

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/228079044>

Usability evaluation: models, methods, and applications

Chapter · January 2010

CITATIONS

15

READS

163

2 authors:



Stefano Federici

University of Perugia. Italy

228 PUBLICATIONS 1,073 CITATIONS

SEE PROFILE



Simone Borsci

University of Twente

81 PUBLICATIONS 458 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



LARTE project [View project](#)



The MATCH Project [View project](#)

International Encyclopedia of Rehabilitation

Copyright © 2010 by the Center for International Rehabilitation Research Information and Exchange (CIRRIE).

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system without the prior written permission of the publisher, except as permitted under the United States Copyright Act of 1976.

Center for International Rehabilitation Research Information and Exchange (CIRRIE)
515 Kimball Tower
University at Buffalo, The State University of New York
Buffalo, NY 14214
E-mail: ub-cirrie@buffalo.edu
Web: <http://cirrie.buffalo.edu>

This publication of the Center for International Rehabilitation Research Information and Exchange is supported by funds received from the National Institute on Disability and Rehabilitation Research of the U.S. Department of Education under grant number H133A050008. The opinions contained in this publication are those of the authors and do not necessarily reflect those of CIRRIE or the Department of Education.

Usability evaluation: models, methods, and applications

Stefano Federici
Department of Human and Education Sciences
University of Perugia, Italy
stefano.federici@unipg.it

Simone Borsci
ECoNA, Interuniversity Centre for Research on Cognitive Processing in Natural and
Artificial Systems
Sapienza University of Rome, Italy
simone.borsci@gmail.com

Introduction

Usability is a recent, even sometimes debated, concept in the human and computer interaction field in which interdisciplinary issues of design engineering and philosophy, cognitive psychology and ergonomics converge. The approach we endorse in order to analyse this interdisciplinary concept is strictly related to the evolution of the evaluation models, methods and applications of usability. At first, we discuss the relation between usability and accessibility as dimensions of the intrasystemic relation between the user and technology, by proposing a definition and a model of evaluation. In the second section, we analyse the mental model's role in system interaction and the distance between them as an outlook to understand usability problems. In the same section, the measures of usability are presented. In the third section, an application of the integrated model of evaluation is described as a step of an assistive technology assessment process in a centre for technical aid. Finally, in the last section, a brief history of human computer interaction evolution and the consequential development of usability and accessibility concepts and methods are presented, in order to understand that usability and accessibility are, at first, linked to technology changes and then strictly related to the spread of technologies.

Definition

Usability is evaluated by the quality of communication (interaction) between a technological product (system) and a user (the one who uses that technological product). The unit of measurement is the user's behaviour (satisfaction, comfort, time spent in performing an action, etc.) in a specific context of use (natural and virtual environment as well as the physical environment where communication between user and technological product takes place). The usability concept and its measurement are strictly connected to that of accessibility ("Web Accessibility"), and the space of the problem, shared by the users, in which the interaction takes place (user technology interaction). Accessibility refers to how a technological product can be used by people regardless of their disability (see here "Web Accessibility"; Web Accessibility Initiative (2010). Usability measures how use is perceived by the user.

Therefore, by improving communication and sharing information among physical, natural and virtual environments, usability is structured on a "User Centred Design", which is an ergonomic approach suited to the biopsychosocial model of disability (WHO 2001). This model complies with the requests and needs of disabled people summed up by the phrase: "nothing about us without us" (Charlton 1998).

The integrated model of usability evaluation

The relation between accessibility and usability is often reduced superficially to that of objectivity and subjectivity (Federici et al. 2005). However, this simplification does not encompass all the

aspects involved in the interaction between technology and user (Annett 2002; Kirakowski 2002). According to this perspective, accessibility refers to the interface code that allows a user to access and achieve the information (e.g. a user can read a text alternative description of a figure with a screen reader). Usability pertains to the subjective perception (satisfaction) of the interface structure's efficiency and effectiveness (e.g. a user is satisfied because they can immediately achieve the information for which they are looking). However, when the relationship between accessibility and usability is defined in this bi-polar way, accessibility might be established as the objective end of the user interaction, while usability might be correlated to subjective aspects, as determined by users' inherent individual differences. From this perspective, a technological product is reduced to a neutral entity that works independently from its user in a neutral environment. As a result, a machine could be perfectly accessible but not usable. Consequently, usability is not only connected to the technological aspects of a machine's functions, but also it pertains to the cognitive and functional aspects of a person's individuality.

In contrast, as Federici et al. (2005) state, when objective and subjective elements refer to accessibility and usability as a user-computer interaction, they cannot be considered as separate entities. Instead they are considered as two different moments, both included in the continuum of empirical observation: each entity is not considered separately from its observer during the interpretative/reconstructive process because the entity is known by the subject only as an observed and perceived object.

From this viewpoint, accessibility and usability are not understood as characteristics of two separate interacting entities but rather as one intrasystemic relation, where both object and subject are just moments in a multiphase process of empirical observation. This prevents the existence of user-less technological products thereby guaranteeing that the accessibility of a machine refers only to the possible entrance and exit of a signal needed to fulfil the task for which it was designed, and that it is in constant relation either to its designer or to its user. In this sense, a machine cannot be accessible and yet unusable at the same time.

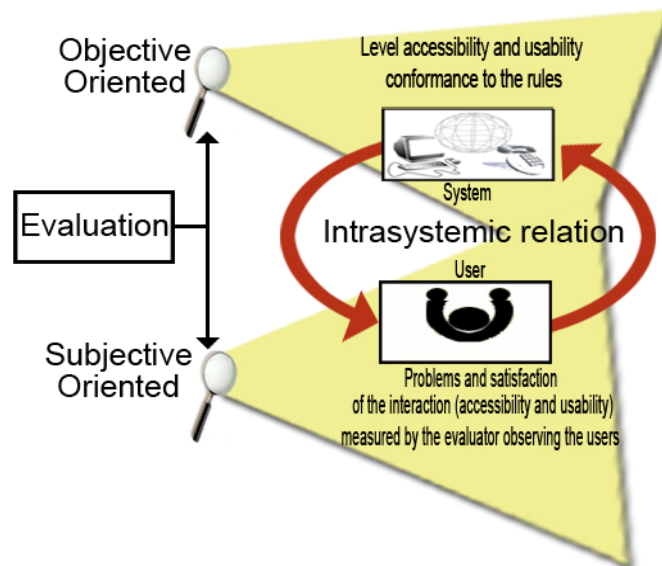


Figure 1 – Integrated model of accessibility and usability evaluation. It shows the possible evaluation perspectives during the evaluation of the intrasystemic dialogue between user and system: the objective-oriented and the subjective-oriented. The interaction evaluation has to take into account not only the properties of a single dimension (the accessibility or the usability), but also the relations that bound the objective part of the interaction to the subjective one (and vice versa). In this context, accessibility and usability are considered as the necessary steps needed for the evaluation of the intrasystemic relation between interface and user.

According to the integrated model, accessibility and usability do not refer to the objective and subjective factors of the user/technology rapport, but rather to a bidirectional way of observing the interaction. Actually, this represents two outlooks from which the one and only observed reality of the user/technology system is drawn. Environment accessibility is therefore based on how it allows the user to initiate and terminate the operation that completes the machine's task (functioning construct), while its usability is based on the user's perception of the user/technology interaction (user performance). The functioning construct of a machine is the basis for standard rules, (e.g. Web Content Accessibility Guidelines) against which accessibility levels are controlled and assessed. The user performance, in relation to the functioning construct of a machine, allows us to deduce scales (e.g. efficiency, satisfaction, cognitive load, helpfulness) of usability scores.

The object of the evaluation cannot be merely reduced to the artefact or to the user: what is evaluated is the functionality of the intrasystemic dialogue between the user (i.e. the subjective dimension of the interaction) and interface (i.e. the objective dimension of the interaction). The accessibility and the usability estimations then need to be understood as the measurements of the users' possibilities to achieve their goals while navigating the given interface. The evaluation of the intrasystemic relation between user and technology includes object-oriented methods as well as subject-oriented methods; still, the overall evaluation cannot be obtained merely by the simple addition of the results coming from the two different methods, but by an evaluation process able to consider and integrate both the accessibility and usability dimensions. An integrated *model of usability evaluation* is compatible with a universal model of disability whereby ability/disability are viewed as a *continuum*. Using ability/disability to refer to an individual, functioning in a real context, can only have a theoretical interest since nobody has a complete absence of disability or complete absence of ability (Bickenbach et al. 1999; WHO 2001; Zola 1989). Therefore, ability or disability is referred to by the activities performed by an individual, originating from the environment and valued by a predetermined *functioning construct*. These activities can change the topology and construct of an environment with respect to the process and measure expected during its functioning.

The usability evaluation method

What the usability evaluates: the problem of use

According to the *integrated model of usability evaluation*, a technological product (e.g. a Web application) is accessible if it meets the success criteria (i.e. WCAG 2.0 guidelines, <http://www.w3.org/TR/WCAG20/>) in such a way as to:

- make content accessible by people regardless of their disability;
- make content usable if a user, regardless of their individual functioning (WHO 2001) and the context of use, perceives no problem in efficiency and effectiveness, and evaluates as satisfactory the use of a technological product (ISO 9241-11).

As Norman (1988) states, an error in the interface is always a "human error" because it depends on an error in the design process. Even when the evaluator detects an error during an objective-oriented analysis, the cause of the error must always be sought in a certain action performed by the designer during the implementation of the interface. During the interaction, the user does not detect an error per se; they can only detect those problems that prevent certain actions in the system being performed. This kind of interaction problem is then observed by the evaluator during the subjective-oriented analysis.

Recently, by enriching Norman's theory on human error, Petrie and Kheir (2007) wrote about the subjective-oriented analysis as follows:

Accessibility and usability problems can be seen as two overlapping sets, which would include three categories:

- Problems that only affect disabled people; these can be termed ‘pure accessibility’ problems;
- Problems that only affect non-disabled people; these can be termed ‘pure usability’ problems;
- Problems that affect both disabled and nondisabled people; these can be termed ‘universal usability’ problems.

In this sense, in order to evaluate properly the intrasystemic relation between user and technology, an interaction evaluation model has to identify: 1) the accessibility conformance by an objective-oriented observation; 2) all kinds of accessibility and usability problems by testing a mixed panels of disabled and non-disabled users with a subjective-oriented analysis; and 3) the user satisfaction in order to complete the subjective-oriented observation.

Usability problems: the distance between the mental models of the designer, the user and the evaluator

Usability problems in interaction are originated by the distance between the models *used to reason with the system, to anticipate its behaviour and to explain why it reacts as it does* (Craik 1943) by the designer of the technological product and by the user of the product. Mental models, considered according to Norman’s definition (1983) as “system causality conveyance”, are those collections of knowledge and skills that lead the subject in the interaction (user) or in the creation (designer) of an interface. From the point of view of the evaluation process, we need to consider that:

- *The developer’s cognitive processes*, involved in the design of the system, are mostly connected to problem-solving strategies to represent knowledge, and to the expertise required in complex task environments. Even though these processes have been analysed, the difficulties due to the “simulation” process have never been properly studied in depth. When designing an interface, the developers simulate how a user would perform in order to achieve their goals; therefore, the designer develops the functions of the system according to their idea of a potential user and of a hypothetical interaction. In this way, the designer is forced to integrate their design skills with their ability to simulate the user’s behaviour. In fact, the application of standard models offered by several international guidelines on accessibility and usability, (even though it is able to, in part, represent the typical user’s behaviour) is not enough to grant the success of a product. Therefore, in order to deliver a satisfactory product, the designer needs to possess, to a certain extent, the ability to “simulate” the possible user’s behaviour. However, the ability to “simulate” someone else’s behaviour consists of one of the hardest and most complex cognitive processes that a human being can perform (Decety and Jackson 2004; Meltzoff and Decety 2003).
- *The user’s cognitive process*: the user’s interaction with the system is quite different from the designer’s process. First of all, it is a fact that by interacting with the interface, the user applies the same cognitive processes used by the designer in the creation of the interface (i.e. problem solving, representation of knowledge and expertise). Thanks to these shared processes, the user is able to “operate” in the interface (i.e. the interface is understandable and usable). However, while the designer applies these shared cognitive processes during the simulation of the behaviour of a hypothetical user (i.e. in the design of the information architecture), the actual user does not need any simulation of the designer’s intention. The user’s cognitive processes are used only to perform actions in the interface and they do not need to be aware of the mental model. Therefore, the actions performed by the user in the

the designer's and the user's mental model cannot be immediately measured, since what we can observe is only the user's interaction with an object created according to the developer's mental model.

In order to analyse the designer's mental model, we need to question them about how they expect a user should interact with the system (usually, this analysis is performed through navigation scenarios). On the other hand, in order to analyse the user's mental model, we only need to observe what they do and how they interact with the system (the user's representation of how the system works and functions). Therefore, to a certain extent, even the user simulates something of how the system works; this kind of simulation, of course, is very different from the designer's simulation: the user does not need to represent and simulate the designer's behaviour or goals, but only how the system works. In this scenario, we can imagine a hypothetical perfect coincidence between the user's expectations about the system and its real functioning (i.e. the system fully satisfies the user), even in the case of an image of a system that does not perfectly express the designer's mental model. In conclusion, while the designer is forced to simulate the behaviour of a hypothetical user, the actual user only needs to simulate the functioning of the system on the basis of their previous experiences and competences. However, even the simulation process of the functioning of a system can generate an interaction distance (i.e. the distance from the user to the developer). This distance usually depends on the fact that people tend to attribute a certain degree of "humanity" to objects, considered as entities capable of performing actions on their own (Gazzaniga 2008). In this scenario, the user (in execution and feedback) would consider the answers that come from the system as the product of an active actor. In general, during the interaction, the user would more easily imagine that the system possesses a certain degree of intentionality rather than try to simulate the designer's model. For example, we often say that "this operative system does not function properly" rather than "this operative system was not designed properly". This fact also helps us to understand why users can only experience problems and not errors: errors are due to bad design or bad implementation of the system, while problems are related to the user's experience of interaction. According to most users, a system does not function when it does not respond properly to their commands. For example, a broken link in the interface would always be experienced by the user as a problem, independent of the fact that the broken link depends on an error in the script (a wrong address or a page that no longer exists) or an error in the pointing procedure of the user (who believes they have clicked the link while actually clicking the background). Even if such errors do exist, they usually remain hidden to the user, who experiences only the "problem" they actually cause. The individual characteristics of each user, as well as their different attributional styles (Abramson et al. 1978; Heider 1958), will then determine how each user will perceive the problem and its causes (e.g. someone could perceive the problem resulting more from an objective error, while someone else could perceive the same problem resulting from a subjective error).

Summarizing, the distance separating the designer and the user in the interaction mostly depends on the different ways of applying their mental model: the designer's simulation of the interaction and the user's interaction with the system. The distance between designer and user can be reduced by the actors' competences to adapt the mental model to the action required (i.e. simulate and interact). The more competent a designer is in simulating the hypothetical user, the smaller the distance separating their respective mental models; the more competent a user is in making system function, the smaller the distance from the conceptual model of the interface (and therefore, from the designer's model).

Both the user and the designer are part of the object we need to measure (the interaction). Therefore, we cannot use, as a standard of measurement, either the expectations of how the system should work (i.e. the designer's perspective), or the experience and satisfaction perceived by the user in the interaction with the system (i.e. the user's perspective). Both these perspectives, in fact, are only a part of what we need to measure. Therefore, we need to find an external unit of measurement able to generalize the relation between the two. This standard unit can be observed only by introducing an external model (i.e. evaluator's model), which is able to evaluate the distance between the two actors involved in the intrasystemic interaction. This model should be created based on the available guidelines and usability evaluation methods (UEMs) for subjective-oriented and objective-oriented observation. Such an evaluator's model will be able to introduce a new conventional unit of measurement, the reliability of which will be granted by the agreement of the international scientific community. Moreover, this new unit of measurement should also respect the principles of economy (i.e. efficiency and efficacy), meaning that it should be able to lessen the costs for the identification of problems in an interaction evaluation.

In this context, the problems are considered as the units of distance between the two mental models (Figure 2). The evaluator's mental model, just like the other two models, is composed of the evaluator's expertise and knowledge. Yet two other components also influence the evaluation process:

1. The international accessibility and design guidelines that determine the standards the evaluator has to take into account when evaluating the interface properties (accessibility and usability);
2. The techniques actually applied by the evaluator to evaluate accessibility, usability and satisfaction. The use of a specific technique forces the evaluator to adapt his or her mental model to the perspective endorsed by the technique. In other words, since the specific techniques used for the evaluation influence the mental model adopted by the evaluator, the evaluation outcome largely depends on the applied techniques.

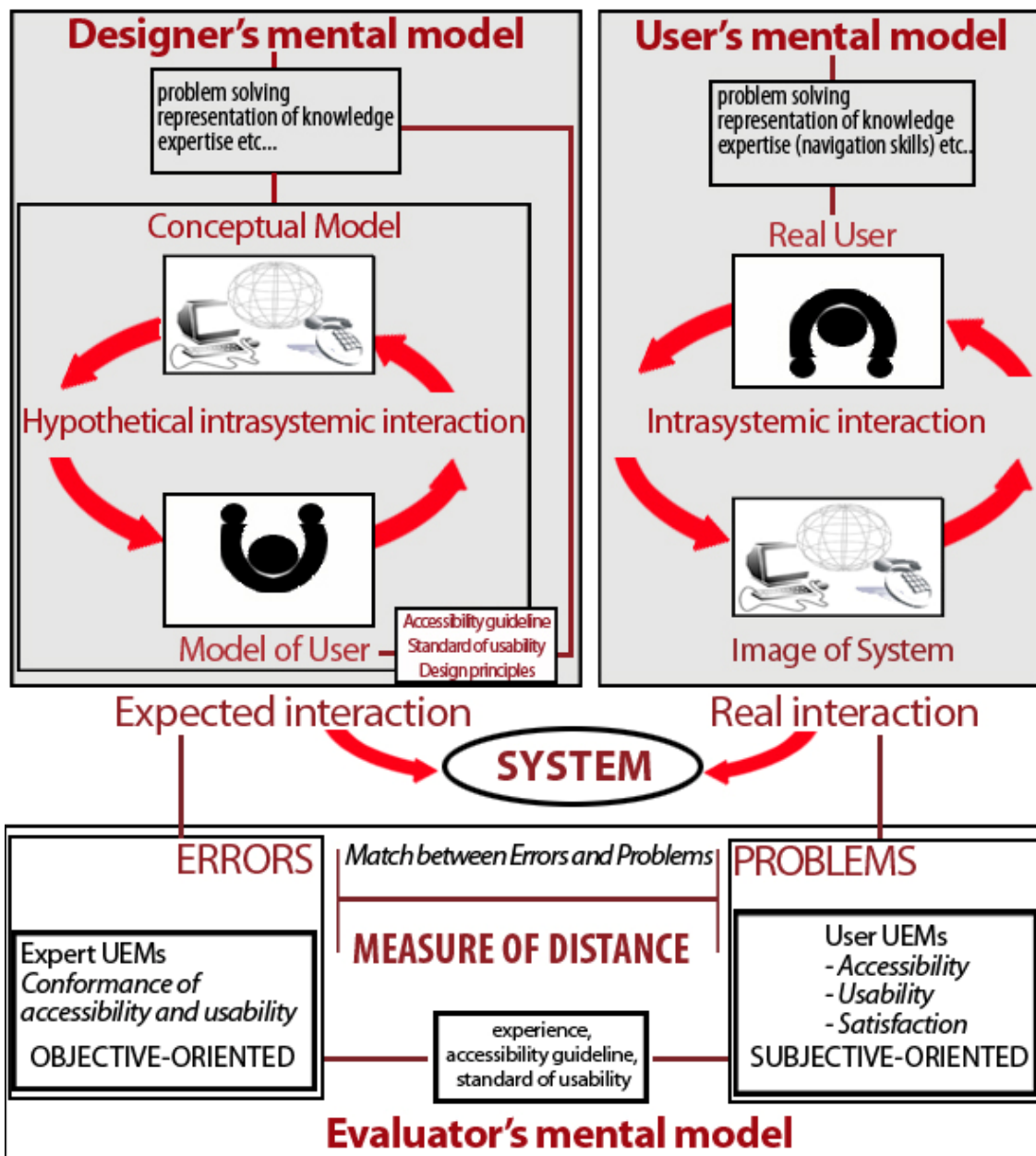


Figure 2 – Integrated model of evaluation and the distance between the mental models. It illustrates the role of the evaluator’s mental model from the perspective of the evaluation: the designer’s mental model is embodied in the system by the conceptual model. The developer designs the system in relation to their experience, representation of knowledge etc. The designer, taking into account standards and guidelines, imagines an expected interaction according to the user model. The real user applies their mental model in the interaction with the image of the system. The user in the “real interaction” experiences problems while the designed system contains the errors. The evaluator’s mental model is involved in the evaluation using the UEM, in order to observe the object, the subject and to measure the distance between the designer’s and the user’s mental models.

At the end of the evaluation process, the evaluator should have obtained: the level of accessibility, the level of usability, the degree of satisfaction and, as an indirect estimation, the measure of the distance between the designer’s and the user’s mental model. This becomes the distance between the technology functions – the conceptual model created by the designer’s mental model – and the function of technology actually perceived by the user.

The evaluator obtains the measure of the interaction distance by matching the errors of the object, analysed by expert analysis (objective-oriented) with the problems observed by the user's evaluation (subjective-oriented). This match shows the distance between the interaction imagined by the designer for a hypothetical user and the interaction perceived by the real user. In this case, the problem, perceived by the real user, is the unit of the interaction distance measurement.

How does the integrated model measure usability?

In order to consider all the aspects of human and computer interaction, an integrated model of usability evaluation is composed of three steps, with different accessibility and usability evaluation methods:

1. *System evaluation (the objective-oriented effectiveness and efficiency of the system)*: the evaluation of the objective aspects of the interface (accessibility and usability). Expert evaluation techniques are used to assess the interface according to a comparison with some standard design models for both accessibility and usability (WCAG rules, heuristics, etc.). The data obtained from these techniques will, in turn, be used as a baseline to be compared with the data collected through the user-based tests. The evaluator, in the objective-oriented observation, measures the conformance with the guidelines (i.e. WCAG, heuristic lists and design principles) of the hypothetical interaction designed by the developer (i.e. the image of the system). This evaluation not only concerns the identification of errors through a conformance analysis, but it is also a reconstruction of the designer's mental model from the evaluator's perspective.
2. *User evaluation of the interaction (the subjective-oriented analysis of effectiveness and efficiency)*: the evaluation of the subjective aspects of the interface. User techniques are used to identify the problems (by distinguishing the *pure accessibility*, the *pure usability* and the *universal usability* problems) as perceived by the user in a real interaction. The data are collected and then matched with those gathered from experts' tests (i.e. the system evaluation) in order to define the efficacy and efficiency of the evaluation process. These measures allow the evaluator to observe the distance between the mental model of the real users (i.e. the real perceived problems) and the reconstructed designer's mental model (i.e. the objective errors identified by the evaluator).
3. *User evaluation of satisfaction (the subjective-oriented analysis of user satisfaction)*: the evaluation of subjective aspects of the interaction (satisfaction). Psychometric tools (e.g. questionnaires) are used to collect qualitative data and information about the system. These data represent an indirect measurement of the efficacy and efficiency of the system, and they are essential for including in the evaluation process. They are subjective aspects perceived by users that cannot be objectively quantified. These measures allow the evaluator to provide an indirect estimate of the distance between the real user and the hypothetical user (i.e. the user model) imagined by the designer. By using the satisfaction analysis, the evaluator can measure the distance between the two mental models, also taking into account (directly or indirectly, depending on the instruments) the user's skills, otherwise excluded from the evaluation.

Each step of the integrated model of the usability evaluation is composed of a set of usability and accessibility evaluation methods. First, in the system evaluation, the *expert evaluators* – who are supposed to detect barriers that usually prevent interaction – test the interface using an analysis of the compliance to the standard guidelines and they apply an expert evaluation technique (e.g. heuristic analysis, cognitive walkthrough, etc.). Second, in the user evaluation of the system, the *users* – who can estimate the extent to which a detected barrier is actually preventing their

navigation – are tested using a specific technique: task analysis, where users think aloud or (for screen-reader users) provide partial concurrent thinking aloud (Federici et al. 2010a; Federici et al. 2010b). Finally, user satisfaction is measured using a questionnaire.

To complete the process, the *coordinator of the evaluation (meta-evaluator)* integrates the results gathered from the expert-based tests with those from the user-based tests, by performing an evaluation of the evaluation (meta-evaluation). This is undertaken by indicating which problems, detected by the expert evaluators, are real ones (i.e. those problems that were also detected during the user-based tests), and identifying which problems, identified by disabled and non-disabled users, address *pure accessibility*, *pure usability* or *universal usability* problems (Petrie and Kheir 2007).

Usability evaluation applications: how can the integrated model be applied in the assessment of technological aid?

The integrated model of usability evaluation as a user-driven model could be helpfully applied, as Scherer and Federici (in press) suggest, in an assistive technology (AT) assessment process in a centre for technical aid. The authors define AT as follows:

The Assistive Technology Assessment is a user-driven process through which the selection of one or more technological aids for an assistive solution is facilitated by the comprehensive utilization of clinical measures, functional analysis, and psycho-socio-environmental evaluations that address, in a specific context of use, the personal well-being of the user through the best matching of user/client and assistive solution.

According to this definition, an assistive solution requires a thorough evaluation of the interaction among users, technologies, the physical surroundings and the environment, by requiring professional skills in usability evaluation methods and techniques.

Let us provide an example of the complexity of accessibility and usability evaluations in relation to the interaction between technology and the environment, by quoting a study by Gossett et al. (2009) where a wheelchair user tries to access a lift:

An example is the treatment of floor surfaces. Carpet minimizes noise, offsetting occasional loud noises. However, carpet can pose problems for wheelchair users by causing their chairs to ‘pull’ like a car out of alignment. In addition, carpet pattern can become an issue for people with seizure disorders. Carpet also accumulates dust and dirt, which along with some cleaning products can pose a problem for individuals with MCS. In the end decision, carpet was chosen that didn’t ‘pull’, had a non-symmetrical pattern and was washable without chemicals.

It could be possible that a good AT, namely accessible and usable, in a good environment has a bad match with an assistive solution for an individual in a specific context, by creating low user satisfaction. In fact, if the environment does not match an AT, which accomplishes the usability and the accessibility rules, the environment is not accessible and usable at all (e.g. the lift cannot contain the wheelchair). However, at the same time, if the environment is usable and accessible but the AT does not match the rule of accessibility and usability, the AT does not allow users to experience a satisfactory interaction with the environment. In this extreme case, the user can access the lift with the wheelchair and the buttons of the lift are usable and accessible, but the wheelchair features do not allow users to push the buttons: in this case, AT does not match the environment. The universal design and the rules of accessibility, usability and sustainability have to work together at an environmental and AT-design level, in order to grant the best match between the user’s AT and the

environment in use. However, without a user-based evaluation, the assessment of the match may not be complete.

In conclusion, as Scherer and Federici (in press) state, the assessment of the interaction between user and environment through the AT is an intrasystemic relation. It can be evaluated by an integrated model of usability and accessibility, which is able to define the quality of the match, or can determine whether the environment has to be changed in order to improve the intrasystemic relation.

Brief history of usability

The historical review of human computer interaction as the process through which the conditions (accessibility and usability) for the dialogue between “humans” and “computers” are developing, may be divided into three periods. During the first period (1950-1963), there was no need for accessibility and usability in the human-computer interface because programmers (the creators of the interface) were at that time, also the users of the software. During the second period (1963-1984), there was no change for the users/programmers; this period was characterized by an evolution of systems and models of interaction. During the third period, (post-1985), accessibility and usability issues became more central, due to the spread of the personal computer and the Internet, with the consequent distinction between users and programmers and the needs of users to access and use the information on the World Wide Web.

1950 to 1963: the programmer is the user. Since the end of the 1950s, technological and computer tools were created with specific functions determined by the different ideal interaction models elaborated by programmers. The principal demand was to get either better calculation performances or better machine functions, with a focus on the management and control of technology. The operator/user had to adapt to the formal rules of the system or technology, which was introduced by the designers, whether it was a calculator, an appliance or industrial machinery. The operator/user could control the technology with a panel meant to be used only for two operations: the correction of the machine’s functions (i.e. debugging) and the input of command lines into the system. The interaction was built up following the ideas of the Command Line Interface (CLI). The consumer/user was forced to learn the commands and input them using a keyboard. The interface was substantially textual and, usually, the interaction was limited to inserting data into the system. In the early 1960s, the demand for a new interaction model started to grow. This new need was made possible by the introduction of new hardware elements (Dix et al. 2004) and by a redefinition of the industrial operators. In fact, in this historical period, the operators of industrial technological products could no longer be considered mere objects, embedded in the assemblage process. Instead, they started to be considered as subjects with rights, as well as consumers of the production process. This historical step marked the passage from operators to users. These new perspectives developed the need to improve the conditions of interactive exchange. There was less physical involvement from people; instead, users interacted with machinery, which asked for instructions and transmitted information about the development process. The ergonomic attention moved from the muscular to the perceptive load.

Users, sitting in front of radar screens, dashboards or command panels, were involved in new interactions with technology. As a consequence, their cognitive workloads were much heavier: attention decreased, while detention time regarding signal and answer times increased. In this new scenario, the reduction of error numbers during the interaction with the system (especially in a work environment) became a major issue for HCI researchers and practitioners. With the progressive automation of the work process, a larger element of information, procedures, strategies and solutions was developed through machines. In this way, operators were released of the workload that was now being undertaken by the machine. This allowed for the possibility to focus their

abilities on complex cognitive tasks. As a consequence, in the 1960s, ergonomic studies switched their attention from the physical structure of the work environment to its psychological and cognitive aspects.

1963 to 1984: evolution of HCI models. In 1963, at the Massachusetts Institute of Technology (MIT), Sutherland (1964), the then Vice President and Fellow at Sun Microsystems, developed the first interactive graphic user interface, Sketchpad. This system consisted of the direct manipulation of some graphic objects through an optic pen with which the user could create and move the graphic elements, receive graphic feedback and modify the interface setup (Sutherland 2003). The idea of direct manipulation helped to overcome the CLI, opening new scenarios for HCI, as Sutherland showed (2003): “The Sketchpad system, by eliminating typed statements (except for legends) in favour of line drawings, opens up a new area of man-machine communication.” The graphic interface development changed the relation between users and technology; in fact, the possibility of a graphic input system without command lines forced users to increase their adaptation abilities to the interface. In 1983, Shneiderman, professor in Computer Science at the Human-Computer Interaction Laboratory at the University of Maryland, defined the main features of direct interface manipulation in graphic interface:

- Continuous representation of the object of interest.
- Physical actions (movement and selection by mouse, joystick, touch screen, etc.) or labelled button presses instead of complex syntax.
- Rapid, incremental, reversible operations whose impact on the object of interest is immediately visible.
- Layered or spiral approach to learning that permits usage with minimal knowledge. Novices can learn a modest and useful set of commands, which they can exercise till they become an ‘expert’ at level I of the system. After obtaining reinforcing feedback from successful operation, users can gracefully expand their knowledge of features and gain fluency.

These principles were further developed by Hutchins, Hollan and Norman (1985). They started from the idea that interaction quality must be linked to the affordances concept (Gibson 1979). Affordances are to be understood as all the latent “action possibilities” in the environment that are objectively measurable and independent from the subjects’ ability to recognize them; even though affordances are not dependant on the subjects’ recognition, they are still related to the actors and their skills. Gibson’s affordance concept was developed to explain subjects’ interactions in physical environments. Hutchinson, Holland and Norman (1985) extended its range to virtual environments also considering, along with physical capabilities, other aspects related to the human-computer interaction, such as: actors’ goals, plans, values, beliefs and past experiences.

An interface, as a place of functions and variables, is designed for operating a system, starting from the user inputs (click/query); in a simple virtual environment (i.e. interface), the user should be able to immediately develop messages or recall useful knowledge from the machine’s memory.

The user, during interaction with a graphic interface like Sketchpad, must work with multiple aspects of knowledge, such as action, objects and manifold levels of syntactic knowledge. Thanks to Shneiderman’s work at the Palo Alto Research Centre and to the subsequent formalization of the Graphical User Interface (GUI), the first window systems were developed in the second half of the 1980s (Smalltalk and InterLisp).

1984 until now: the personal computer and Internet era. In 1984, at MIT, client/server architecture was produced in order to work flexibly on an interactive windows information system. It was named X Window System. In 1985, the GUI system became available to regular consumers when

the first version of MS-Windows was released. Finally, during the 1990s, the WIMP (Windows, Icons, Menus and Pointer) type of interface became the most operative system, until today. The transformations in hardware and software over the last thirty years have promoted the development of graphic elements and have imposed the use of an interaction code, which is based on symbolic and spatial elements and not only on a linear language. Indeed, in the WIMP interface, the information organized in the virtual space, is spread both by the succession and by the order of content (i.e. the drop-down menus, the toolbars and the textual guides). At the same time, the information follows the rules of temporality, irreversibility, horizontality, uniformity, causality and fragmentation of the written language (De Kerckhove 1995). In fact, users interact with information both through graphic content and through organization of space and form (i.e. the concept of a desktop, the icons, the radial menus and the interaction through the mouse, etc.).

Thereby, users must extend their cognitive faculties from their logical-analytical, linear and sequential abilities to figurative, spatial, gestaltic and circular abilities (De Kerckhove 1995). While in the CLI, both the designer and user are bound by the same communicative code, which is logical and analytical (i.e. typed text), in the WIMP, the information is also communicated through graphic-spatial codes. The WIMP communicative code no longer coincides with the system code. The information relevance of an icon, for example, is not only linked to its content and functions, but also to its position on the screen or to the environment during the interface. The information content of CLI was entirely lacking of graphic context and the facilitation for the users' interactions could be reduced to few ergonomic rules (size of the screen, size and shape of the fonts, brightness, etc.). On the other hand, in the WIMP model, the content is linked to the graphic-spatial context, so introducing the user to greater possibilities for the interpretation of content and environment (position, clarity of the symbol, graphic effects, etc.). With WIMP interfaces, in order to guarantee the functionality of the environment, the developers' function is no longer limited to the verification of the syntactic correctness of the code. In this way, the users' interpretative analysis of the code becomes part of the design process itself.

The historical transition from CLI to the GUI and the diffusion of personal computers raised usability and accessibility problems, and as a consequence, researchers developed new evaluation methodologies in order to develop friendly systems that were able to support performance and the spread of information on the World Wide Web. The evaluation of these products, and in particular the evaluation of the interaction with users, became an important way to develop the technology. In fact, from the 1980s, the first kind of usability test, known as "laboratory usability testing", quickly became the primary usability evaluation method for examining a new or modified interface. Developers considered laboratory testing as a way to minimize costs of service calls, to increase sales through the design of a more competitive product (by minimizing risk) and creating a historical record of usability benchmarks for future releases (Rubin 1994). In addition to the users' subjective evaluations, laboratory testing measures speed, accuracy and errors of users' performances. Methods for collecting data beyond user performance included: verbal protocols (Ericsson and Simon 1984, 1987), critical incident reporting (del Galdo et al. 1986) and user satisfaction ratings (Chin et al. 1988). More recently, in the 1990s, developers explored other evaluation methods in an attempt to decrease further the costs and time required for traditional usability testing. In addition, since usability testing tended to occur late in the design process, developers were motivated to search for new methods that could be used with the prototypes developed in the early design process (Bradford 1994; Marchetti 1994).

Some of the most popular expert-based UEMs include guidelines for the evaluators based on the rules of the interaction design. Examples include those by Smith and Mosier (1986), which created guidelines for the software of the United State Air force, heuristic evaluation (Nielsen and Molich 1990), cognitive walkthroughs (Lewis et al. 1990; Wharton et al. 1992), usability walkthroughs

(Bias 1991), formal usability inspections (Kahn and Prail 1994) and heuristic walkthroughs (Sears 1997). Practitioners are far from settled on a uniform UEM, and researchers are far from agreeing on a standard for evaluating and comparing UEMs. In order to overcome these difficulties and misunderstandings between different UEMs, Gray and Salzman (1998) supported strongly the introduction of comparative studies in the usability evaluation field. As the subsequent debate finally demonstrated (Olson and Moran 1998), only the application of a rigorous experimental methodology could lead to a future formalization of several standard rules for usability evaluation. The comparative review carried out by Hartson et al. (2001) gives us an accurate analysis of expert and user based evaluation methods, but this analysis is not extended to the model-based methods and properties. Among the model-based evaluation technique, GOMS (Goals, Operators, Methods and Selection rules) is one of the most well known. This technique was developed by three pioneers of HCI: Card, Senior Research Fellow at Xerox PARC; Moran, distinguished engineer at the IBM Almaden Research Center and Newel, a researcher in computer science and cognitive psychology at the RAND Corporation (1983). This evaluation technique introduced an ideal model of the human process of information elaboration able to represent the human perception and elaboration of external stimuli, and to predict the human performance during the interaction with the technology (Kieras 2003).

Another model, strictly related to the GOMS, is the cognitive complexity theory (CCT) (Bovair et al. 1990). This could predict performances and learning time of users during an interaction with a certain technology. However, Polson compared the results achieved with the CCT model with those obtained with an expert-based model, and underlined that the use of this model can supplement, but not replace, cognitive walkthrough (Polson et al. 1992).

In discussing these techniques, it may be necessary to underline the important role played by Fitts' law (1954). This law is a model of human movement which predict the time required for a subject to move rapidly to a target area, expressed as a logarithmic function of the distance to the target area and its size. Fitts' law was applied on different studies of HCI in order to calculate the index of performance during the interaction movements, which manipulate the interfaces through different kinds of mice and pointers (MacKenzie 1992). As Dillon (2001) states:

“Finally, there are good reasons for thinking that the best approach to evaluating usability is to combine methods – e.g., using the expert-based approach to identify problems and inform the design of a user-based test scenario, since the overlap between the outputs of these methods is only partial, and a user-based test normally cannot cover as much of the interface as an expert-based method”

According to the latest studies in the scientific literature, the integration of evaluation models emerges as the only possible solution in order to take account of the evaluation process and the totality of aspects involved in the HCI.

Today, the debate on usability measurements is centred on the possibility of the evaluation methods' integration and the role played by subjectivity in the assessment of human-computer interaction, as we have discussed in the previous sections.

Concluding remarks

Usability and accessibility represent two distinct concepts, in a similar way to how human interaction is only approached from the perspective of the engineering design process of a system. Conversely, when an evaluation perspective is implemented, the boundaries between usability and

accessibility become two different moments, both included in the continuum of empirical observation. In fact, according to the integrated model, accessibility and usability do not refer to the objective and subjective factors of the user/technology rapport, but rather to a bidirectional way of observing the interaction.

In this context, our proposal of an integrated evaluation model represents the perspective of the evaluator, aiming to assess all the aspects of the intrasystemic relation by focusing on the user's behaviour. According to our model of usability evaluation (user-driven), the role of users, with regard to their individual functioning, becomes central since the identification of problems is the ultimate unit of measurement of usability assessment.

References

- Abramson L, Seligman M, Teasdale J. 1978. Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology* 87(1):49-74.
- Annett J. 2002. Subjective rating scales: Science or art? *Ergonomics* 45(14):966-987.
- Bias R. 1991. Walkthroughs: Efficient collaborative testing. *IEEE Software* 8(5):94-95.
- Bickenbach JE, Chatterji S, Badley EM, Üstün TB. 1999. Models of disablement, universalism and the international classification of impairments, disabilities and handicaps. *Social Science and Medicine* 48(9):1173-1187.
- Bovair S, Kieras DE, Polson PG. 1990. The acquisition and performance of text-editing skill: a cognitive complexity analysis. *Human- Computer Interaction* 5(1):1-48.
- Bradford JS. 1994. Evaluating high-level design: Synergistic use of inspection and usability methods for evaluating early software designs. In: J Nielsen, RL Mack, editors. *Usability inspection methods*. New York: Wiley & Sons. p. 235–253.
- Card SK, Newell A, Moran TP. 1983. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc.
- Charlton JI. 1998. *Nothing About Us Without Us: Disability Oppression and Empowerment*. Berkeley and Los Angeles: University of California Press.
- Chin JP, Diehl VA, Norman KL. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of the SIGCHI conference on Human factors in computing systems*. Washington (DC): ACM. p. 213-218.
- Craik K. 1943. *The Nature of Exploration*. Cambridge (UK): Cambridge University Press.
- De Kerckhove D. 1995. *The skin of culture: Investigating the new electronic reality*. Toronto: Somerville.
- Decety J, Jackson PL. 2004. The Functional Architecture of Human Empathy. *Behavioral & Cognitive Neuroscience Reviews* 3(2):71-100.
- del Galdo EM, Williges RC, Williges BH, Wixon DR. 1986. An Evaluation of Critical Incidents for Software Documentation Design. *Human Factors and Ergonomics Society Annual Meeting Proceedings* 30(1):19-23(5).

- Dillon A. 2001. Beyond Usability: Process, Outcome, and Affect in Human Computer Interactions [Au delà de la convivialité: processus, résultats, et affect, dans les interactions personne-machine]. *Canadian journal of information and library science* 26(4):57-69.
- Dix A, Finlay J, Abowd DA, Beale R. 2004. *Interazione uomo macchina*. Milano: Mc Graw-Hill.
- Ericsson KA, Simon HA. 1984. *Protocol analysis: Verbal reports as data*. Cambridge (MA): MIT Press.
- Ericsson KA, Simon HA. 1987. Verbal reports on thinking. In: C Faerch, G Kasper, editors. *Introspection in Second Language Research*. Clevedon: Multilingual Matters. p. 24-53.
- Federici S, Borsci S, Mele ML. 2010a. Usability evaluation with screen reader users: A video presentation of the PCTA's experimental setting and rules. *Cognitive Processing* 11(3).
- Federici S, Borsci S, Stamerra G. 2010b. Web usability evaluation with screen reader users: Implementation of the Partial Concurrent Thinking Aloud technique. *Cognitive Processing* 11(3).
- Federici S, Micangeli A, Ruspantini I, Borgianni S, Pasqualotto E, Olivetti Belardinelli M. 2005. Checking an integrated model of web accessibility and usability evaluation for disabled people. *Disability & Rehabilitation: An International Multidisciplinary Journal* 27(13):781-790.
- Fitts PM. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology: General* 47(6):381-391.
- Gazzaniga MS. 2008. *Human: The Science Behind What Makes Your Brain Unique*. New York: Harper Collins.
- Gibson JJ. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Gray WD, Salzman MC. 1998. Damaged merchandise? a review of experiments that compare usability evaluation methods. *Human- Computer Interaction* 13(3):203-261.
- Hartson HR, Andre TS, Williges RC. 2001. Criteria For Evaluating Usability Evaluation Methods. *International Journal of Human-Computer Interaction* 13(4):373-410.
- Heider F. 1958. *The Psychology of Interpersonal Relations*. Hillsdale (NJ): Lawrence Erlbaum Associates.
- Hutchins EL, Hollan JD, Norman DA. 1985. Direct Manipulation Interfaces. *Human-Computer Interaction* 1(4):311-338.
- International Standards Organization (ISO). 1998. ISO 9241-11: Ergonomic requirements for office work with visual display terminals. Geneva: International Organization for Standardization.
- Kahn MJ, Prail A. 1994. Formal Usability Inspections. In: J Nielsen, RL Mack, editors. *Usability Inspection Methods*. New York: John Wiley & Sons. p. 141-172.

- Kieras D. 2003. GOMS Models for Task Analysis. In: D Diaper, N Stanton, editors. *The Handbook of Task Analysis for Human-Computer Interaction*. New York: Lawrence Erlbaum Associates. p. 83-116.
- Kirakowski J. 2002. Is ergonomics empirical? *Ergonomics* 45(14-15):995-997.
- Lewis C, Polson PG, Wharton C, Rieman J. 1990. Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*. Seattle (WA): ACM. p. 235-242.
- MacKenzie IS. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction* 7(1):91-139.
- Marchetti R. 1994. Using usability inspections to find usability problems early in the lifecycle. In: *Pacific Northwest Software Quality Conference; 1994*. Palo Alto (CA): Hewlett Packard. p. 1-19.
- Meltzoff AN, Decety J. 2003. What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London* 358(1431):491-500.
- Nielsen J, Molich R. 1990. Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*. Seattle (WA): ACM. p. 249-256.
- Norman DA. 1983. Some Observations on Mental Models. In: Gentner D, Steven A, editors. *Mental Models*. Hillsdale (NJ): Lawrence Earlbaum Associates. p. 7-14.
- Norman DA. 1988. *The psychology of everyday things*. New York: Basic Books.
- Olson GM, Moran TP. 1998. Commentary on "Damaged merchandise?". *Human-Computer Interaction* 13(3):263-323.
- Petrie H, Kheir O. 2007. The relationship between accessibility and usability of websites. *Proceedings of the SIGCHI conference on Human factors in computing systems*. San Jose (CA): ACM. p. 397-406.
- Polson PG, Lewis C, Rieman J, Wharton C. 1992. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies* 36(5):741-773.
- Rubin J. 1994. *Handbook of usability testing: How to plan, design, and conduct effective tests*. New York: Wiley and Sons.
- Scherer MJ, Federici S. in press. *Assistive Technology Assessment: A Handbook for Professionals in Disability, Rehabilitation and Health Professions*. London: CRC Press.
- Sears A. 1997. Heuristic Walkthroughs: Finding the Problems Without the Noise. *International Journal of Human-Computer Interaction* 9(3):213-234.

- Shneiderman B. 1983. Direct Manipulation: A Step Beyond Programming Languages. *Computer* 16(8):57-69.
- Smith SL, Mosier JN. 1986. Guidelines for designing user interface software. Bedford (MA): MITRE Corporation. No. ESD- TR-86-278/MTR 10090.
- Sutherland IE. 1964. Sketch pad a man-machine graphical communication system. In. *Proceedings of the SHARE design automation workshop*; 1964 1964: ACM Press.
- Sutherland IE. 2003. Sketchpad: A man-machine graphical communication system [Internet]. Cambridge: University of Cambridge; 2003 [cited 2010 June 3].
- Web Accessibility Initiative (WAI). 2010. The World Wide Web Consortium (W3C) [Internet]. [cited 2010 June 3]. Available from: <http://www.w3.org/>
- Wharton C, Bradford J, Jeffries R, Franzke M. 1992. Applying cognitive walkthroughs to more complex user interfaces: experiences, issues, and recommendations. Proceedings of the SIGCHI conference on Human factors in computing systems. Monterey (CA): ACM. p. 381-388.
- World Health Organization (WHO). 2001. ICF: International Classification of Functioning, Disability, and Health. Geneva: WHO.
- Zola IK. 1989. Toward the Necessary Universalizing of a Disability Policy. *Milbank Quarterly* 67 supplement:401-428.