality of software or websites. The entire PET-USES questionnaire (including all modules) and its items can be found in the appendix.

The focus of the scales are the following privacy-critic areas:

- Data Management: The extent to which the system makes it easier to store and organize personal information. This scale can be used to evaluate all types of identity management software and services.
- Credential Management: The extent to which the system makes it easier to store and organize credentials and other certificates. This scale can be used to evaluate identity management systems that include issued claim credentials (e.g. the Higgins project[6]).
- PrivPrefs: This scale is designed to measure the extent to which the system makes it easier to set general and excessive levels for data release policies and to what extent the user is informed of unwanted data dissemination. Thus, an aspect of this scale is the decision support qualities of the system.
- Recipient Evaluation: the extent to which the system helps users to evaluate the data recipients' credibility and trustworthiness. This scale can also be regarded in terms of decision support.
- Data Release: The extent to which the system clarifies what personal information is being released and who is the recipient of the data.
- History: The extent to which the system can show the user when, what and, to whom personal information has been released and thus provide an overview of what data any given service provider might have accumulated.

Effectiveness and efficiency are often measured in a more objective fashion than the user self-estimations of the PET-USES. The effectiveness of a given interface can for instance be measured in terms of task completion time and efficiency in terms of quality of task solution [11] and, of course, optimally usability evaluations should be comprised of a combination of self-estimation and more objective measurements. It should, however, be pointed out that these types of measurement requires fully functional interfaces and both logging of behavior and knowledge of desired outcomes whilst the PET-USES can be used in a much earlier stage to measure users perception as estimates of effectiveness and efficiency.

---

[5] PrivPrefs (Privacy Preferences) is a method that is currently being investigated in the PrimeLife project for defining personal privacy preferences (see for example [12]) which will be used for automated evaluations of the appropriateness of data-requests. The PrivPrefs are evaluations of polices as defined in P3P (http://www.w3.org/TR/2006/NOTE-P3P11-20061113/#P3PPolicies).

[6] http://wiki.eclipse.org/Password_Cards#Required_Claim_Types

Practical considerations such as time and effort to answer the questions can prevent the PET-USES to measure all of the categories mentioned in Table 1 in separate scales and hence several of the categories will have to be combined into more general domains.

4 Discussion

A fundamental principle of self-estimation scales is that all questions are thought to measure an aspect of construct of interest. Thus, all questions are thought to be exchangeable with other questions that also measure a related aspect of the same construct. An important aspect of this fact is that the focus of measurement is rather the aggregated data of all the questions of a module than on the specific questions themselves. This idea, of course, is analogous to any type of sampling and point estimation. As with any sampling procedure more observations lead to better estimations. However, when it comes to self-ratings, time constraints are the biggest obstacles to extensive sampling.

In essence, all self rating scales are constructed in a similar fashion containing a stimuli and some way for the participant to rate this stimuli. The PET-USES is constructed as a number of Likert scales [13]. Thus, in accordance with the principle of Likert scales, the stimuli used are a number of statements and users are asked to rate to what extent they agree or disagree with these statements. The response format used in the PET-USES is a five point scale. Thus it is possible for the user to respond in a neutral fashion unlike in a forced choice scale. It should be noted that it is possible to utilize other response formats such as any number of values or a Visual Analogue Scale [14]. There is quite some debate (see for example [13, 15]) over the level of measurement of the added values of the Likert scales and if they should be treated as ordinal or interval data. The basic argument for viewing the scales as being ordinal is that it is impossible to create a subjective scale with equal distances between response options. On the other hand, it is possible to argue that there are in fact equal distances between the response options as respondents are using the numbers one to five, not the verbal descriptions. The main reason for purporting the notion of the scales being interval is of course the possibility to use parametrical tests.

4.1 Evaluating scales

All measurement needs to be evaluated in terms of reliability and validity. As the individual questions of a scale are thought to measure the same construct, the most fundamental evaluation of a scale is one of internal consistency. The basic principle is that respondents should answer the questions in a coherent manner, that is, if a respondent scores high on one statement of a given scale s/he should score rather high on other statements measuring the same underlying construct. If this is not the case, the items are thought to measure different constructs. Additionally, as different underlying constructs are supposed to be independent from each other, items measuring different underlying constructs should not correlate highly. The statistical

technique used to estimate internal consistency is Chronbach's alpha and factor analysis to assess the underlying constructs as such [16].

Tests such as Chronbach's Alpha, however, say nothing about what the test actually measures even though it might measure it satisfactory. In order to understand what a scale actually measures we need to assess its external validity. As some aspects of the PET-USES measure constructs which are also possible to measure in other ways, the questionnaire should be evaluated against these criteria. For instance, as the sub-scales Ease of Learning and Ease of Use both are aspects of Efficiency they could be correlated with measurements of quality of solutions or such.

4.2 When to use the PET-USES

The main reason for conducting usability tests is to discriminate between usable and not usable interfaces either during the design process or in comparisons between different systems. Typical use-cases for the PET-USES include both of these scenarios. Thus, PET-USES can be used both in order to compare the perceived usability strengths and weaknesses between different interfaces, and, in order to aid interface designers during the design process through administrating the test at various steps in the process. However, as during all statistical testing, the possibility to find significant results is dependent on the power of the investigation. As usual there are only two ways to achieve statistical power: a bigger sample or a bigger effects size. When it comes to comparing existing interfaces a bigger effects size can be achieved both by choosing interfaces that are evaluated as being extremely good and bad and by inviting more of the current user base into the evaluation. During interface design, especially during fast iterations, the differences between versions are usually quite small and the tested user group rather small and hence the power of a test such as the PET-USES will become quite small. This should be taken into consideration when planning when to use the PET-USES as it will be more useful evaluating clear steps in the design process. In order to gain power by adding more respondents without having to do a great number of complete user tests it is possible to do large group walkthroughs of screen recordings. An additional feature of this method is that it is possible to do user tests on interfaces without any functionality.

So far the usage of PET-USES is rather limited, but it has been incorporated in usability studies performed by Center for Usability Research and Engineering (CURE). Although not enough data has yet been collected for formal statistical evaluation of the PET-USES, feedback from both practitioners and users show that the test is easy to use.

4.3 The CURE web service

In order to facilitate both the use and the evaluation of the PET-USES, a web service is set up at CURE[7]. The site enables research companies to use the PET-USES questionnaire for their evaluations and will be open to all who wish to use the PET-

[7] http://pet-uses.cure.at/

USES on the premises that the collected PET-USES data will be used to gather feedback and further develop the questionnaire and its scales. In addition to using the scales of the PET-USES researchers in this area will have the possibility to suggest new modules for inclusion in the sub-scale battery to reflect the ever changing field of PETs. Data provided on the website will be anonymized and treated confidentially. Only those conducting the research and the creators of the PET-USES (i.e. Karlstad University and CURE) will have access to the data provided. Users of the site who wish to retain data from other sources than the PET-USES are of course allowed to do so, but in order to evaluate the PET-USES users are encouraged to provide data, such as the maturity of the tested system or correlations with other measurements as a part of the validation of the test.

5 Conclusion

The PET-USES presented in this paper is a questionnaire which focuses on measures of both aspects of General Usability and specifically tailored scales that measure the usability of PET solutions. The test is grounded in current views on usability and the experience so far of using the test show that both practitioners and users report that the PET-USES is easy to use. The CURE web service for using the PET-USES is open to PET researchers who wish to evaluate PET UIs.

References

1. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität, In: Ziegler, J., Szwillus, G. (eds.): Mensch & Computer 2003, Interaktion in Bewegung, pp. 187--196 (2003)
2. Brooke, J.: SUS: a “quick and dirty” usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) Usability Evaluation in Industry, pp. 189--194. Taylor & Francis, London (1996)
3. Tullis, T.S., Stetson, J.N.: A Comparison of Questionnaires for Assessing Website Usability. Usability Professional Association Conference (2004)
7. ISO. Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11: guidance on usability-Part 11: guidance on usability (ISO 9241-11:1998) (1998)
8. Patrick, A.S., Kenny, S., Holmes, C., van Breukelen, M.: Human Computer Interaction. Handbook for Privacy and Privacy-Enhancing Technologies: College bescherming persoonsgegevens, pp. 249--290 (2003)
9. Pettersson, J.S. (ed.): HCI Guidelines, PRIME Deliverable, D6.1.c, https://www.prime-project.eu/prime_products/reports/arch/pub_del_D06.1.c_ec_wp06.1_V4_final.pdf (2005)
10. Hornbæk, K.: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Human-Computer Studies 64, pp. 79--102 (2006)
11. Frøkjær, E., Hertzum, M., Hornbæk, K.: Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? Proceedings of the ACM CHI 2000 Conference on Human Factors in Computing Systems, 1-6 April 2000, pp. 345--352, ACM Press, New York (2000)
13. Edmondson, D.: Likert Scales: A History. CHARM – the Conference on Historical Analysis and Research in Marketing, pp. 127--132 (2005)

14.Butler, P.V.: Linear Analogue Self-Assessment and Procrustean Measurement: A Critical Review of Visual Analogue Scaling in Pain Assessment. J. Clin. Psychol. Med. Settings 19, pp. 111--129 (1997)
15.Göb, R., McCollin, C., Ramalhoto, M.: Ordinal Methodology in the Analysis of Likert Scales. Quality and Quantity 41, pp. 601--626 (2007)
16.Anastasi, A., Urbina, S.: Psychological Testing. 7th edition. Prentice Hall, Upper Saddle River, NJ (1997)
4. van Blarkom, G.W., Borking, J.J., Olk, J.G.E.: PET. In: van Blarkom, G.W., Borking, J.J., Olk, J.G.E. (eds.) Handbook of privacy and privacy-enhancing technologies (the case of intelligent software agents). http://www.andrewpatrick.ca/pisa/handbook/handbook.html (2003)
5. Harbird, R.: Privacy enhancing technologies. In: Information Commissioner's Office, UK, Privacy by design. Wycliffe House, Cheshire (2008)
6. Koorn, R., van Gils, H., ter Hart, J., Overbeek, P., Tellegen, R.: Privacy-enhancing technologies: White paper for decision makers. The Netherlands: Ministry of Interior and Kingdom Relations (2004)
12.Bergmann, M.: Generic Predefined Privacy Preferences for Online Applications. In: Fischer-Hübner, S., Duquenoy, P., Zuccato, A., Martucci, L. (eds.) The Future of Identity in the Information Society Boston, pp. 259--273, Springer (2008)

Appendix: PET-USES [1.0]

Modules

The PET-USES questionnaire comprises the following modules (the detailed content can be seen in the Appendix):

Part I – Usability:
- Ease of Learning
- Ease of Use
- User Value

Part II – PET-related aspects:
- Data Management
- Credential Management
- PrivPrefs
- Recipient Evaluation
- Data Release
- History

Instructions

This test is designed to measure your experience with the system you've tested today. Your answers will be used to evaluate the system so please answer the questions as truthfully as you can. As the questions are designed to measure various aspects of the systems usability there are no right or wrong answers. Please use the scale below to indicate to what extent you disagree or agree to the statements that follow.

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree

General Usability

1. I found it easy to learn how to use the *system*	1 2 3 4 5
2. I had to learn a lot in order to use the *system*	1 2 3 4 5
3. I keep forgetting how to do things with this *system*	1 2 3 4 5
4. I need a lot of assistance to use this *system*	1 2 3 4 5
5. I find the *system* interface easy to use	1 2 3 4 5
6. I find the organisation of the *system* interface understandable	1 2 3 4 5
7. I get confused by the *system* interface	1 2 3 4 5
8. I find it very difficult to work with the *system*	1 2 3 4 5
9. I find that the benefits of using the *system* are bigger then the effort of using it	1 2 3 4 5

10. I would like to use this *system* regularly	1 2 3 4 5
Data Management	
11. I get a clear view of my personal *data* from the system	1 2 3 4 5
12. I find organising my personal *data* easy with this system	1 2 3 4 5
13. I find keeping track of various user names and passwords is easy with this *system*	1 2 3 4 5
Credential Management	
14. I find it easy to add personally issued credentials into the *system*	1 2 3 4 5
15. I find it easy to add / import certificates into the *system*	1 2 3 4 5
16. I find it easy to manage my credentials with this *system*	1 2 3 4 5
PrivPrefs	
17. I find it easy to use settings for how much or how little *data* to be released with this *system*	1 2 3 4 5
18. I find that the *system* helps me understand the effects of different privacy settings	1 2 3 4 5
19. I feel safer knowing that I will be notified by the *system* if I'm about to release more personal *data* than my chosen preference	1 2 3 4 5
Recipient Evaluation	
20. The *system* makes it easy for me to decide if it is safe to release my data	1 2 3 4 5
21. I don't understand how the *system* determines if a data recipient is trustworthy	1 2 3 4 5
22. I feel safer releasing my personal data when the *system* states it's OK	1 2 3 4 5
Data Release	
23. I know what personal information I'm releasing when I'm using this *system*	1 2 3 4 5
24. The system makes it easy to decide how much or how little *data* to release in a given transaction	1 2 3 4 5
25. I get help from the system to understand who will receive my *data*	1 2 3 4 5
History	
26. I can easily find out who has received my personal *data* with this *system*	1 2 3 4 5
27. I get a good view of who knows what about me from this *system*	1 2 3 4 5
28. I can easily see how much I've used a particular user name with this *system*	1 2 3 4 5

Headings and numerals are mainly for presentational purposes and thus optional during the use of PET-USES. Items 2, 3, 7, 8, and 21 should be reversed before summated.