Investigating Usability Metrics for the Design and Development of Applications for the Elderly

Andreas Holzinger[1], Gig Searle[1],
Thomas Kleinberger[2], Ahmed Seffah[3], and Homa Javahery[3]

[1] Medical University Graz, A-8036 Graz, Austria
Institute for Medical Informatics, Statistics & Documentation (IMI)
Research Unit HCI4MED
andreas.holzinger@meduni-graz.at,
gig.searle@meduni-graz.at
[2] Fraunhofer Institute for Experimental Software Engineering (IESE)
67663 Kaiserslautern, Germany
thomas.kleinberger@iese.fraunhofer.de
[3] Concordia University, Montreal, QC, Canada, H3G 1M8
Human-Centered Software Engineering Group
seffah@cse.concordia.ca,
h_javahe@cs.concordia.ca

Abstract. Metrics-based benchmarks are crucial for measuring usability, particularly for special end user groups such as older adults. So far, usability metrics that are accessible and useful for software developers are still missing. Although usability standards are continually being developed and adapted to the rapid change in both software and computing hardware, the increasing diversity of elderly populations, both culturally and educationally, requires the development of a specific set of criteria. This paper studies two different aspects of usability for the elderly; *passive* and *active* interaction. It explores the limitations of active interaction, its potential causes and results. The long term goal is the development of new methods to counteract potential negative bias with regard to passive interaction. More specifically, we are investigating achievable metrics for the evaluation of passive technology trustworthiness and usability while categorizing them according to applicability for usability testing.

Keywords: Human–Computer Interaction, Usability Engineering, Metrics, User Interfaces, Older Adults.

1 Introduction and Motivation for Research

There is accordance that in most industrialized countries the demographical, structural and societal trends are moving towards an increase in the elderly population and single households. This has dramatic effects on health care, emergency medical services and the individuals themselves [1]. Scientists agree that, besides new forms of living together (generations, communities), information technology (IT) is a promising way to cope with such effects to enable elderly to stay longer in their preferred environment. Consequently,

K. Miesenberger et al. (Eds.): ICCHP 2008, LNCS 5105, pp. 98–105, 2008.

IT must be developed to support user needs and emerging requirements. Whatever the circumstances and the degree of acceptance by a user of innovative technology, a clear benefit must be offered, whether in a physical, medical or emotional respect. Such an objective can be achieved at several levels:

- The design needs to be adapted to the end user's *physical impairments;*
- The interface must offer a relative degree of *familiarity* to overcome any reservations felt by the end user;
- The *benefit* of using the device must be appreciable, and the balance between intuitive use and practicable teaching methods, addressing the learning needs of this age group, must be established [2].

Quantitative research methods are common in most sciences. Based on the popular quote *"You cannot control what you cannot measure"* by DeMarco (1982) [3] computer scientists have tried to apply similar approaches to software development. Consequently, software metrics were developed to measure *some* properties of the software. However, although numerical precision is at the core of these measures, current practices that employ existing usability engineering methods are not numerically precise [4]. Sauro (2006) points out that this does not mean that usability is not a science and cannot be precisely measured; however, he also emphasizes that there is still much to be done [5]. During the past 15 years, the Human–Computer Interaction (HCI) community has developed a large variety of User–Centered Design (UCD) techniques for qualitatively assessing usability [6], [7], [8]. Although there are many of these methods available, they are underused and difficult to understand by software development teams and organizations [9]. This is because these techniques have been developed independently from the software engineering (SE) community, which has its own quantitative techniques and tools for measuring software quality including usability, e.g. experimental evaluations [20]. The obvious question is: Where should UCD techniques and the related usability measures be considered in the existing software development life cycle to maximize benefits gained from both approaches? Avoiding obstacles like those defined by Seffah & Metzker [10] is the first step towards a harmonious coexistence of UCD techniques and the entire software engineering life cycle. A first milestone to reach this goal is to have a consolidated usability metrics model [11] which will:

- Reduce the cost of usability testing by providing a basis for understanding and comparing various usability metrics;
- Complement more subjective, expert-based evaluations of usability;
- Provide a basis for clearer communication about usability measurement between software developers and usability experts;
- Promote usability measurement practices that are more accessible to software developers who usually have a limited background in HCI or Usability Engineering.

2 Expectations and Preconceptions

The basic use of computers is not beyond the understanding of the elderly, although this fact is not always appreciated by care givers [12]. However, the elderly are often of the opinion that the benefits associated with computer use fail to outweigh the necessary effort. One of the impediments faced by the elderly is not having grown up with computers or computer languages and some, when faced with the technological terminology frequently used in both handbooks and interfaces, feel overtaxed. Interfaces based on the experience, language and expectations of this group, which offer to support their special requirements while taking physical deficiencies into account, can supply the necessary incentives. Many of the devices on the market, which are designed for use by care givers for the prevention and alleviation of possible dangers to the elderly, require trust not only on the part of the care givers but also on the part of the elderly. This trust is given more easily when the beneficiaries understand and appreciate their functioning and benefit. The rewards of gaining a new skill have also proved to increase the quality of their own lives and extend the period of their independence [13].

Metrics for the evaluation of trustworthiness and acceptance of passive technology for the elderly must be approached from the view of the elderly. By working together with the end users, we are able quantify their questions in the form of usability metrics (see table 1).

Table 1. Analogies between user anxiety and metrics

Questions	Resulting Metric
Can I trust it	How can passive technology be made more **trustworthy**?
Can I switch it off/on?	How can we make it more **controllable**
Can I understand it?	How can we improve **understanding** of the principles and functionality, without too many confusing details
Will it obey me?	Can we remove the Frankenstein element; turn it from “magic” to machine, thereby inspiring **confidence**?
Who can see me?	Can we counteract the Big Brother element; replace the fear of *being* controlled with a feeling of *being in* **control**?
Do I really need this?	Explanation of benefits and purposes, **appropriateness** of the measures taken.

3 History and State of the Art

The ISO/IEC 9126-4 (2001) standard identifies usability as one of six different software quality attributes [14]; and the ISO 9241-11 (1998) standard defines usability in terms of efficiency, effectiveness and user satisfaction. User-specific goals can be achieved in a specified context of use; for example, as indicated by directive 90/270/EEC of the Council of the European Union (1990) on minimum safety and health requirements for work with computers. Although there is consensus on software measurables such as system reaction speed and performance times, there is less agreement on the measurability of subjective values such as usability. For example, the criteria *satisfaction* has been defined both subjectively (Shneiderman) and objectively (Nielsen).

The model *Metrics for Usability Standards in Computing (MUSiC)* [15] was concerned with defining measures of software usability, many of which were integrated into the original ISO 9241 standard. According to Bevan (1995), [15], the assumption that quality is an attribute of the product is misleading: The attributes required for quality will depend on how the product is used; i.e. quality of use is defined as the extent to which a product *satisfies stated and implied needs when used under stated conditions.* Consequently, quality of use can be applied to measure usability as the extent to which specific goals can be achieved, concerning effectiveness, efficiency and satisfaction by particular end users carrying out specific tasks in specific environments [15]. Table 2 shows some examples of metrics and the way usability has been defined in a non-consistent manner. This makes it difficult to apply metrics.

Table 2. Usability attributes of various well-known standards or models

Constantine & Lockwood (1999)	ISO 9241-11 (1998)	Shneiderman (1992)	Nielsen (1993)	Preece et. al. (1994)	Shackel (1991)
Efficiency in use	Efficiency	Speed of performance	Efficiency of use	Throughput	Effectiveness (Speed)
Learnability		Time to learn	Learnability (Ease of learning)	Learnability (Ease of learning)	Learnability (Time to learn)
Remember-ability		Retention over time	Memorability		Learnability (Retention)
Reliability		Rote of errors	Errors /Safety	Throughput	Effectiveness (Errors)
User Satis-faction	Satisfaction (Comfort and accept-ability of use.)	Subjective satisfaction	Objective Satisfaction	Attitude	Attitude

4 Usability Metrics

4.1 Metrics

A measure or metric corresponds to a mapping function that assigns a number or symbol to an attribute of an entity. Metrics can be based on some functions that are defined in terms of a formula or just simple countable data. Such countable metrics may be extracted from raw data collected from sources such as log files, video observations, interviews, or surveys. Calculable (refined) metrics are the results of mathematical calculations, algorithms, or heuristics that use raw observational data and countable metrics. A typical example is the *Quality in Use Integrated Map* (QUIM), which is a hierarchical model [16]. The model represents its structure in five different levels which are called factors, criteria, metrics, data and artifacts. The relationship between the elements of these layers is an n-m relationship. Factors in this model are applicable as generic characteristics to all software products and the model is open to context-specific characteristics. The model refines factors into criteria then maps the criteria into usability metrics.

4.2 Application

Since there are two definite categories of technology in question for the application area of information technology for elderly people, it would appear that two different metric tables must be considered.

The first category deals with technology designed to be ubiquitous, non-obtrusive and largely automatic; primarily for the use of care givers and designed to avoid, or at least facilitate, emergency situations. This category can be referred to as *passive technology*. Here, the usability metrics would be similar to standard usage, since it can be assumed that the care givers will have the advantage of being instructed on the correct usage of the software. Since the end users are the care givers, it may also be assumed that, except in very few cases, they will not themselves be suffering from any mental or physical disability.

The second category requires the active cooperation of the elderly as end users. For example, forwarding medical data regularly to a general practitioner's (GP's) database via a mobile phone; contacting chemists for prescription delivery; taking advantage of eCommerce and eGovernment [17] facilities to enable the elderly to retain their independence and to use the internet to keep in contact with other elderly members of the community (chat rooms) and interest groups. This second category, which we can refer to as *active technology*, requires not only the established standards but also those usability criteria which apply to individuals with slight disabilities or special needs. However, this is not sufficient in itself since it fails to take the social environment, needs and opinions of the elderly into account.

The switch from *passive technology* to *active technology* will enable the elderly to retain a larger measure of independence for a longer length of time, thereby reducing the strain on the care giver. It will also require them to learn a new set of habits and abilities. Since it is unreasonable to expect the elderly, possibly ill or disabled individuals to acquire a completely new mental model, the technology must be adjusted to conform to their requirements, preferences and thought patterns [18].

In addition, it will be necessary to overcome any misgivings they may have about their own capabilities, apprehensions about the devices, and possible fears of isolation or worries about cost.

While Sambasivan & Moore [19] correctly state *consistency* and *standards* as an important factor in their metrics and justified this by the statement: "*Consistency and standards – Not having to wonder what different words, situations, or actions mean. Does the system follow conventions?*"; the conventions themselves are not questioned.

Although, consistency is important within a closed system, it is exactly these conventions which we must question when designing for the elderly. Icons, methodologies and metaphors vary between users according to age, culture and previous experience. Familiarity (mentioned in QUIM) can be achieved by semantic conformity, positive nomenclature, acceptance of bias, etc. It is important to ensure that the user is comfortable with the words, situations and actions, since these are parameters that promote acceptability.

Table 3. Relationship between Factors and Criteria (sub-categories)

Relationship between Factors and Criteria (sub-categories)											
	Factors										
Criteria	Efficiency	Effectiveness	Satisfaction	Productivity	Learnability	Safety	Trustfulness	Accessibility	Universality	Usefulness	Acceptability
Likeability			+								+
Controllability							+	+	+	+	+
Simplicity					+			+	+		+
Privacy							+		+	+	+
Security						+	+				+
Familiarity					+		+				+
…											
Safety											+
Discretion											+
Dependability											+
Non-obtrusiveness											+
Appropriateness											+
Understandability											+
Trustworthiness											+

This quantification can be achieved by consolidating these metrics with the QUIM structure [10] in the form of an 11th factor: **acceptability**.

While this factor has many criteria in common with the other factors such as likeability, simplicity, privacy, security, and familiarity, it also introduces a few new criteria specifically for the elderly population: Safety, discretion, dependability, non-obtrusiveness, appropriateness, understandability, and trustworthiness (or the ability to inspire confidence). The result is the extended table 3. Measurements of these factors can be achieved by including the end users in the design. Sociological and philological studies, as well as investigation into the desires and expectations of the end user will provide milestones to integrate applicable usability criteria into the design, while helping to overcome possible distrust.

5 Conclusion and Future Outlook

The rapid advances being made daily mean that we are unable to anticipate exactly what aids will be available and how these will affect lives. In order to prepare the elderly population to live longer and more independent lives with the help of information technology, we must first introduce them not to any particular device but to the concept of modern engineering. They must be willing to judge innovations on their merits rather than rejecting them, out of hand. Awareness and acceptance can be fostered and increased by education and example. Industry must be cognizant of the fact that awareness training must go hand-in-hand with good design and that knowledge of the end users is as important as functionality, since without the end user's cooperation, functionality will be ineffective. In Future extensive work is needed to enhance the quality of technology development for the elderly people.

Acknowledgements

This work has been partly funded by the European Commission under the project no. FP6-IST-2005-045056 Emergency Monitoring and Prevention (EMERGE) [21].

References

1. Kleinberger, T., Becker, M., Ras, E., Holzinger, A., Müller, P.: Ambient Intelligence in Assisted Living: Enable Elderly People to Handle Future Interfaces. In: Universal Access to Ambient Interaction. LNCS, vol. 4555, pp. 103–112. Springer, Heidelberg (2007)
2. Holzinger, A., Searle, G., Nischelwitzer, A.: On some Aspects of Improving Mobile Applications for the Elderly. In: Stephanidis, C. (ed.) Coping with Diversity in Universal Access, Research and Development Methods in Universal Access. LNCS, vol. 4554, pp. 923–932. Springer, Heidelberg (2007)
3. De Marco, T.: Controlling software projects. Management, measurement & estimation. Barry W. Boehm Yourdon Press, Englewood Cliffs (1982)
4. Sauro, J.: The user is in the numbers. Interactions 13(6), 22–25 (2006)
5. Sauro, J.: Quantifying usability. Interactions 13(6), 20–21 (2006)
6. Norman, D.A., Draper, S.: User Centered System Design. Erlbaum, Hillsdale (1986)
7. Karat, J.: User centered design: quality or quackery? Interactions 3(4), 18–20 (1996)
8. Vredenburg, K., Isensee, S., Righi, C.: User Centered Design: an integrated approach. Prentice Hall, Upper Saddle River (2002)
9. Holzinger, A.: Usability Engineering for Software Developers. Communications of the ACM 48(1), 71–74 (2005)
10. Seffah, A., Metzker, E.: The obstacles and myths of usability and software engineering. Communications of the ACM 47(12), 71–76 (2004)
11. Seffah, A., Donyaee, M., Kline, R.B., Padda, H.K.: Usability measurement and metrics: A consolidated model. Software Quality Journal 14(2), 159–178 (2006)
12. Dickinson, A., Goodman, J., Syme, A., Eisma, R., Tiwari, L., Mival, O., Newell, A.: Domesticating technology. In: Stephanidis, C. (ed.) In-home requirements gathering with frail older people. HCI International 2003. LEA, pp. 827–831 (2003)
13. Inglis, E.A., Szymkowiak, A., Gregor, P., Newell, A.F., Hine, N., Wilson, B.A., Evans, J., Shah, P.: Usable technology? Challenges in designing a memory aid with current electronic devices. Neuropsychological Rehabilitation 14(1-2), 77–87 (2004)
14. Bevan, N.: Quality in use: Meeting user needs for quality. Journal of Systems and Software 49(1), 89–96 (1999)
15. Bevan, N.: Measuring Usability as Quality of Use. Software Quality Journal 4(2), 115–130 (1995)
16. Seffah, A., Kececi, N., Donyaee, M.: QUIM: A Framework for Quantifying Usability Metrics. In: Software Quality Models Second Asia-Pacific Conference on Quality Software (APAQS 2001), pp. 311–318 (2001)

17. Phang, C.W., Sutanto, J., Kankanhalli, A., Li, Y., Tan, B.C.Y., Teo, H.H.: Senior citizens' acceptance of information systems: A study in the context of e-government services. IEEE Transactions On Engineering Management 53(4), 555–569 (2006)
18. Hsu, Y.C.: The effects of metaphors on novice and expert learners' performance and mental-model development. Interacting with Computers 18(4), 770–792 (2006)
19. Sambasivan, N., Jackson, M.M.: Designing Pervasive Brain-Computer Interfaces. In: Holzinger, A. (ed.) USAB 2007. LNCS, vol. 4799, pp. 267–272. Springer, Heidelberg (2007)
20. Basili, V.R.: Software Modeling and Measurement: The Goal Question Metric Paradigm. Computer Science Technical Report, CS-TR-2956, University of Maryland, College Park (MD) (September 1992)
21. http://www.emerge-project.eu