Universal Access in the Information Society (2020) 19:515–525
https://doi.org/10.1007/s10209-019-00659-w

LONG PAPER

# Validation of a usability assessment instrument according to the evaluators' perspective about the users' performance

**Ana Isabel Martins**[1] · **Alexandra Queirós**[1,2] · **Nelson Pacheco Rocha**[1,3]

Published online: 1 July 2019


**Abstract**
Technologies for ageing in place may help older adults in their homes to overcome multiple impairments and to promote their autonomy and independence. When conceptualizing technologies for ageing in place, the International Classification of Functioning, Disability and Health (ICF) is a key element due to its functioning and disability framework with consolidated concepts and terminologies. Based on the ICF conceptual framework, the article presents the ICF based Usability Scale (ICF-US) to evaluate the usability of technologies for ageing in place and reports a study aiming: (1) to validate the ICF-US to evaluate usability according to the evaluators' perspective about the users' performance and (2) to evaluate the utility and applicability of the ICF-US. Two observational studies evolving 184 participants were conducted to: (1) assess the validity and reliability of the ICF-US to evaluate usability according to the evaluators' perspective about the users' performance and (2) to verify the utility and applicability of the ICF-US. The results suggest that the ICF-US is valid and reliable and can be used in different stages of technological developments without losing its discriminatory capacity.

**Keywords** Technologies for ageing in place · Usability assessment · International classification of functioning · Disability and health · ICF based Usability Scale · Scale validation

## 1 Introduction

The active ageing vision aims to contribute for a healthy, autonomous and independent life expectation with quality as people get older, including those who are vulnerable, disabled or in need of care [1–3]. The term active refers not only to the ability to be physically active or have an occupation, but also be able to participate in social, economic, cultural, civil or spiritual matters.

✉ Ana Isabel Martins
anaisabelmartins@ua.pt

Alexandra Queirós
alexandra@ua.pt

Nelson Pacheco Rocha
npr@ua.pt

[1] Institute of Electronics and Telematics Engineering of Aveiro, Campus Universitário, 3810 Aveiro, Portugal

[2] Health Sciences School, University of Aveiro, Campus Universitário, 3810 Aveiro, Portugal

[3] Department of Medical Sciences, University of Aveiro, Campus Universitário, 3810 Aveiro, Portugal

In this context, the technological solutions might be used to promote human functioning and to mitigate disabilities, particularly those resulting from the natural ageing process. This perspective is evident in the development of technologies for ageing in place aiming to help older adults in their homes [4, 5].

Technologies for ageing in place, by their nature and complexity, require efficient development methods and a deep knowledge of the end users. One obvious way to get knowledge about the potential end users is to involve them in the development process, in order to consider all the required aspects, such as the personal characteristics, including functioning and incapacity.

Although there are several models to represent and explain how human functioning and incapacity interact, the International Classification of Functioning, Disability and Health (ICF) [6] is assuming an increasing importance since it presents a comprehensive framework for functioning and disability with consolidated concepts and terminologies. Additionally, the ICF conceptual framework allows a multidisciplinary approach centred on the individuals [6].

The purpose of the study reported in the present paper is to validate a usability assessment instrument based on the conceptual framework of ICF, the ICF based Usability Scale (ICF-US), to evaluate the usability of technologies for ageing in place according the evaluators' perspective about the users' performance. In addition to the validation of the referred instrument, this article also reports on the evaluation of its utility and applicability.

In addition to this introductory section, the paper comprises four more sections: Related work, Material and methods, Results and Discussion and conclusion.

## 2 Related work

### 2.1 Technologies for Ageing in Place

Ageing in place is a popular term in current ageing policy, and it is defined as remaining living in the community, with some level of independence, rather than in residential care [7]. This concept is about enabling older adults to maintain independence, autonomy and connection to social support, including friends and family. Having people remaining in their homes and communities for as long as possible also avoids the cost of institutionalization and is therefore favoured by policymakers, health providers and by many older adults themselves [8].

Moreover, ageing well at home means living with quality as long as possible and promotes the use of technological solutions (i.e. technologies for ageing in place [4]) to ensure autonomy (i.e. ability to control, cope with and make personal decisions on a day-to-day basis) and independence (i.e. the ability to perform functions related to daily living with no, any or little help from others) [3–9].

Technologies for ageing in place might contribute, with different roles, to promote care services evolution, from a medical approach to individual-centric operational models, in which the individuals become active partners in their own care processes [10, 11]. Although focalized in specific aspects, different groups of technologies might contribute to this idealized model.

Among these groups of technologies, mobile health (mhealth) (i.e. the use of mobile communication devices, such as smartphones or tablets to support health services, both in terms of disease and well-being management [12, 13]), and Ambient Assisted Living (AAL) (i.e. the application of the Ambient Intelligence concept to enable elderly with specific demands to live longer in their natural environment [14, 15]) have been objects of relevant research.

Furthermore, the advances on sensing technology make it possible the development of mobile and wearable sensors able to monitor continuously physiological parameters, activities and behaviours in out-hospital settings. Therefore, a typical application of technologies for ageing in place consists in monitoring health conditions or the progress of some illness, namely chronic diseases (e.g. patients suffering from diabetes mellitus can be connected to their caregiver by using some mobile technology or ambulatory devices, such as glucometers). Additionally, monitoring physiological parameters together with daily activities (i.e. identifying consistency and completeness in these activities) to assess, in a naturalistic and continuous way, health and cognitive status (e.g. changes in movement patterns, number of outgoings or sleep rhythm [14, 16]) might help to automate assistance and prevents accidents or diseases exacerbation.

As a consequence of relevant technological developments, a significant number of systematic reviews report evidence of the advantages and disadvantages of the use monitoring devices to support patients with chronic diseases (e.g. diabetes [17–23], heart failure [24–26] or chronic obstructive pulmonary disease [27, 28]).

In some instances, technologies for ageing in place might promote the engagement with primary care, replace time-consuming visits and provide rehabilitation care or assistance to other health-related interventions [29, 30]. This might benefit specialities that require frequent follow-up care. For instance, specialized training systems useful to treat stroke patients can be controlled by remote physiotherapists with access to the results, namely in terms of exercise levels.

In the future, AAL and smart health applications may increase the living span of disabled people and elderly, which is a major challenge itself that will, necessarily, change the way primary care institutions operate [31].

Furthermore, it should be understood how technologies for ageing in place might facilitate the individuals to be actively involved in their health and care pathway. Concerning patients with chronic diseases, applications might allow them to receive information to better control their diseases. For instance, educational information about pain (e.g. general information, symptoms or causes) can be provided, as well as information related to individual health conditions or pain relief (e.g. relaxation techniques or pain reduction techniques, such as acupressure), through a variety of media, including images, video or animations [32]. Additionally, applications can be used for cognitive rehabilitation and to support older adults suffering from cognitive decline [33].

Other applications promote self-management, namely lifestyle management, prescriptions reminders, care appointments management or health care record access (e.g. putting patients with the ability to securely share their health information with clinicians or others, as needed), or help the patients to contribute with observations of their daily living (e.g. personal health records) [34].

Technologies for ageing in place can also support other types of applications, namely preventive applications, applications to enhance the communication between care providers and patients and between caregivers or applications to support frail citizens [23, 35–37].

Most of the applications related to technologies for ageing in place intend to use explicit and implicit interaction forms, and, consequently, the interaction mechanisms are very demanding. Moreover, their acceptance by the end users should be assessed carefully. Therefore, usability is a critical issue, which highlights the need of comprehensive usability assessment instruments [38].

In this field, it should be noted that older users' acceptance of software solutions is influenced by several factors including perceived ease of use, perceived usefulness and previous exposure to technology [39, 40].

Different types of instruments allow the collection of quantitative and qualitative data related to characteristics, thoughts, feelings, perceptions, behaviours or attitudes of the end users. Examples include general purpose usability instruments such as the Post-Study System Usability Questionnaire (PSSUQ) [41] and the System Usability Scale (SUS) [42].

The PSSUQ is a usability evaluation questionnaire developed by International Business Machines (IBM). It is composed of 19 items developed to assess the user satisfaction with the system usability. The PSSUQ is widely used and robust in psychometric terms [43, 44].

The System Usability Scale (SUS) was developed by John Brooke more than 25 years ago as part of a usability engineering programme (1986) as a 'quick and dirty' survey scale that would allow the usability practitioner to quickly and easily assess the usability of a given product or service. SUS is composed of 10 statements that are scored on a 5 point likert scale of strength of agreement. Its final score can range from 0 to 100, where higher scores indicate better usability [45, 46].

The principal value of this usability instruments is that they provide a single reference score for participants' view of a product's usability.

### 2.2 International classification of functioning, disability and health

The decline of human functioning is a common problem, frequently associated with changes related to age, health status or social factors [47]. When the decline is intense, there is a decrease in the functional reserve, and the person becomes more vulnerable to the onset of chronic diseases [48] which impacts their autonomy and independence, and, thus, their quality of life. With the age increase, a set of sensory modifications tend to emerge, such as loss of hearing and visual acuity. These types of changes have a real impact on the individuals' life and are reflected in the level of human functioning [49, 50].

A formal terminology able to characterize the human functioning and incapacity is a major challenge, because there is still a conceptual ambiguity, demonstrated, for example, by the multiplicity of existing concepts related to functioning [51–53].

The ICF [6], adopted at the 54th World Health Assembly in 2001, attempts to provide a coherent view of biopsychosocial health and health-related states [54]. The title itself reflects the priority given to functioning as a health component over the consequences of disease. Additionally, the nomenclature being used focuses on a positive approach to body structures and functions, activities and participation [6, 55].

The ICF does not classify people, but rather interprets their characteristics and the influences of the environment (i.e. contextual factors including personal or environmental factors), which allows to properly describe functional states. Hence, the ICF model does not emphasize the individual as a disabled person, even if temporary, but the components that favour or hinder the execution of activities or participation. The functioning and the disability are considered as a result of a dynamic interaction between health conditions and the contextual factors, whether personal or environmental factors [54, 56].

According to the model underlying the ICF, different environments may have a different impact on people with the same health condition. An environment with or without barriers (i.e. environmental factors that, by its absence or presence, restrict the human functioning and cause disability) can limit or facilitate individuals' performance and participation [54]. Thus, the environment in which people live and conduct their lives has a direct influence on their functioning, and in this sense, it can be enhanced by facilitators such as technologies for ageing in place that meet the needs and characteristics of end users.

Facilitators are environmental factors that, by their absence or presence, improve human functioning and reduce the disability of the persons by increasing their performance and participation. These factors include aspects such as an accessible physical environment, availability of appropriate assistive technology, positive attitude towards disability, as well as services, systems and policies aimed at improving the involvement of all people, irrespective of their health conditions in all areas of life.

Considering the use of technologies for ageing in place aims to improve individual performance in carrying out activities and to facilitate participation, and then the technologies for ageing in place are, in the context of the ICF perspective, environmental factors. Thus, when developing and evaluating technologies for ageing in place, it is essential to consider, primarily, human functioning and environmental factors that influence functioning and, secondly, standardized concepts that provide a common language

**Table 1** ICF-US I items [59]

| How would you rate the application regarding |
|---|
| Ease of use |
| The satisfaction with its use |
| The learning easiness |
| The achievement of expected results |
| The similarity in the operation mode in the different tasks |
| The possibility to interact in various ways |
| The understanding of the messages presented |
| The application responses to your actions |
| The knowledge of what was happening in the application during the utilization |
| Overall, I consider that the application was… |

**Table 2** ICF-US answer key [59]

| Barrier | | | Facilitator | | |
|---|---|---|---|---|---|
| Big | Medium | Small | Small | Medium | Big |
| −3 | −2 | −1 | 1 | 2 | 3 |

to all stakeholders [57]. Consequently, ICF might have an important role in supporting the development and evaluation of technologies for ageing in place, namely in what concerns usability assessments [58].

The authors have proposed the ICF based Usability Scale (ICF-US), a usability assessment instrument with two subscales, the ICF-US I and ICF-US II [59]. The ICF-US I as a self-perceived usability rating scale (i.e. filled in according to the users' opinion) proved to be reliable and has construct validity when correlated with other self-perceived usability rating scales [60], namely the PSSUQ [41] and the SUS [42]. Additionally, the ICF-US II is a questionnaire that allows a more detailed usability evaluation, by classifying the components of a prototype as barriers or facilitators to identify their strengths (facilitators) and weaknesses (barriers).

# 3 Materials and methods

## 3.1 ICF based Usability Scale (ICF-US)

The ICF-US I is composed by a set of items associated with different usability principles (Table 1), and its answer key (Table 2) is based on the ICF environmental factors qualifiers (facilitators and barriers), being all items scored from −3 to 3. The value 3 is the highest score (facilitator), and −3 is the lowest (barrier).

If a participant does not answer to an item or classify it as not applicable (NA), then this item score should be filled up with the average value of the remaining items, rounded to the units. The final score of the ICF-US I is calculated by adding the scores of all items. A value above 10 means good usability. On the other hand, a value below 10 indicates poor usability and, in that case, the ICF-US II should be applied to identify what should be improved.

If there is an opportunity for improvement, it is important to define the priorities of subsequent developments, considering the aspects acting as barriers and facilitators.

The ICF-US II has the same answer key as the ICF-US I (Table 2), and its objective is to gather information to understand what must be modified (barriers) and what should be established as good practice (facilitators). It consists of a set of items that identify different components (e.g. sound, image or touch) of the prototype being evaluated.

The ICF-US II items depend on the technological solutions being evaluated, so its instantiation must be based on the particular solution components, in accordance with specific guidelines created to ensure that the instrument can be used by anyone else than the authors themselves [61].

Each component is then classified as a barrier or facilitator. When a component is classified as a barrier, the evaluator must identify the feature that is causing that classification. For example, the reason why a specific interface can be considered as a barrier could be the size, sharpness or contrast of the icons, or any other reason (Fig. 1). This procedure should be performed for all identified barriers, which allows listing the components that need to be improved and the reasons why [62].

## 3.2 Study design

The ICF-US was developed to assess self-perceived usability [60]. However, it is relevant to assess usability according to the evaluators' perspective about the users' performance, in order to make a more objective evaluation. Therefore, the study reported in the present paper aimed at the following objectives:

- To validate the adequacy of the ICF-US I to evaluate the usability of technologies for ageing in place reflecting the opinions of the evaluators based on the observation of the users' performance, instead of self-perceived usability;
- To validate the ICF-US II, based on the observation of the user's performance, as a complement of the ICF-US I, particularly by identifying potential barriers or facilitators of technologies for ageing in place;
- To evaluate the utility and applicability of the ICF-US according the evaluators' perspective about the users' performance.

The authors designed two observational studies to achieve these objectives:

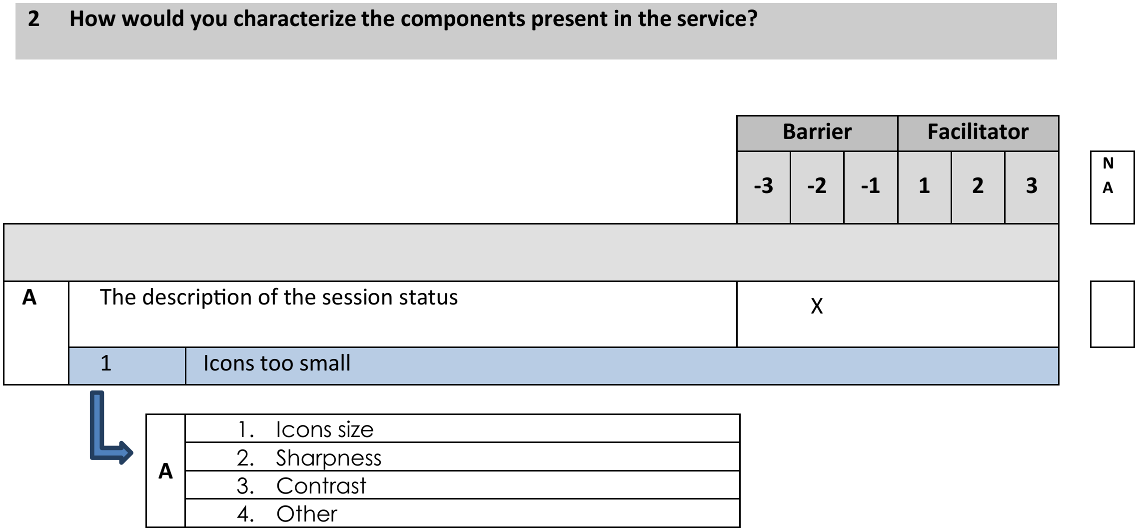


**Fig. 1** Extract from the ICF-US II [59]

- Validity and reliability: the first observational study was designed to evaluate the validity and reliability of the ICF-US I when evaluating usability according the evaluators' perspective about the users' performance, as well as to verify if the results of ICF-US II to complement ICF-US I are correlated with the information gathered through usability evaluation techniques based on the users' performance;
- Utility and applicability: the second observational study intended to evaluate the utility and applicability of the ICF-US instrument.

Although the data collected in the two observational studies are not sensitive, the principles underlying the Helsinki Declaration were considered [63], and the necessary authorizations were requested, all collected data were anonymized, and all participants signed an informed consent prior to data collection.

### 3.3 Validity and reliability

The ICF-US instrument validation was held in a social solidarity institution during December 2014 using an interactive television application to support home telecare, the aal@meo [62]. At that time, the aal@meo was a prototype at an intermediate state of development (i.e. requiring further tests and improvements) aiming to be integrated in a commercial service of an Internet Protocol television operator.

The authors used a convenience sample including users above 18 years old, able to read, understand and sign the informed consent. The only exclusion criterion was the presence of limitation in the fine movements of the arms that prevented the use of conventional software applications.

Data collection included the filling in of two usability assessment tools (ICF-US and PSSUQ), user performance registration and the record of critical incidents. The PSSUQ is a usability evaluation questionnaire [43], which was translated and adapted from the cultural and linguistic point of view to European Portuguese [64]. It consists of 19 items aimed at addressing five usability characteristics of a system: rapid completion of the task, ease of learning, high-quality documentation, online information and functional adequacy [43]. The performance evaluation consisted of the registration of the success or failure rate in carrying out tasks, the execution time in seconds and the total number of errors. The critical incident registration aimed at the systematic identification of any event outside normality, by the observer, considering details such as, for example, difficulty of interaction with the application.

An evaluator and an observer were involved in the evaluation sessions that had the following structure:

- Introduction—the evaluator applied a sociodemographic questionnaire and then delivered a session script, explaining orally all the information in it;
- Test—the participant performed the tasks included in the session script. At the same time, the observer recorded the performance of the participant in a performance evaluation grid and noted any critical incident that happened;
- Interview—the evaluator conducted a final interview, applied the PSSUQ and thanked the participation;
- Completion—the evaluator filled in the ICF-US according to his opinion about the participant performance.

When filling in the ICF-US I and II, it was up to the evaluator to decide what qualifier should be allocated to each item.

This decision relied on the observation of the participant interaction with the aal@meo and in the interview to clarify the issues that had caused doubts. For example, if during the test a participant showed feelings of frustration, disgust and confusion, then the evaluator had sufficient evidence to qualify the item related to the satisfaction without having to ask the participant. Whenever there was a lack of information, the evaluator used the interview to gather information to support the assignment of the qualifier.

Construct validity was assessed through the correlation between the ICF-US I and the PSSUQ.

Moreover, the results of the application of ICF-US II were correlated with the performance records, particularly the critical incident data. Two independent researchers analysed the correlation between the results of the ICF-US II and the critical incident data. The analysis was performed using the Statistical Package for Social Sciences (SPSS) to calculate a Spearman correlation coefficient, and the level of significance established was $p < 0.05$.

### 3.4 Utility and applicability

Concerning the observational study to verify the utility and applicability of the ICF-US, another technology for ageing in place was used: the Brain on Track, a computerized web-based self-administered test intended for longitudinal cognitive testing [65].

Brain on Track was in a state of advanced development, properly consolidated and ready to be disseminated to the general population. The Brain on Track resulted from an iterative process of development and evaluation in a hospital environment and its clinical validation for the Portuguese population was carried out in a research cohort [65].

The observational study was held on the community as a real context evaluation, from February 2015 to July 2017.

The sample selection was performed using as inclusion criteria age above 50 years. All adults able to fill in the assessment instruments were eligible to participate if they gave written informed consent and were 50 years or older (consecutive recruitment). The written informed consent was obtained prior to data collection.

Since the experiment was planned to provide a real context evaluation, the assessment was performed without a task script and without any interference from the evaluator in the course of the participants' interaction with Brain on Track. Therefore, the unpredictability was considerable superior to the unpredictability of an evaluation performed in a laboratory.

Data collection procedure was performed using the following structure:

- Introduction—the evaluator applied a sociodemographic questionnaire;
- Test—the participant interacted freely with the Brain on Track by performing a cognitive monitoring test;
- Completion—the evaluator filled the ICF-US based on his own opinion about the participant performance.

Regarding data analysis, to describe and characterize the subjects that constitute the population sample, central tendency and dispersion measures were used, including mean, range and standard deviation (SD). Statistical analyses were performed with Microsoft Excel.

## 4 Results

### 4.1 Validity and reliability

Thirty participants were selected for the assessment of the validity and reliability of ICF-US when used to evaluate usability according the evaluators' perspective about the users' performance. The participants had an average age of 58 years (SD = 3.3), maximum of 67 and a minimum of 54 years.

The correlation of the ICF-US I with PSSUQ (− 0.75) was significant, as well as the correlation between the ICF-US I and the performance records, namely with the success/failure rate (0.74) and the number of errors (− 0.72) (Table 3).

The ICF-US I had a strong negative correlation with PSSUQ because a higher score in ICF-US I corresponds to a lower score in the PSSUQ, where lower values indicate better usability. Similarly, the correlation between the ICF-US I and the success/failure rate was also negative, which makes sense because a higher score in the ICF-US I corresponds to a minor number of errors.

To verify the agreement between the results in the ICF-US II and the critical incident data, the authors created a table to classify each item of the ICF-US II (from − 3 to 3) and, in the case of a barrier, the indication of the reason why the item was classified as a barrier. Then, the critical incident report was fully transcript and matched with the item or items of the ICF-US II that addressed the same issue.

The classification of each item was based on the following criteria:

**Table 3** Correlation between the ICF-US I, PSSUQ and the performance records [59]

| | Correlation with the ICF-US I |
|---|---|
| PSSUQ | − 0.75 ($p < 0.05$) |
| Performance (success/failure) | 0.74 ($p < 0.05$) |
| Performance (number of errors) | − 0.72 ($p < 0.05$) |

- Concordance barrier—whenever the evaluator classified the item as a barrier and the observer registered a related critical incident;
- No concordance barrier—whenever the evaluator classified the item as a barrier, but the observer did not register any related critical incident;
- Facilitator without corresponding critical incident—whenever the evaluator classified the item as a facilitator, but the observer did not record any related critical incident;
- Concordance facilitator—whenever the evaluator classified the item as a facilitator and the observer registered critical incidents that are related;
- Frail concordance facilitator—whenever the evaluator classified the item as a facilitator, but the related critical incidents do not clarify on the facility/difficulty in performing the task. For example, sometimes the evaluator assigned a score of 1 (small facilitator) and the critical incident recorded by the observer was something like 'he needed minor help to start the task but then managed to do it without any problem';
- No concordance facilitator—whenever the evaluator classified the item as a facilitator, but the observer registered some critical incidents.

The analysis was performed individually by two researchers, which confronted the doubtful items, thus reaching to a result about the agreement between the ICF-US II classification and the related critical incident data.

This result is significant (Table 4), with a concordance rate of 95.8% (i.e. whenever there was a critical incident regarding a component, in general, the ICF-US II reported a barrier for that component). Whenever the evaluator described the item as a facilitator, in 80.2% of the cases, there was no critical incident related, which is expected because this technique only previses the registration when something outside normal happens.

The percentage of facilitators that had a weak agreement was very low, 5.6%, representing only 14 cases. A more detailed analysis showed that the cases marked as weak agreement were mostly rated as facilitator level 1 (12 cases), indicating that the correlation between the evaluator and the observer is weaker when it comes to aspects classified as small facilitators. Therefore, ICF-US II has enough sensitivity to discriminate environmental factors.

These results also show that there is a correlation between objective usability measures, namely users' performance (success/failure rate and number of errors) and the ICF-US I when the filling is based on the perspective of the evaluators about the users' performance. There is also a very similar match to critical incident data.

The execution time of each task was also collected. However, execution times were not considered for the analysis because they were influenced by other circumstances (e.g. the delay of the communication network).

## 4.2 Utility and applicability

For the verification of the utility and applicability of the ICF-US, the number of participants selected was 154. The participants had an average age of 58 years (SD = 7.21), maximum of 83 and a minimum of 50 years (Table 5).

Although focusing in technologies for ageing in place, testing and gathering suggestions during the development of technological solutions with adults (that are not elderly yet) may represent an added value considering that they will be the users of these solutions in a short–medium term future.

Considering as an inclusion criteria age over 50 allows to capture the opinion of future elderly that may be different from the current elderly.

Considering the results of ICF-US I, the application was a facilitator for 118 participants and a barrier for 36 participants. The mean score of all participants was 15.02, in a range from −30.00 to 30.00 (SD = 13.15), which indicates that in general the application was considered a medium facilitator. The participant with the highest score had 30.00, and the participant with the lowest score had −30.00.

**Table 5** Sample characterization

| | | |
|---|---|---|
| Participants | 154 | |
| Gender *n* (%) | | |
| Female | 106 | (68.8) |
| Male | 48 | (31.2) |
| Age | | |
| Mean (DP) | 58 | (7.21) |
| Min–Max | 50–83 | |

**Table 4** Analysis of agreement results [59]

| Barrier | | Facilitator | | | |
|---|---|---|---|---|---|
| Concordance | 95.8% | Concordance | 99.2% | Facilitator without corresponding critical incidents | 80.2% |
| | | | | Concordance facilitator | 13.5% |
| | | | | Frail concordance facilitator | 5.6% |
| No concordance | 4.2% | No concordance | 0.8% | | |

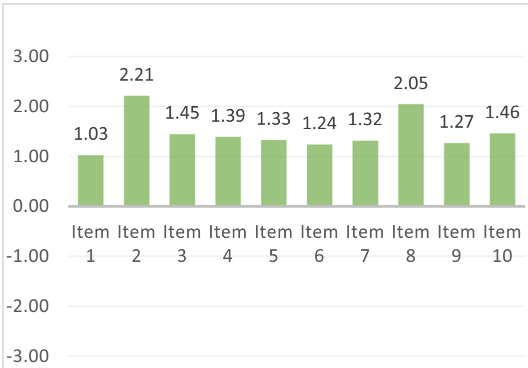


**Fig. 2** ICF-US I item scores

**Table 6** ICF-US I score analysis according with age

| Age Group | Average score ICF-US I | Number of participants |
|---|---|---|
| 50–55 | 16.52 | 71 |
| 56–60 | 15.11 | 38 |
| 61–65 | 13.35 | 23 |
| 67–70 | 12.11 | 9 |
| 71–75 | 11.5 | 10 |
| 76+ | 9.33 | 3 |

Observing the items individually, it is important to classify each item as a barrier or a facilitator (Fig. 2).

All items were considered as facilitators. The items that were the biggest facilitators were item 2 'The satisfaction with its use' (2.22) and item 8 'The application responses to your actions' (2.05). Considering that all items were identified as facilitators, it appears that most of the participants easily understood the information included in the application, were able to perform tasks related to the interaction, learning, execution and repetition of actions and, moreover, were satisfied.

A detailed analysis according to age was performed in order to understand how age influences the usability spectrum. There was a downward trend in the ICF-US I score as the participants' age increases. While participants in the age group 50–55 had an average score of 16.52, the older age groups had an average of 11.55 (age group 71–75) and 9.33 (age group 76+) (Table 6).

In order to allow the identification of potential barriers and facilitators, the ICF-US II must be instantiated to each specific application. For the instantiation of the ICF-US II, the components presented on Table 7 were considered.

The results of the three parts associated with the instantiation of ICF-US II were analysed according to the mean of responses (in a range from − 3.00 to 3.00):

- The mean value for the evaluation of application components was − 0.30 (SD = 1.46);

**Table 7** Instantiation of ICF-US II

| |
|---|
| Part 1—application components |
| How do you characterize the components present in the application? |
| The login to the system was… |
| The instructions were… |
| The timer was… |
| The action buttons were (start, pause, expand)… |
| Part 2—detailed usability |
| How do you characterize the text? |
| The size was… |
| The font was… |
| The color of the text was… |
| How do you characterize the image? |
| The size was… |
| The colors were… |
| The contrast was… |
| How do you characterize the audio feature? |
| The instructions were… |
| The sound feedback of the answers (right vs. wrong) was? |
| How do you characterize the understanding of the application's operation mode? |
| The application operation mode was… |
| Part 3—overall usability |
| In general, how would you characterize the progress of the session? |

- The mean value for the evaluation of detailed usability was 0.27 (SD = 1.11);
- The mean value for the overall usability of the application was − 1.27 (SD = 2.05).

Observing the items of ICF-US II individually, it is important to classify each item as a barrier or a facilitator. The barriers, in order of severity, were: (1) 'The login to the system' (− 2.14); (2) 'The progress of the session' (− 1.68); (3) 'The application operation mode' (− 1.63); (4) 'The instructions' (− 1.10); (5) 'The action buttons' (− 0.04).

In turn, the facilitators were: (1) 'The text size' (1.67); (2) 'The image contrast' (1.66); (3) 'The colours of the text' (2:36); (4) 'The sound feedback of the answers' (1.53); (5) The 'font' (1.50); (6) 'The image size' (1.44), (7) 'The image colours' (1.32), (8) 'The audio of the instructions' (0.23); (9) 'The timer' (0.15).

Considering the items identified as a barrier, it appears that most participants had difficulty with tasks related to logging in the system, the application operation mode, the clarity of the instructions and action buttons.

## 5 Discussion and conclusion

The results of the first observational study (validity and reliability) suggest appropriate values for the validity and reliability of the ICF-US I when used to evaluate usability according to the evaluators' perspective about the users' performance [59], which corroborates the results obtained by the same scale to assess self-perceived usability [60]. In addition, the ICF-US II also presents a high agreement rate and correlation with the critical incident data.

Furthermore, the results of the first observational study also point that when the filling of the ICF-US I is based on the evaluators' perspective about the users' performance, there is a strong correlation with objective usability evaluation measures, including the success/failure rates and the number of errors in each task.

Thus, the ICF-US I, in addition to being a valid and reliable measure for assessing self-perceived usability of technologies for ageing in place, when filled to reflect the opinion of the evaluator, is as accurate as other objective evaluation measures with higher complexity.

This overcomes a difficulty reported in the literature related to the fact that the opinion of the users, collected through the filling of generic usability scales, does not fully reflect users' performance. In fact, the work developed by Bangor and colleagues [66] examined 200 studies that evaluated the usability using the SUS and found that the results do not follow the full performance spectrum.

Regarding specifically the ICF-US II, the fact it presented an agreement rate above 95% with critical incident data appears to be an indicator that, in case of limited resources to conduct the evaluation, particularly in terms of human resources, it can be done with only one evaluator. Ideally, there should be an evaluator and an observer, even for the sake of information redundancy, though if this is not affordable, using the ICF-US makes possible to perform the same evaluation, almost without loss of information [59].

The critical incident technique involves recording all the situations that deviate from normality. There is no structure or standardized procedure for the registration of critical incidents. On the other hand, the ICF-US II is built for each application in order to address the different components, the user interaction and the usability issues. Consequently, when the ICF-US II is filled in shortly after the test session, the evaluator is able to fill it easily without incurring losses of information. In addition, when the data is processed, the information is already structured and organized, in opposition to what happens with the critical incident technique, whose data analysis is typically done by creating classes and counting the number of occurrences of each critical incident [59]. For this reason, it is important that the components of ICF-US II are well defined, preferably by a professional with accurate knowledge about the technological solution to be evaluated and an expert in usability evaluation.

The second observational study (utility and applicability) shows that even for a technological solution that has been evaluated several times before and that presented good usability, the Brain on Track, the application of ICF-US I and ICF-US II was able to identify barriers, such as the login method.

Therefore, another interesting result is that the ICF-US can be used in different stages of development without losing its discriminatory capacity. Even when evaluating the usability of a valid and consolidated system, with usability tests performed, and in its final stage of development, which was the case of Brain on Track, the ICF was able to discern between age groups, proving its usefulness and applicability. The discriminatory capacity of ICF-US was also evaluated with different groups of participants, and the results show that older people have lower usability scores, which is a predictable result since for the more advanced ages people tend to have lower technology literacy and more difficulty when interacting with information technologies.

The application of ICF-US in two observational studies assessing different technologies at different stages of development shows the plasticity of this usability assessment instrument and its ability to adapt to the various technological development phases. This indicates that both components of ICF-US (i.e. ICF-US I and ICF-US II) can be useful tools for usability evaluation of technologies for ageing in place according to the evaluators' perspective about the users' performance.

Comparing with other usability evaluation instruments, the main advantage of using the ICF-US is related to the fact that this instrument is based on a conceptual model supported by concepts and terminologies established by the World Health Organization, which means that they are universally accepted. This may facilitate the consolidation of knowledge, which is essential to promote strategic planning, technological innovation and the involvement of different stakeholders in the design and development cycles of ageing in place technologies.

**Acknowledgements** This study was funded by the Project QREN No. 13852 I&D «AAL4ALL—Ambient Assisted Living for All», UI 127/94—IEETA co-funded by COMPETE and FEDER.

## References

1. World Health Organization, Growing older-Staying Well. Ageing and physical activity in everyday life, Geneve (1998)
2. World Health Organization, Health Promotion Glossary, Genebra (1998)
3. World Health Organization, Active ageing: a policy framework. A contribution of the World Health Organization to the second United Nations World Assembly on Aging, Genebra (2002)
4. Connelly, K., ur Rehman Laghari, K., Mokhtari, M., Falk, T.H.: Approaches to understanding the impact of technologies for aging in place: a mini-review. Gerontology **60**(3), 282–288 (2014)
5. Pastalan, L.A.: Aging in place: the role of housing and social supports. Haworth Press, Philadelphia (1990)
6. World Health Organization. International classification of functioning, disability and health: ICF. Geneva: World Health Organization (2001)
7. Davey, J.A.: New Zealand Institute for Research on Ageing, Accommodation options for older people in Aotearoa/New Zealand. Centre for Housing Research (2004)
8. World Health Organization, Global Age-friendly cities: a guide, Geneva (2007)
9. Walker, P.R.: Report from the world NGO forum on ageing and the united nations second world assembly on ageing (2002)
10. Korhonen, I., Bardram, J.E.: Guest editorial: introduction to the special section on pervasive healthcare. IEEE Trans. Inf. Technol. Biomed. **8**(3), 229–234 (2004)
11. Korhonen, I., Parkka, J., Van Gils, M.: Health monitoring in the home of the future. IEEE Eng. Med. Biol. Mag. **22**(3), 66–73 (2003)
12. Boulos, M.N.K., Brewer, A.C., Karimkhani, C., Buller, D.B., Dellavalle, R.P.: Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J. Public Health Inform. **5**(3), 229 (2014)
13. Mosa, A.S.M., Yoo, I., Sheets, L.: A systematic review of healthcare applications for smartphones. BMC Med. Inform. Decis. Mak. **12**(1), 67 (2012)
14. Rashidi, P., Mihailidis, A.: A survey on ambient-assisted living tools for older adults. IEEE J. Biomed. Heal. Inform. **17**(3), 579–590 (2013)
15. Queirós, A., Silva, A., Alvarelhão, J., Rocha, N.P., Teixeira, A.: Usability, accessibility and ambient-assisted living: a systematic literature review. Int. J. UAIS **14**(1), 57–66 (2015)
16. Suzuki, T., Murase, S., Tanaka, T., Okazawa, T.: New approach for the early detection of dementia by recording in-house activities. Telemed. e-Health **13**(1), 41–44 (2007)
17. Or, C.K.L., Tao, D.: Does the use of consumer health information technology improve outcomes in the patient self-management of diabetes? A meta-analysis and narrative review of randomized controlled trials. Int. J. Med. Inform. **83**(5), 320–329 (2014)
18. Farmer, A.J., McSharry, J., Rowbotham, S., McGowan, L., Ricci-Cabello, I., French, D.P.: Effects of interventions promoting monitoring of medication use and brief messaging on medication adherence for people with Type 2 diabetes: a systematic review of randomized trials. Diabet. Med. **33**(5), 565–579 (2016)
19. Huang, Z., Tao, H., Meng, Q., Jing, L.: Management of endocrine disease: effects of telecare intervention on glycemic control in type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. Eur. J. Endocrinol. **172**(3), R93–R101 (2015)
20. Hunt, C.W.: Technology and diabetes self-management: an integrative review. World J. Diabetes **6**(2), 225 (2015)
21. Riazi, H., Larijani, B., Langarizadeh, M., Shahmoradi, L.: Managing diabetes mellitus using information technology: a systematic review. J. Diabetes Metab. Disord. **14**, 49 (2015)
22. Tildesley, H.D., Po, M.D., Ross, S.A.: Internet blood glucose monitoring systems provide lasting glycemic benefit in Type 1 and 2 diabetes. Med. Clin. North Am. **99**(1), 17–33 (2015)
23. Arambepola, C., Ricci-Cabello, I., Manikavasagam, P., Roberts, N., French, D.P., Farmer, A.: The Impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with Type 2 diabetes: a systematic literature review and meta-analysis of controlled trials. J. Med. Internet Res. **18**(4), e86 (2016)
24. Nakamura, N., Koga, T., Iseki, H.: A meta-analysis of remote patient monitoring for chronic heart failure patients. J. Telemed. Telecare **20**(1), 11–17 (2014)
25. Conway, A., Inglis, S.C., Clark, R.A.: Effective technologies for noninvasive remote monitoring in heart failure. Telemed. e-Health **20**(6), 531–538 (2014)
26. Grustam, A.S., Severens, J.L., van Nijnatten, J., Koymans, R., Vrijhoef, H.J.M.: Cost-effectiveness of telehealth interventions for chronic heart failure patients: a literature review. Int. J. Technol. Assess. Health Care **30**(01), 59–68 (2014)
27. Pedone, C., Lelli, D.: Systematic review of telemonitoring in COPD: an update. Pneumonol. Alergol. Pol. **83**(6), 476–484 (2015)
28. Lundell, S., Holmner, Å., Rehn, B., Nyberg, A., Wadell, K.: Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respir. Med. **109**(1), 11–26 (2015)
29. Alcañiz, M., Botella, C., Baños, R.M., Zaragoza, I., Guixeres, J.: The Intelligent e-Therapy system: a new paradigm for telepsychology and cybertherapy. Br. J. Guid. Counc. **37**(3), 287–296 (2009)
30. Singh, D., Kropf, J., Hanke, S., Holzinger, A.: Ambient Assisted Living Technologies from the Perspectives of Older People and Professionals, pp. 255–266. Springer, Cham (2017)
31. Holzinger, A., Röcker, C., Ziefle, M.: From smart health to smart hospitals. In: Smart Health: State-of-the-Art and Beyond, LNCS, Lecture Notes in Computer Science, pp. 1–20 (2015)
32. Rosser, B.A., Eccleston, C.: Smartphone applications for pain management. J. Telemed. Telecare **17**(6), 308–312 (2011)
33. Cruz, V.T., et al.: A rehabilitation tool designed for intensive web-based cognitive training: description and usability study. JMIR Res. Protoc. **2**(2), e59 (2013)
34. Roehrs, A., da Costa, C.A., da Righi, R., de Oliveira, K.S.F.: Personal health records: a systematic literature review. J. Med. Internet Res. **19**(1), e13 (2017)
35. Chen, Y.-R.R., Schulz, P.J.: The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J. Med. Internet Res. **18**(1), e18 (2016)

36. Matthews, J., Win, K.T., Oinas-Kukkonen, H., Freeman, M.: Persuasive technology in mobile applications promoting physical activity: a systematic review. J. Med. Syst. **40**(3), 72 (2016)
37. Shingleton, R.M., Palfai, T.P.: Technology-delivered adaptations of motivational interviewing for health-related behaviors: a systematic review of the current research. Patient Educ. Couns. **99**(1), 17–35 (2016)
38. Hoppe, A.: Technological stress: mental strain of younger and older users if technology fails. Springer, Berlin (2011)
39. Holzinger, A., Searle, G., Wernbacher, M.: The effect of previous exposure to technology on acceptance and its importance in usability and accessibility engineering. Univers. Access Inf. Soc. **10**(3), 245–260 (2011)
40. Holzinger, A., Schaupp, K., Eder-Halbedl, W.: An investigation on acceptance of ubiquitous devices for the elderly in a geriatric hospital environment: using the example of person tracking. In: Velasco, C., Weber, G., Barroso, J., Mohamad, Y., Paredes, H. (eds.) Computers helping people with special needs. Springer, Berlin, pp. 22–29 (2008)
41. Lewis, J.R.: Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int. J. Hum. Comput. Interact. **14**(3&4), 463–488 (2002)
42. Brooke, J.: SUS-A quick and dirty usability scale. In: Jordan, P., Thomas, B., Weerdmeester, B., McClelland, I. (eds.) Usability evaluation in industry, pp. 189–194. Taylor and Francis, London (1996)
43. Lewis, J.R.: Psychometric evaluation of the post-study system usability questionaire: the PSSUQ. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 1259–1263 (1992)
44. Rosa, A.F., Martins, A.I., Costa, V., Queirós, A., Silva, A., Rocha, N.P.: European Portuguese validation of the post-study system usability questionnaire (PSSUQ). In: 2015 10th Iberian Conference on Information Systems and Technologies, CISTI 2015 (2015)
45. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the System Usability Scale. Int. J. Hum. Comput. Interact. **24**(6), 574–594 (2008)
46. Martins, A.I., Rosa, A.F., Queirós, A., Silva, A., Rocha, N.P.: European Portuguese validation of the system usability scale (SUS). In: Procedia Computer Science, vol. 67 (2015)
47. Fernández, L.K.: Valoración geriátrica integral. El Resid. **5**(2), 55–65 (2010)
48. Netto, M.P., Ponte, J.: Envelhecimento: desafio na transição do século. In: Netto, M.P. (ed.) Gerontologia—a Velhice e o Envelhecimento em Visão Globalizada, 1st edn, pp. 3–12. Atheneu, São Paulo (2000)
49. Keller, B.K., Morton, J.L., Thomas, V.S., Potter, J.F.: The effect of visual and hearing impairments on functional status. J. Am. Geriatr. Soc. **47**(11), 1319–1325 (1999)
50. Debevc, M., Kosec, P., Holzinger, A.: Improving multimodal web accessibility for deaf people: sign language interpreter module. Multimed. Tools Appl. **54**(1), 181–199 (2011)
51. Di Nubila, H.B.V., Buchalla, C.M.: O papel das Classificações da OMS—CID e CIF nas definições de deficiência e incapacidade. Rev. Bras. Epidemiol. **11**(2), 324–335 (2008)
52. Stucki, G., Reinhardt, J.D., Grimby, G., Melvin, J.: Developing 'human functioning and rehabilitation research' from the comprehensive perspective. J. Rehabil. Med. **39**(9), 665–671 (2007)
53. Schneider, R., Marcolin, D., Dalacorte, R.: Avaliação funcional de idosos. Sci. Med. (Porto. Alegre) **18**(1), 4–9 (2008)
54. Jette, A.M.: Toward a common language for function, disability, and health. Phys. Ther. **86**(5), 726–734 (2006)
55. Nordenfelt, L.: Action theory, disability and ICF. Disabil. Rehabil. **25**(18), 1075–1079 (2003)
56. Federici, S., Meloni, F., LoPresti, A.: International literature review on WHODAS II. Life Span Disabil. **12**(1), 83–110 (2009)
57. Queirós, A., Alvarelhão, J., Amaro, A., Teixeira, A., Rocha, N.: The international classification of functioning, disability of health as a conceptual framework for the design, development and evaluation of AAL services for older adults. In: Workshop AAL. Rome, Italy, pp. 46–59 (2011)
58. Queirós, A., Silva, A., Alvarelhão, J., Rocha, N.: Um modelo conceptual para o ambient assisted living. In: Teixeira, A., Queirós, A., Rocha, N. (eds.) Laboratório Vivo de Usabilidade, 1st edn, pp. 89–99. Arc Pubishing, Aveiro (2013)
59. Martins, A.I., Queirós, A., Silva, A.G., Rocha, N.P.: ICF based usability scale: evaluating usability according to the evaluators' perspective about the users' performance. In: Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion—DSAI 2016, pp. 378–383 (2016)
60. Martins, A.I., Rosa, A.F., Queirós, A., Silva, A., Rocha, N.P.: Definition and Validation of the ICF—usability scale. Proc. Comput. Sci. **67**, 132–139 (2015)
61. Martins, A.I., Queirós, A., Cerqueira, M.: Metodologia Living Usability Lab. In: Teixeira, A., Queirós, A., Rocha, N.P. (eds.) Laboratório Vivo de Usabilidade, 1st edn, pp. 215–233. Arc Pubishing, Aveiro (2013)
62. Ribeiro, V.S., Martins, A.I., Queirós, A., Silva, A.G., Rocha, N.P.: AAL@MEO: interactive digital-TV to support home care. Stud. Health Technol. Inform. **217**, 1024–1029 (2015)
63. World MedicalAssociation: Declaration of Helsinki. Law Med Health Care A Publ Am Soc Law Med **19**(3-4), 264–265 (1991)
64. Martins, A.I., Rosa, A.F., Queirós, A., Silva, A., Rocha, N.P.: European Portuguese validation of the system usability scale (SUS). Proc. Comput. Sci. **67**, 293–300 (2015)
65. Ruano, L., et al.: Development of a self-administered web-based test for longitudinal cognitive assessment. Sci. Rep. **6**(1), 19114 (2016)
66. Nokelainen, P.: An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. Educ. Technol. Soc. **9**(2), 178–197 (2006)