Usability as a Key Quality Characteristic for Developing CAAD Tools and Environments

Burak Pak[1], Johan Verbeke[2]
[1,2]Sint-Lucas School of Architecture, Faculty of Architecture and Arts, Association KU Leuven Belgium
[1]http://blog.associatie.kuleuven.be/burakpak
[1]burak.pak@architectuur.sintlucas.wenk.be, [2]johan.verbeke@architectuur.sintlucas.wenk.be

Abstract. *In this paper, we will stress the importance of usability as a key quality characteristic for the Computer Aided Architectural Design (CAAD) software prototypes. We claim that usability evaluation practices can assist the integration of human factors and the accommodation of local differences. These practices are not solely limited to interface tests, but they can also provide valuable information on the possible added values of CAAD software prototypes, increase the overall product quality and thus contribute to the sustainable development of the CAAD research field.*
In this context, we aim to initiate a constructive discussion on this topic by reviewing various usability frameworks and highlighting possible opportunities and challenges of applicable evaluation methods. Consequent to this discussion, we will elaborate on our recent findings relating to the reliability and effectiveness of particular evaluation methods applied to a web-based geographic virtual environment prototype.
In conclusion, we will introduce a new "design usability" framework that is suitable for CAAD software development; which suggests a variety of design usability quality characteristics, cost-effective evaluation methods and possible influence factors in the evaluation process.
Keywords. *Usability; Quality in Use; Evaluation; CAAD Software Development; Human Factors.*

INTRODUCTION

Fifteen years ago, in his seminal paper "CAAD's Seven Deadly Sins", professor Tom Maver (1995) presented a critical view on the direction of research and development in CAAD. He introduced seven topics of criticism: *overestimating the short term impacts and underestimating the longer term impacts (macro-myopia), re-visiting ideas (déjà vu), absence of a core research discipline (xenophilia), discarding fitness-for-purpose, cost-effectiveness and environmental sustainability (unsustainability), generation of hypotheses without rudimentary testing (failure to validate), failure to criticize and finally, insufficient investigation of usability and functionality in teaching or practice (failure to evaluate).*

Since then, CAAD research has made significant progress. The last decade has witnessed a widespread use of scripting languages, open source software and application programming interfaces followed by an admirable number of researchers who developed, shared, tested and implemented new CAAD tools. Quite often, these tools have been evaluated and presented together with novel design products in the form of case studies as "proofs of concepts". These studies are essential because they demonstrate the utility of the tools by offering visible evidence related to the integration of experimental CAAD research into design practices and/or the profession. Thus, they inspire future studies.

On the other hand, evaluation of usability still remains a difficult and ill-defined topic, especially when it comes to the experimental CAAD tools and design environments.

USABILITY AND QUALITY IN USE CHARACTERISTICS IN VARIOUS QUALITY FRAMEWORKS AND MODELS

In this section, we will critically review and compare definitions of usability characteristics in various frameworks with the purpose of initiating a discussion on the applicability of these definitions to the CAAD field.

Our comprehensive literature study on usability points out to an inflation of definitions and methods proposed for software usability evaluation. Authorities such as the International Standards Organization (ISO) issued more than 50 standards related to software usability and Human Computer Interaction (HCI) (Bevan, 2006). Various other descriptive theoretical usability frameworks were also suggested by Nielsen and Mack (1994), Norman (2004), Davis et al. (1989), Venkatesh et al. (2003).

To begin with, HCI-related ISO standards are primarily shaped around four topics: quality in use, product quality, process quality and organizational capability. Concerning the usability evaluation of CAAD software prototypes, "quality in use" stands out as the most relevant quality characteristic described in ISO/IEC 25000 Series (2006).

In this framework (Figure 1), quality in use is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO/IEC 25062, 2006).

Besides the international standards organizations, Jakob Nielsen is a leading figure who has produced a significant number of publications on usability evaluation and engineering. In 1994, Nielsen and Mack proposed "a framework of system acceptability" to describe a complete set of quality characteristics (Figure 2). They defined five different usability attributes: learnability, use efficiency, memorability, freedom from errors and subjective pleasantness. The "system acceptability" in Nielsen and
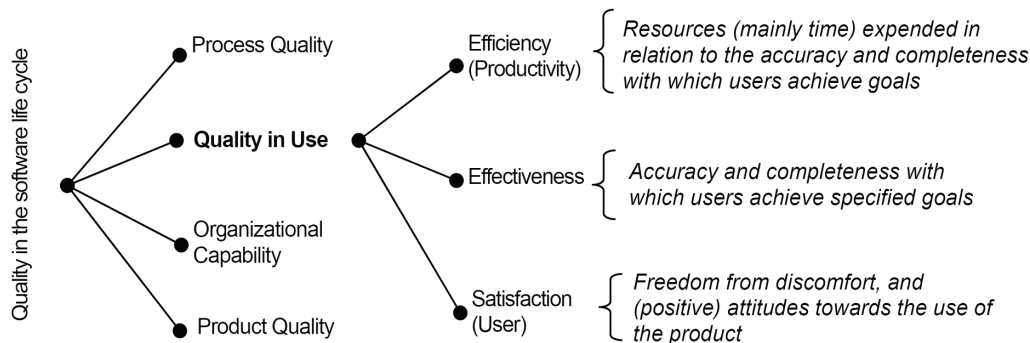

Figure 1
A concept map (created by the authors) illustrating the Quality in Use Characteristics in ISO/IEC 25000 Series Software Product Quality Requirements and Evaluation Standards.

Mack's (1994) framework can be compared to "quality in the software life cycle" (ISO 25000 Series, 2006), whereas "practical acceptability" and "utility characteristics" contain similar topics as "product quality" in ISO 25000 Series. Overall, it would not be wrong to state that these two frameworks are complementary in their interpretation of usability.

In Nielsen and Mack's framework, subjective pleasantness refers directly to user attitudes and satisfaction. It is suggested to be measured by post-task satisfaction surveys. The usability attribute efficiency is similar to productivity in ISO 25000 Series and described as "the time required to accomplish the designated tasks" whereas memorability refers to the efficiency of the casual users that are away from the system for a specified amount of time (Table 1).

Learnability is related to the time needed for the users to perform the required number of tasks (the required number of tasks is a threshold value determined by the evaluators). Nielsen and Mack's

(1994) final attribute error is defined as any action that does not reach the desired goal. The system's error rate is calculated by counting the number of such actions made by users while performing the designated tasks.

There are various other incomparable usability frameworks that introduce different quality characteristics with different scopes. For instance, in the User Experience Model (UX), Norman (2004) asserts emotions and aesthetics as separate (but dependent) quality characteristics besides utility and usability.

On the other hand, the "Technology Acceptance Model" by Davis et al. (1989) suggests perceived usefulness as primary indicators for system usability. The "system usability" in this framework refers to acceptance and use. It is interpreted as a behavior that is predicted rather than observed.

In the extended version of this theory, the "Unified Theory of Acceptance and Use of Technology" Venkatesh et al. (2003) included performance expectancy, effort expectancy, social influence, facilitating conditions, voluntariness of use, gender, age and experience to these criteria (Figure 3). This theory is important because it stresses the possible influences of the social, individual and facilitating factors on system usability.

This is in line with the ACCOLADE project (2001) where it was found that social and behavioral factors are utterly important when using software for collaborative action.

Figure 2
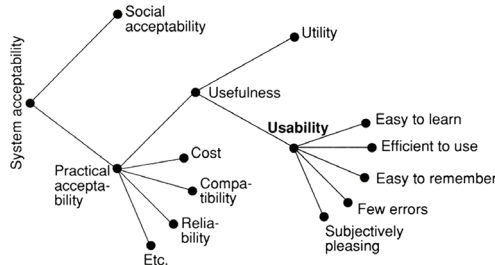*The framework of system acceptability quality characteristics (Nielsen and Mack, 1994)*

Table 1
Comparison of ISO 25000 "Quality in Use Characteristics" and Nielsen and Mack's (1994) "Usability Attributes".

ISO 25000 Series "Quality in Use Characteristics"	Nielsen and Mack's (1994) framework "Usability Attributes"
User Satisfaction	Subjective Pleasantness
Efficiency	Efficiency of use
No direct equivalence	Memorability Memorability is the change in efficiency after the casual users are away from the system for some time
Effectiveness Task completion rates by time, freedom from errors, learnability, number of assists…	No direct equivalence ~Learnability ~Freedom from errors

Based on the theories and models on usability and quality in use characteristics reviewed above, it is possible to construct specific measures and metrics applicable in CAAD software development and evaluate them using a variety of methods. In the following topic, we will discuss the opportunities and challenges of these methods.

USABILITY OF USABILITY EVALUATION METHODS FOR THE DEVELOPMENT OF CAAD TOOLS AND ENVIRONMENTS: OPPORTUNITIES AND CHALLENGES

A comprehensive literature review reveals that there is a variety of usability evaluation methods. Each of these methods provides valuable opportunities but also present challenges for usability evaluation (Table 2). Moreover, some of these methods may be considered as more suitable to be utilized in the early stages of CAAD software development -especially in the academic field- as they are relatively easy to conduct, efficient, effective and reliable; thus satisfactory.

To begin with, *task observation* is an effective method for evaluating how well the software facilitates users to accomplish a number of tasks. In this method, the evaluators choose around ten vital tasks to be completed by the representative users. These tasks are then given to the users in a preferably controlled space (such as a fixed laboratory or a conference room) and they are observed by the evaluator and/or a video camera. The evaluator times and records the specific indicators either during the test or after the test.

This method can be performed with a limited number of participants. In an experimental study, Lewis (1994) found that only eight evaluators are sufficient to detect 95% of the problems with an individual detection rate of 0.45. In this context, task analysis is an easy-to-use method for evaluating the user and software performance, observing how the interactions are related to the relevant tasks and prioritizing possible functionalities.

Logging users' interactions is another effective usability evaluation method. The strength of this method comes from the fact that it can be applied to a large number of actual users (although analyzing them may take some time). In this sense, through logging, it is possible to find usability issues which cannot be revealed through observation. Moreover, use logs are valuable sources especially when combined with task observations and other collected data.
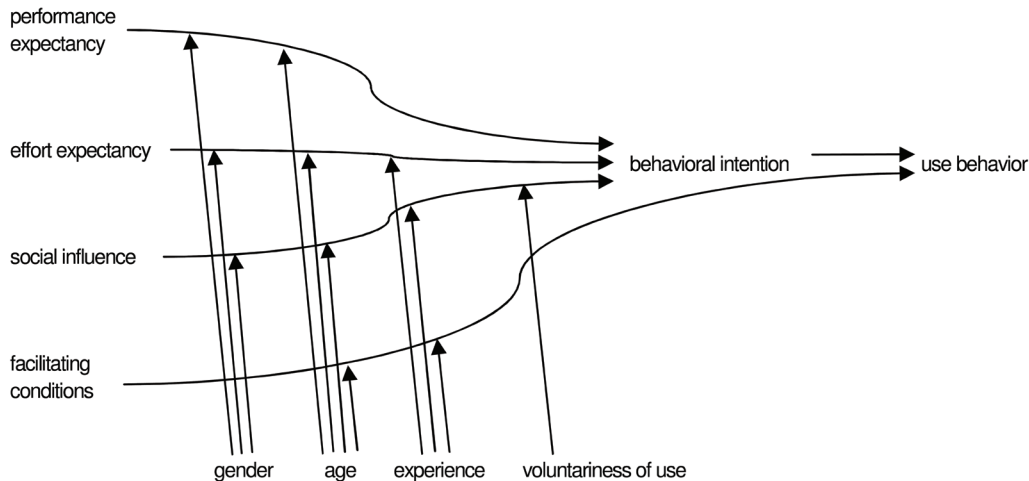

Figure 3
Unified Theory of Acceptance and Use of Technology (Venkatesh et al, 2003) based on (Davis et al., 1989).

Questionnaires and surveys can be utilized for various purposes related to usability (ISO/IEC 25062, 2006). A common practice is the user satisfaction assessment. In the last thirty years, numerous user satisfaction questionnaires are developed and tested by established researchers. Most of these questionnaires are well-documented and publicly available. Therefore, conducting such studies is not so difficult. In addition, questionnaires are efficient tools for collecting information on user characteristics; which is essential for profiling the users and determining the possible influence factors. Furthermore, questionnaires can be conducted online, saving plenty of resources and making this method even more cost-effective.

Interviewing is another beneficial method which is often used as a follow-up measurement tool in combination with other methods (Shuy et al., 2001). When performed rigorously, interviews are useful for collecting information on users' experiences and ideas. In particular, follow-up interviews are highly complementary with task observation and questionnaire methods. However, this method can be very time-consuming especially when conducted with a high number of users.

Focus group studies are moderated roundtable discussions that are conducted with carefully selected participants to collect information on their ideas and experiences. They are useful for obtaining various perspectives on use case scenarios, functionalities and the design of the interfaces. In this sense, focus group discussions are valuable in the exploratory stages of CAAD software development; but leading the group in an efficient way and keeping the discussion on track or "focused" is a challenging task. Through these methods, it is possible to obtain high quality feedback in a limited period of time.

Table 2
Opportunities and challenges of usability evaluation methods; based on (Nielsen and Mack, 1994), (ISO/IEC 25062, 2006), (Shneiderman and Plaisant, 2005), Usability Body of Knowledge [1].

Method	Opportunities	Challenges
Task Observation	Evaluates the extent to which the software facilitates users to accomplish their tasks	Sensitive to user profiles, requires careful setup and task definition
Logging use	Provides a valuable data record of the real use and interfaces; powerful when combined with the content	Possible decrease in efficiency and issues related to privacy may arise in certain cases
Questionnaires	Measures users' attitude towards the software	Bias should be minimized in the design
Interviews	Powerful as a follow-up tool; valuable for understanding user experiences	Provides incomparable information; possible interference with the outcomes
Focus groups	Obtaining various perspectives; valuable in the exploratory stages	Requires efficient moderation; hard to keep a focus
Ethnography	Allows long-term information collection in a natural setting	Time and resource consuming; high observer bias involved
Cultural Probes	Reveals cultural aspects and values	Based on self reports; difficult to interpret
Think-aloud method	Reveals users' thinking processes, misconceptions can be identified, provides comparable data	Negative effects on efficiency and performance; takes long time to analyze
Eye Tracking	Provides objective, accurate, visual and time-based data on interface use	It is difficult to infer causality between the collected data and usability problems
Model Based Evaluation (KLM/ GOMS)	Can be used in the early development phase, formative and summative evaluation	Based on an expert user model, not real users; predicts mostly quantitative measures
Inspection: Heuristic Evaluation and Walkthroughs	When applied prior to other methods, it can reduce the number of user errors	Fails to include real users; relies on expert knowledge; can possibly hinder innovation and creativity

In contrast, we can reference various other methods that are valuable but not so cost-effective in early CAAD software prototype development. Among those are *ethnography and cultural probes* which require long-term commitment from the users (Gaver et al., 1999). Overall, these two methods give high quality results but these are only suggested when the CAAD software developers have enough time and resources to execute them.

Similarly, *think-aloud* is a reliable but not so cost-effective research method. It can provide critical insight into the users' thinking processes (Ericsson and Simon, 1993) and help evaluators to identify misconceptions.

On the other hand, it takes a lot of effort and time to make a pilot study, design the experiment and build a coding scheme, conduct the real experiment, transcribe, segment and codify the verbalizations and perform statistical analysis.

Eye tracking is another valuable usability evaluation method that provides accurate and time-based data on interface use (Nielsen and Pernice, 2010). However, it is difficult to infer causality between the data collected from the experiments and usability problems. Furthermore, sophisticated equipment, software, and training are required to conduct eye tracking experiments; which makes this method expensive.

In the future, eye tracking is expected to become more affordable and accessible as eye tracking equipment and technologies are developed further which may increase the cost effectiveness of this method.

Goals, Operators, Methods, and Selection rules (GOMS) and Keystroke Level Modeling (KLM) are predictive methods based on the human information processor model for human computer interaction observation (Card et al., 1983). Relying on expert user models (not real users), these methods estimate metrics such as the time required to learn the system use and execute specific tasks. In this context, they can be highly usable for developing interfaces for users (such as pilots) who meet certain physical and mental requirements and follow guidelines to execute defined tasks. In contrast, the profiles of architectural designers are highly heterogeneous which makes their performance difficult to predict.

As a final topic to be reviewed, *Heuristic Evaluation* is a method that heavily relies on "inspector" knowledge. This is a questionable approach to the evaluation of CAAD software prototypes because it insulates the users from the development process and replaces them with highly normative heuristic guidelines which are open to discussion.

In this sense, use of heuristic guidelines and creative design may be contradictory because they limit the potentials and possibilities (Burmester and Machate, 2003). Another well-known negative aspect of this method is that it is costly to employ usability experts.

In conclusion, each usability evaluation method has different flaws and limitations and they detect different usability problems. The best practice is to combine various evaluation methodologies to evaluate usability (ISO/IEC 25062, 2006).

Task observation, questionnaires, interviews, focus groups and logging use are efficient, effective and reliable; thus satisfactory methods for the early development of CAAD software, even at the concept generation stage (Pak and Verbeke, 2011). These methods can be easily customized to fit into the problem area, depending on the characteristics of the software environment.

CASE STUDY: FOCUS GROUPS, QUESTIONNAIRES AND USE LOGGING AS USABILITY EVALUATION METHODS IN THE CAAD CONTEXT

In this section, we will briefly discuss the effectiveness and reliability of various usability methods based on the usability studies that we have recently conducted for the evaluation of a web-based geographic virtual environment prototype; primarily developed in the framework of a long-term research project (Pak and Verbeke, 2011).

For evaluating the usability of this virtual environment prototype, we arranged focus group meetings, employed questionnaires for collecting user characteristics, determining user satisfaction levels and attitudes towards the system (task observation studies are in progress). Moreover, we logged the users' activities for further assessment.

Focus group meetings have significantly contributed to the early design development of our prototype. During the two meetings conducted with two different groups of experts, we have gained a clear insight into their expectations. Besides being established experts, the participants were also possible future users of the virtual environment that we are developing. In this sense, we were able to conceptualize numerous novel and critical ideas about possible improvements and new features. We believe that the characteristics of the participants played a positive role in these meetings. Moreover, these meetings were also useful for promotion and social networking.

In addition to the focus group meetings, we employed two types of questionnaires for usability evaluation: user attitude and user satisfaction.

The user attitude questionnaire included eleven Likert scale questions related to the goals of our study; an open-ended question for comments and three questions related to the computer and language skills of the students. The user satisfaction questionnaire was a standard after scenario questionnaire (ASQ) developed by Lewis (1991). Both of these were answered online by 25 students who used the web-based geographic virtual environment prototype for eight weeks in an international design studio context.

Overall, both questionnaires were valuable tools for gathering the user feedback in a formal and holistic manner. Furthermore, with the open-ended questions we were able to receive constructive criticisms.

The responses to these questionnaires effectively illustrated the (positive) attitude towards our virtual environment prototype and (high) user satisfaction levels.

In order to test the reliability of these questionnaires, we have presented the same questions to the same students four months later in print format. The comparative analysis indicates a high level of correlation between the two measurements (Figure 4). These results can be considered as suggestive evidence for the reliability of the questionnaires as evaluation tools in CAAD context.

Logging system use was an effective method for usability evaluation of the virtual environment prototype. We were able to record various user activities and compare them with the results from other measurement methods. For instance, 87% of the users in the attitude questionnaire strongly, mostly or somewhat agreed that they have actively collaborated with each other. This result was consistent with the use logs which indicated that 79% of the activities were collaborative (this study is also available in the proceedings book as an individual publication by the authors).

Moreover, we have detected strong correlations (r>.71 with 95% confidence) between user computer skills, language skills, design studio grades and satisfaction levels; which suggests that the individual characteristics of the users may influence the usability evaluation process.

A COMPREHENSIVE FRAMEWORK OF DESIGN USABILITY

Our observations and the comprehensive background study that we have reported in the previous topics illustrate that usability -especially in the CAAD field- cannot be reduced to a "software quality". Individual characteristics of the designers and the design context should also be taken into account as major influence factors. With this motivation, we are offering a novel framework for an extended understanding of usability in design (Figure 5).

In this framework, we describe design usability as a multivariate quality emerging from the interactions between the designer, CAAD software and the design context ("designer" in this sense does not necessarily need to be an architect; this definition can include all actors involved in the design process
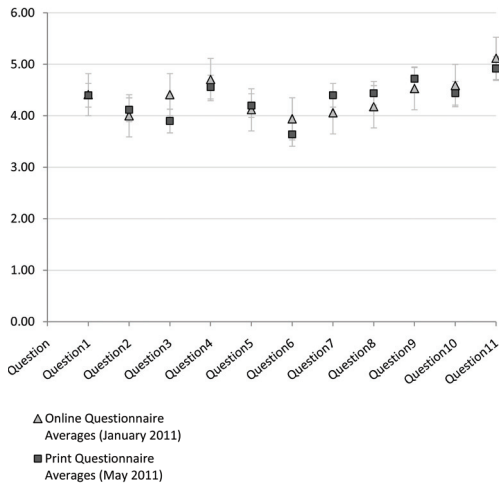
Online Questionnaire
Averages (January 2011)

Print Questionnaire
Averages (May 2011)

reflected our recent findings relating to the reliability and effectiveness of particular evaluation methods. As a result of this discussion, we introduced a new design usability framework that is suitable for CAAD software prototype development; which suggests a variety of design usability quality characteristics and possible influence factors in the evaluation process.

The framework that we have proposed in the previous section represents design usability as a multivariate quality emerging from the interactions between the designer, CAAD software and the design context. We claim that design should be a sustainable and social practice, so should the CAAD software and usability evaluation criteria be.

In this sense, the proposed framework includes two new and essential concepts, "use sustainability" and "sociability". With the rapid evolution of information and communication technologies, "use sustainability" is a critical issue to be addressed. Any software or script can ultimately become unmaintainable, inefficient, unreliable and dysfunctional in a relatively short time period (one or two years). Ensuring use sustainability is the responsibility of the developers as well as the design actors. Moreover, in the age of sociable media, "sociability" emerges as an essential quality for the future software. There are unlimited opportunities for promoting conversation and interactions between the parties involved in the design processes. Software designers can make use of social networking tools and strategies to ensure the sociability of CAAD software.

We suggest the proposed framework and suggested evaluation methods as a basis for conducting cost-effective, easy to set-up and reliable usability tests; which can improve the value of the software products by accommodating local differences and integrating human factors. In our usability study (that we have presented in the preceding section), we have observed that usability evaluation is not solely limited to interface testing, but it can also provide valuable information on the possible added values of CAAD software prototypes and increase the overall product quality.

such as the clients and engineers).

The proposed design usability characteristics include (but are not limited to): designer satisfaction, effectiveness, efficiency, freedom from errors, learnability, memorability, use sustainability and sociability.
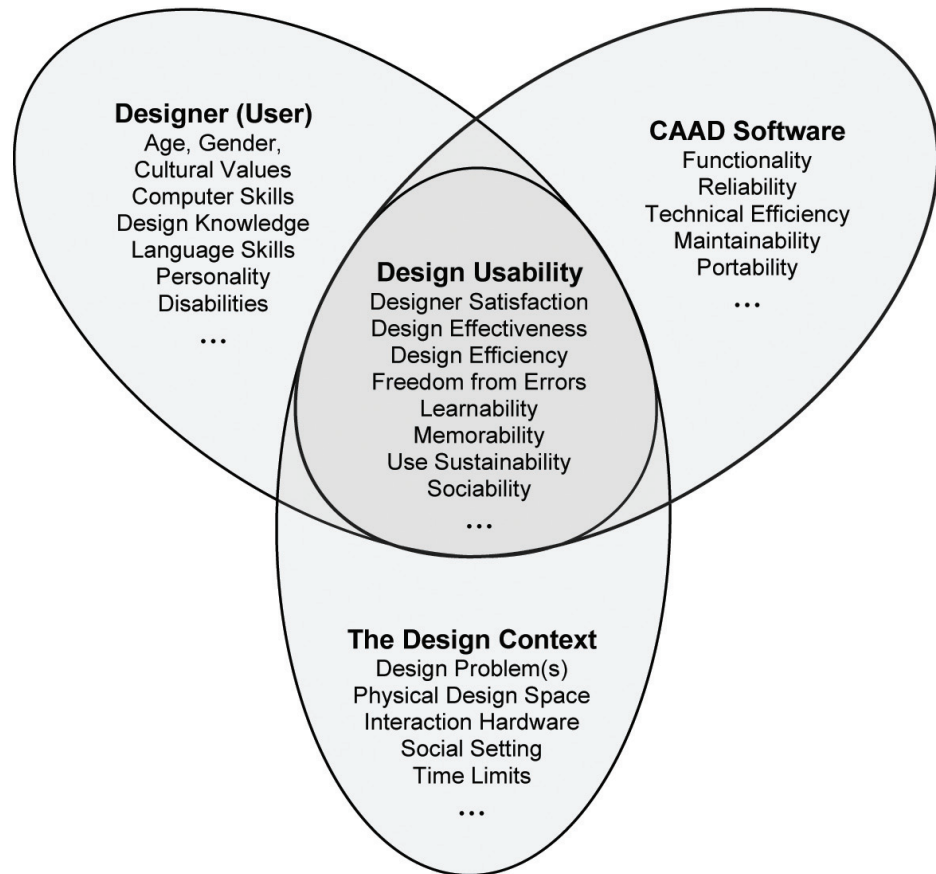
The first six characteristics are derived from the ISO Standards (2006) and the framework of system acceptability quality characteristics by Nielsen and Mack (1994); whereas "use sustainability" and "sociability" are added as possible attributes of next generation CAAD software.

Use sustainability refers to the potential of the CAAD software for the long-term maintenance of the design usability for a defined group of designers, in a specified design context. Sociability can be described as the quality of the options offered by the CAAD software to the designers that allows them to reflect their design progress through social networks and receive feedback.

CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we initiated a constructive discussion on design usability evaluation by reviewing various definitions of usability and highlighting possible opportunities and challenges of CAAD-applicable evaluation methods. Moreover, we

Designer (User)
Age, Gender,
Cultural Values
Computer Skills
Design Knowledge
Language Skills
Personality
Disabilities
...

CAAD Software
Functionality
Reliability
Technical Efficiency
Maintainability
Portability
...

Design Usability
Designer Satisfaction
Design Effectiveness
Design Efficiency
Freedom from Errors
Learnability
Memorability
Use Sustainability
Sociability
...

The Design Context
Design Problem(s)
Physical Design Space
Interaction Hardware
Social Setting
Time Limits
...

As a future recommendation -considering the complex and time consuming nature of the CAAD software prototype development and evaluation- we propose to employ open-source strategies and tools to enhance the sustainability of CAAD software. As an example, a web based medium (or the "CUMINCAD of source code") can be created for sharing and collaboratively developing software prototypes under the Creative Commons license. This environment may also be designed to provide web tools for "community-based" or "crowdsourced" usability testing; which can assist the sustainable development of the CAAD research field and enhance its sociability.

REFERENCES
Burmester, M, Machate, J 2003, 'Creative design of interactive products and use of usability guidelines–a contradiction?' in *Human-Computer Interaction: Theory and Practice (Part 1)*, Lawrence Erlbaum, New Jersey, pp. 43–48.

Bevan, N 2006, 'International Standards for Human Computer Interaction', in Claude Ghaoui (Ed.) E*ncyclopedia of Human Computer Interaction,* Idea Group Publishing, London, pp. 662–372.

Card, S K, Moran, TP, Newell A 1983, *The Psychology of Human-Computer Interaction,* Lawrence Erlbaum Associates Inc., Hillsdale, New Jersey.

Davis, F D 1989, '*Perceived usefulness, perceived ease of use, and user acceptance of information technology'*, MIS Quarterly, 13(3), pp. 319–340

Ericsson, KA, Simon, HA 1993, *Protocol analysis;* Verbal reports as data, Bradford books/MIT Press, Cambridge, Massachusetts.

Gaver, B, Dunne, T, Pacenti, E 1999, 'Design: Cultural probes', *Interactions Journal,* vol. 6, no.1, pp. 21–29.

ISO/IEC 25062 2006, *Software engineering: Software product Quality Requirements and Evaluation Common Industry Format for usability test reports,* ISO/IEC, Geneva.

Lewis, JR 1991, 'Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ', *SIGCHI Bulletin* no. 23, pp. 78–81.

Lewis, JR 1994, 'Sample sizes for usability studies: Additional considerations', *The Journal of the Human Factors and Ergonomics Society,* vol. 36, no. 2, pp. 368–378.

Maver, T W 1995, 'CAAD's Seven Deadly Sins', *Sixth International Conference on Computer-Aided Architectural Design Futures*, Singapore.

Nielsen, J, Mack, R L 1994, *Usability Inspection Methods*, Wiley & Sons, New York.

Nielsen, J, and Pernice, K 2010, *Eye tracking Web Usability,* New Riders, Berkeley, CA.

Norman, D 2004, *Emotional Design*, Basic Books, New York.

Pak, B, Verbeke, J 2011, *A Virtual Environment Model for Analysis and Evaluation of Alternative Urban Development Projects for the Brussels-Capital Region (Project Report),* INNOV-IRIS: Institute for Research and Innovation, Brussels.

Pak, B, Verbeke, J 2010, 'A Virtual Environment Model for Analysis and Evaluation of Future Development Projects for a European Capital', *Education and Research in Computer Aided Architectural Design in Europe (eCAADe) Conference*, ETH (Eidgenössische Technische Hochschule), Zürich.

Shneiderman, B, Plaisant, C 2005, *Designing the User Interface,* Boston, Pearson.

Shuy, RW 2001, 'In-person versus phone interviewing', in J F Bugrium, J A Holstein (Eds.), *Handbook of interview research: Context and method,* Sage Publications, Thousand Oaks, California, pp. 537–555.

Stellingwerff, M, Verbeke, J 2001, *{ACCOLADE} Architecture, Collaboration, Design Euro-Workshop,* Delft University Press Science, Delft.

Venkatesh, V, Morris, MG, Davis, FD, Davis, GB 2003, 'User acceptance of information technology: Toward a unified view', *MIS Quarterly*, vol. 227, no.3, pp. 425–478.

 [1] tinyurl.com/usabilitybodyofknowledge.