Usability evaluation of a prototype design tool for uncertainty propagation and sensitivity analysis

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

USABILITY EVALUATION OF A PROTOTYPE DESIGN TOOL FOR UNCERTAINTY PROPAGATION AND SENSITIVITY ANALYSIS

Christian Struck[1, 2], Jan Hensen[2] and Wim Plokker[3]

[1] Lucerne University of Applied Sciences and Arts, Switzerland
[2] Department of Architecture Building and Planning, Technische Universiteit Eindhoven, The Netherlands
[3] Vabi Software, Delft, The Netherlands

ABSTRACT

Software developments in the domain of building performance simulation (BPS) targeting the early design stages of a building need to address two points successfully to be adopted in design practice: (1) facilitating communication between multiple engineering disciplines and (2) the limited amount of design information. The authors consider the limited amount of design information available not as a limit but as a design uncertainty. To focus the designer's attention the approach chosen here is to extend the capabilities of existing simulation tools with uncertainty and sensitivity analysis.

The development of software goes as any product development through stages as design, synthesis and analysis and involves numerous design iterations. The analysis of the prototypical tool extension includes verification und usability evaluation.

Whilst the verification of prototypical design tools is necessary to ensure the added analysis functionality is implemented correctly the usability evaluation is to ensure the proposed feature meets the demand of the potential user group.

A heuristic usability evaluation was conducted with six expert practitioners using a paper prototype. The quantitative feedback to heuristics as design guidance, process integration, and application confirmed the potential of the tool extension to support design practice. The usability evaluation indicated that expert practitioners would encourage the use of uncertainty propagation and sensitivity analysis if tool extensions to BPS-tools were available. The experts assess uncertainty propagation and sensitivity analysis to add value by reducing the risk of technical design decisions and limiting the extent of design iterations.

USABILITY ENGINEERING

The use of building performance simulation (BPS) tools during the early design stages is limited. The author's hypothize that the use of uncertainty propagation and sensitivity analysis techniques allow to focus the designer's attention to the key-design parameters with respect to their influence on important performance indicators. To test the hypothesis usability evaluation was conducted.

Usability engineering considers usability as important aspect in the process of designing software and interfaces. Usability is discussed since the 1970's and concerns improving the interaction of products and users. Efforts have been published in different fields such as the design of appliances (Sauer et al., 2010), medical interactive systems (Bastien, 2009), software-, website- and interface design, building simulation (Folmer and Bosch, 2004, Holzinger, 2005, Nielsen and Landauer, 1993, Hopfe and Hensen, 2009) among others. Definitions for usability are given by different sources as Shackel (2009), Nielsen (1993), ISO9241-11 (1998) and Holzinger (2005). Holzinger defines usability based on work of Bevan (1995) as…

…the ease of use and acceptability of as system for a particular class of users carrying out a specific task in a specific environment. Ease of use affects the user's performance and their satisfaction, while acceptability affects whether the product is used."

Different authors categorize usability engineering methods and tools differently. Quesenbery (2008) uses four categories: exploratory research, benchmark metrics, diagnostic evaluation and summative testing. Holzinger (2005) differentiates inspection methods - without end-users and test methods - with end-users. He subdivides test methods into thinking aloud, field observation and questionnaires, and inspection methods into heuristic evaluation, cognitive walkthroughs and action analysis. Folmer and Bosch (2004) identify usability testing, usability inspection and usability inquiry.

Usability testing requires representative users to work with the product on typical tasks. That can be done on either the final product or a not yet finished model of the product. Methods include thinking aloud and question asking protocols.

Usability inspection requires either experts, developers or other professionals to assess a prototype to follow established usability principles.

Typical methods are heuristic analysis and cognitive walkthroughs.

Usability inquiry is concerned with obtaining information about the likes, dislikes, needs and users understanding of a product in real operation. Information can be gathered by field observations and surveys.

Heuristic evaluation

Nielsen (1993) argues that heuristic evaluation is the most common informal method. It is based on usability specialists assessing if the presented product design follows established principles. Before conducting a usability test a number of points need to be considered:

- Type of participants, expert versus novice designer,

- Prototype fidelity, paper prototype versus fully functional product,

- Number of participants, only a few versus a large number.

Type of participants

Participants can be chosen based on a number of characteristics such as competence, attitude, state and personality. In the context of this work a competent person is a person highly skilled; knowledgeable and able in the domain of climate engineering. Literature provides statements about what type of participants is best suited under the given circumstances. In a recent study Sauer et al. (2010) conclude that when aiming at gaining an overview of usability problems experts provide a more comprehensive list. If the aim was to identify the most severe usability problems as quickly as possible novice users perform better. Severe usability problems are considered problems that would prevent the completion of a specific task. It was found that expert would point to problems relating to efficiency and functionality as they were able to adopt compensating strategies experienced problems. Further to being able to identify to actual problems experts were able to point to potential usability problems.

Prototype fidelity

The prototype fidelity depends on the stage of the development process. Whilst in the late stages fully functional prototypes might be available, which is not the case during the early design stages. The work by Sauer et al. (2010) suggests that subjective usability ratings were unaffected by prototype fidelity. The users seem to be able to compensate the lack of system and environmental feedback of low fidelity prototypes with their mental model of the product. That is as they use their understanding of the product to predict its performance. Low fidelity prototypes are suitable for usability testing taking into account the following three points:

- Overestimation of available user controls

- Limitation of number of measured usability metrics

- "Deficiency compensation" might lead to more positively rating the low fidelity prototype than the final product

Number of participants

Literature gives no clear answer to the question how many participants to involve. Nielsen and Landauer (1993) propose a model to predict the eventual number of problems that will be found by a usability study. Holzinger (2005) states that inspection methods require 1-5 participants, whilst test methods require 4 till well over-30 participants. Usability can be considered from at least five perspectives as: learnability, efficiency, memorability, error rate, and satisfaction.

1. Learnability is important so that the user can rapidly begin working with the system product.

2. Memorability considers allowing the occasional user to return to the product without having to relearn its use.

3. Error rate focuses on eliminating the risk of making severe mistakes, and minimizing the general error rate while using the product.

4. Satisfaction aims at making the system pleasant to use.

5. Efficiency concerns enabling the user who has learned the product to attain a high level of productivity. The latter is considered here.

Only a few publications relate usability engineering to process stages of designing software. Folmer and Bosch (2004) conclude their comprehensive review that none of the usability engineering methods as testing, inspection or inquiry have the capacity to support software architecture. However, the authors consider software architecture as part of the software development process, whereas it would be advantageous, applying the systems theory perspective and decomposing the process into components. By doing so it is possible to apply common process stages to engineering software architecture, and considering the use of traditional usability engineering techniques.

Other practitioner's guides as by Dumas and Redish (1999) suggest applying usability engineering techniques throughout the design process from conceptual design to detailed design. Dumas and Redish (1999) suggest:

- Focusing early and continuously on the user;

- Considering all aspects of usability;

- Testing versions of the product with users early and continuously;

- Iterating the design according to the feedback.

For the early design stages of software developments it suggests the following:

1. Make use of low fidelity prototypes as it allows the rapid turnaround of design iterations.

2. Focus on key usability aspects/ heuristics of the development, which is here providing means to enhance the efficiency of the building design process by providing design guidance.

3. Target domain experts as they have the ability to give feedback on the potential of the prototype to increase the process efficiency by adding value and new functionality.

Heuristic evaluation is a widely used informal usability engineering method (Holzinger, 2005). It requires the careful selection of usability heuristics. Different sets of usability heuristics have been applied in different studies. Nielsen (1994) compares seven sets used for high fidelity prototypes and identifies nine heuristics: (1) visibility of system status, (2) match between system and real world, (3) user control and freedom, (4) consistency and standards, (5) error prevention, (6) recognition rather than recall, (7) flexibility and efficiency of use, (8) aesthetic and minimalist design and (9) helping users recognise, diagnose and recover form errors.

METHODOLOGY

Prototype design tool and case study

To test the hypothesis a prototypical and case specific design tool was developed which integrates VA114 (Vabi Software BV, 2010) , an integrated BPS-tool and Simlab3.2.5 (EU Joint Research Centre, 2009) a statistical analysis tool in the Matlab environment.

Uncertainty analysis is expected to support quantifying the performance variation of a design concept relative to a design requirement. Using statistical analysis techniques the sensitivity of performance requirements to input parameters can be established. The knowledge about sensitivities is expected to support communicating the building performance between project stakeholders. For the analysis of the case study Latin hypercube sampling and regression analysis was chosen.

The case study comprises of a representative office building. The office space considered for the comfort assessment is visualized in Figure 1.
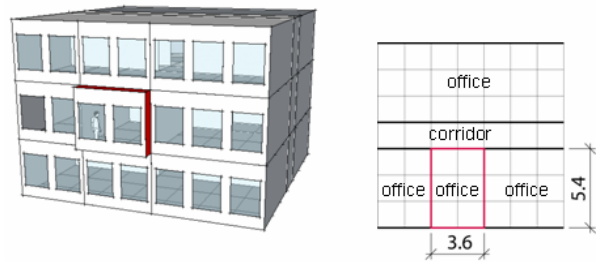

Figure 1 Case study - office room

Two design alternatives (DA1&2) were investigated for finding the most appropriate compromise for the conflicting design requirements: (1) good thermal comfort at workplace level and (2) low final energy demand for heating and cooling at building level.

DA 1 represents a space conditioning concept integrating floor heating and cooling, mechanical supply & extract of the minimum fresh air rate, radiator heating and a 1m deep overhang at the highly insulated façade.

DA 2 is a space conditioning concept based on hybrid ventilation scheme; natural air supply with mechanical extract, a 4-pipe fan coil unit for heating and cooling, internal blinds and moderate insulated façade.

The uncertainty of final energy use und comfort for both DA's to variations of parameters as internal gains, g-value, glass to wall ratio, u-value wall, ventilation air volume, infiltration rate, u-value window ratio, acoustic to thermally active ceiling area and orientation was assessed (see Figures 2&3).
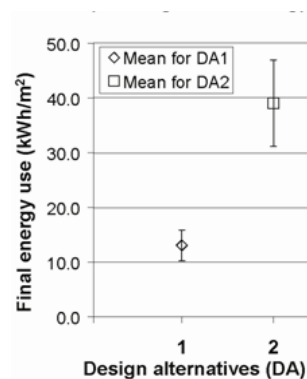
Figure 2 Final building energy use (heating & cooling) for two design alternatives
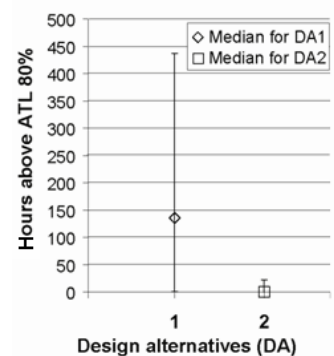
Figure 3 Number of hours ATL 80% for two design alternatives

DA1 shows the lowest final energy use and the potential to also provide very low hours above the adaptive temperature limit above 80%. This is the starting point for the sensitivity analysis. The question is: "Which parameters have the biggest impact on the number of hours above the adaptive

temperature limit (ATL) of 80% for DA1?". Using regression analysis it could be identified that the three most important parameters are: internal gains, g-value, glass to wall ratio. The three least important parameters are u-value window, ratio acoustic to thermally active ceiling area and orientation (see Figure 4).
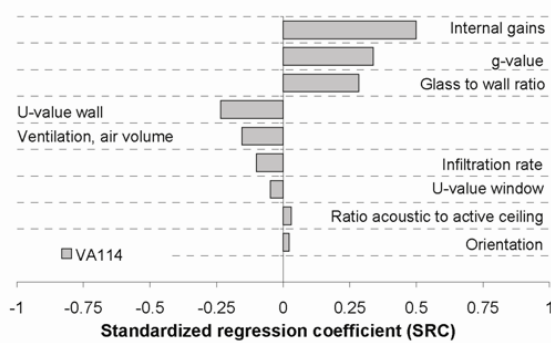

Figure 4 Parameter sensitivity to number of hours above the adaptive temperature limit of 80% by standardized regression coefficient.

Heuristic evaluation of a prototype design tool

The knowledge gained by developing the prototype design tool and conducting the analysis of the case study represented the input to produce the prototype for the usability evaluation. The effort resulted in a low-fidelity paper-prototype. The paper prototypes visualize the case based analysis workflow using a BPS-tool expanded with the capability for uncertainty propagation and sensitivity analysis. Attention was paid to the definition of design problems, data input and presentation of analysis data.

Three evaluation heuristics are particularly important: (1) provision of design support; (2) design process integration and (3) perceived applicability.

The paper prototype was exposed to six expert practitioners in the field of environmental engineering in form of a PowerPoint presentation.

The usability of the final prototype was evaluated in individual sessions with expert practitioners in August 2009. The experts, exposed to paper prototypes, were engaged in a discussion to learn about their perception. The individual usability problems were noted and categorized. The notes were fed back to the experts to confirm and were necessary to complement the list of recorded usability problems. To objectively assess the value of the final prototype for supporting the design process it was necessary to quantitatively evaluate the cumulative response of the expert practitioner's. This was achieved by scaling qualitative feedback. The Likert-scale (Hinkin, 1995) was applied, as it allows

scaling the expert response to items as closed questions and statements.

statements, closed- and open ended questions. Three questions were assigned to the identified heuristics. Mean scores and variances were computed for the response scores.

The audio track of the evaluation meetings was recorded to supplement the analysis. First the practitioners were exposed to the theoretical background of the prototype. Thereafter, the prototypes capability providing support was illustrated for solving a realistic design problem for a virtual building. Two design alternatives were evaluated. Subsequently the experts were asked to score the response to key-questions making use of the Likert-scale. The questions were answered un-anonymously.

RESULTS AND DISCUSSION

Qualitative Results

The feedback to the evaluation was exhaustive. The following points were identified motivating the use of building performance simulation to support conceptual design:

- The aim during building design is to match design concepts to user requirements expressed in form of performance indicators (limits). BPS allows quantifying conflicting performance indicators as, e.g. are energy use and comfort.

- Climate engineers have limited influence on the architectural design. Tools that provide useable interfaces, advanced analysis techniques as uncertainty propagation and sensitivity analysis and intelligent means to present results are expected to significantly improve the communication within the design team, especially between architects and climate engineers.

- During design BPS has great potential to shape the expectations of the design team with regards to building performance. Realistic expectations, based on quantitative performance information, help reducing costly design iterations.

- The use of performance simulation provides the opportunity to break away from peak performance design moving towards designing for typical performance, e.g. during midseason. When aiming at reducing energy consumption for heating and cooling the integrated system efficiency shows the biggest saving potential.

The following points relating to usability were identified:

1. Application of the prototype (Do what the user wants)

Building performance simulation is used when the effort justifies the results by supporting finding a better design solution and to communicate concept advantages and disadvantages.

The prototype is required to enable an early design robustness assessment to parameters as user behavior and climate variations. The robustness of the system has consequences on plant room- and riser sizes.

2. Variable analysis focus (Do what the user wants)

Support the shift in analysis focus away from peak load towards typical performance design.

Variable analysis zoom-in level are required to be able to quickly change focus between workspace, zone and/or building level depended on the considered performance indicator.

3. Knowledge accessibility (Interoperability)

Access to knowledge based limits and restrictions for performance indicators and design variables to support definition of parameter input ranges. Decision making and knowledge management systems are expected to be advantageous.

4. Representation of analysis input (Speak the language)

The input to analysis tools is ideally of the same type and format as used by practitioners to synthesize integrated design concepts, such as subsystems, components and variables. System theory is expected to provide the required theoretical framework. The dominant type is determined by the design stage, e.g. concept or detailed, and type of project, e.g. new build or refurbishment.

5. Extend of component models and parameters (Quality assurance/ error management)

The minimum requirement for the prototype is to allow synthesizing alternative integrated design concepts from components as: façade, lighting, energy generation and distribution, shading devices, structure and ventilation. The expert's opinions diverge on the required extend of the application. Whilst some statements request extension to, e.g. model transportation systems in buildings, others advise to withdraw from moving towards all integrated simulation models. The following components are assessed particularly important: suspended ceiling and raised floor constructions, displacement ventilation and day lighting systems Infiltration rate.

6. Modeling detail during design (Efficiency of use)

The level of modeling detail should be consistent for models targeting early und late design stages. Still the requirements of the different design stages should be maintained. During concept design the interest lies in evaluating multiple feasible system combinations. Therefore representative system settings need to be available. During detailed design the aim is to manipulate the system parameter to optimize its integrated performance.

7. Design parameter ranges (Quality assurance/ error handling)

To support the conceptual design, extensions to BPS-tools should be able to allow an impact assessment of design parameters within realistic ranges as: ratio net to gross floor area, impact of HVAC systems on aesthetic, HVAC control settings, occupancy pattern, properties of glass, lighting control, convection factor of blinds.

8. Performance indicators

It was pointed out that the weight of performance metrics in design decisions is not constant during design process. It changes during the design and construction process. Examples are: "Costs are not important as long the budget is not exceeded.", or "Variations in daylight levels are accepted as long it is not dark.".

To communicate costs and performance the following metrics are considered: running-, investment-, life cycle- and maintenance costs.

Improving thermal comfort as it relates to productivity. Thermal comfort is particularly crucial in mid-season.

Three metrics are in use to communicate thermal comfort adaptive temperature limits, overheating hours and weighted overheating hours.

The value of CO_2 emissions to communicate building performance is controversial among the practitioners. Whilst CO_2 emissions might be a performance requirement for public clients, private clients tend to be less considerate.

Practitioner's asses cost saving potential the highest by improving space use and considerate investments. The saving potential is assessed small for reducing energy consumption. When targeting energy efficiency the potential lies in the midseason.

It was suggested including productivity to cost function – extra annual income due to enhanced indoor air quality equals annual depreciation of building and components.

9. Analysis process

It was explained that it would help decision making to screen many input (design) parameters early.

Building design doesn't typically follow prescriptive stages. Design decisions are taken on specific subjects. Those subjects might or might not fall into the timeline of traditional design process descriptions.

To take advantage of uncertainty propagation and sensitivity analysis during the early design stages practitioners require a two phased approach. The first phase, an initial qualitative concept comparison followed by quantitative parameter impact assessment as second phase.

10. Analysis methods specific result representation

Experts assess knowledge of parameter non-linearity of less valuable. More crucial is knowledge about parameter impact (sensitivity).

Result presentations should indicate relationships for easily deriving deign information. "If an engineer cannot explain analysis results, it will be rejected by the design team.

Quantitative results

The feedback from practitioners to the quantitative heuristic prototype evaluation showed different characteristics.
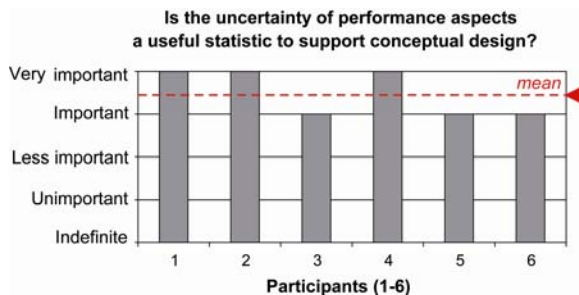

Figure 5 Design support - Usefulness of parameter uncertainty as statistic to support design

It varied over the full range of the scale in some instances, whilst being tightly scattered around the mean of the scale in others (see figures 5 & 6).
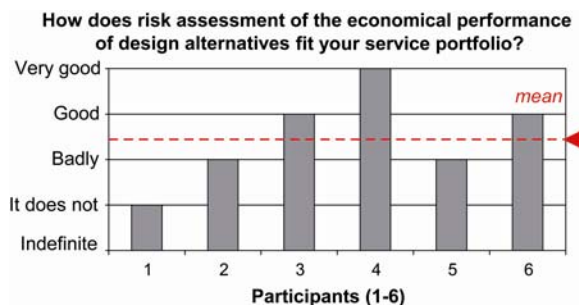

Figure 6 Application – Risk assessment of economical performance of design alternatives

The results from the heuristic specific questions are presented in tables 1-3. The mean score and variance across the expert's feedback is presented. The mean score gives the perception of design experts. The variance is used as metric to evaluate the agreement or disagreement between the experts. Whilst a low variance (<0.5) indicates good agreement between the experts on the mean score, a high variance (>1.0) indicates disagreement.

Table 1 Design support – Quantitative results

DESIGN SUPPORT	MEAN SCORE/ MAX. SCORE	VARIANCE OF SCORES
Is the uncertainty of performance aspects a useful statistic to support conceptual design?	3,5/4	0,3
Has uncertainty propagation and sensitivity analysis (UP/SA) prototype the potential to add value to the design process by generating extra design information?	4/4	0,0
Would you benefit from applying uncertainty propagation and sensitivity analysis to your current projects?	4/4	0,0

The expert's scores to questions targeting the design support are in all three instances very high. The variance is negligible small.

Table 2 Process integration – Quantitative results

PROCESS INTEGRATION	MEAN SCORE/ MAX. SCORE	VARIANCE OF SCORES
How do you assess the potential of the prototype for UP/SA to reduce time turning over simulation projects?	3,2/5	1,4
How do you assess the potential reducing design iterations using UP/SA?	4/5	0,4
Is the UP/SA analysis process transparent enough to be able to communicate its advantages and disadvantages to the design team?	2,7/4	0,2

The scores for process integration were medium to high. The highest variance is noticeable for the potential of UP/SA to reduce time turning over simulation projects indicating different opinions among the experts. The analysis process transparency received medium scores with low variance as the experts stated that there is no need to communicate details of the analysis process to the design team.

Table 3 Application – Quantitative results

APPLICATION	MEAN SCORE/ MAX. SCORE	VARIANCE OF SCORES
How does risk assessment of the economical performance of design alternatives fit your service portfolio?	2,5/4	1,1
How does the risk assessment of technical design decisions fit your service portfolio	3,75/4	0,175
Which project stakeholder will benefit most from employing BPS extended with UP/SA?	Design team *	/

* Discrete choice: Other possibilities were client, occupant, climate engineer or others.

The expert's scores to the risk assessment of economic performance of design alternatives are medium with a high variance indicating different opinions. However, there is a high score with low variance noticeable for the risk assessment of technical design decisions. The experts agree that the design team will benefit most from the capability of uncertainty propagation and sensitivity analysis.

Whilst the low-fidelity prototype limited the evaluation heuristics to the assessment of the potential for design support, design process integration and application later higher-fidelity might allow addressing a broader range of heuristics as indicated in section "Heuristic evaluation".

The use of the 5 point Likert-scale works to obtain quantitative results from the heuristic evaluation.

CONCLUSIONS

The application of heuristic usability evaluation is feasible and useful and its fits the character of the tools early design phase. The literature review revealed that there are only a few publications available on the subject of evaluating the usability of BPS-tools on the potential user group. The application of heuristic usability evaluation was found suitable for a low-fidelity prototype design tool. The Feedback to key-heuristics indicates

strength and weaknesses of the concept of the prototype design tool. Strengths are:

- Uncertainty is assessed a useful statistic to provide design support.

- Experts agree that UA/SA have potential to add value to the design process by generating extra design information.

- Experts would use tools for UA/SA on current projects if they were available.

- Strong focus on risk assessment of technical design decisions rather than on economical performance of design alternatives.

- Experts assess the entire design team will benefit from the analysis rather than individual stakeholders.

Potential weaknesses are:

- Perceived effort to turn over simulation projects.

- Transparency of analysis process.

This first usability evaluation of a low-fidelity prototype design tool was concluded successfully. As the evaluation was not conducted on a final product it does not conclude the process. Further rounds of user evaluations have to be conducted expanding on both the considered target group and number of evaluation heuristics as the development progresses. That is to ensure that the user demands on the final product are met.

REFERENCES

BASTIEN, J. M. C. (2009) Usability testing: a review of some methodological and technical aspects of the method. *International Journal of Medical Informatics,* In Press, Corrected Proof.

BEVAN, N. (1995) Measuring usability as quality of use. *Software Quality Journal,* 4**,** 115-130.

DUMAS, J. S. & REDISH, J. C. (1999) *A Practical Guide To Usability Testing* Exeter, Intellect Books.

EU JOINT RESEARCH CENTRE (2009) SimLab3.2.5.

FOLMER, E. & BOSCH, J. (2004) Architecting for usability: a survey. *Journal of Systems and Software,* 70**,** 61-78.

HINKIN, T. R. (1995) A Review of Scale Development Practices in the Study of Organisations. *Journal of Management,* 21**,** 967-988.

HOLZINGER, A. (2005) Usability engineering methods for software developers. *Commun. ACM,* 48**,** 71-74.

HOPFE, C. J. & HENSEN, J. (2009) Experiences testing enhanced building performance

simulation prototypes on potential user group. *11th IBPSA Building Simulation Conference.* Glasgow, UK, International Building Performance Simulation Association.

ISO9241-11 (1998) Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability.

NIELSEN, J. (1993) *Usability engineering,* Chestnut Hill, MA, Academic Press.

NIELSEN, J. (1994) Enhancing the explanatory power of usability heuristics. *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence.* Boston, Massachusetts, United States, ACM.

NIELSEN, J. & LANDAUER, T. (1993) A mathematical model of the finding of usability problems. IN ASHLUND, S., MULLET, K., HENDERSON, A., HOLLNAGEL, E. & WHITE, T. (Eds.) *INTERCHI 93: Conference on Human Factors in Computing Systems.* Amstedam, The Netherlands, Association for Computing Machinery (ACM), New York (USA).

QUESENBERY, W. (2008) Choosing the Right Usability Technique: Getting the Answers You Need. *User Friendly 2008 - Innovation for Asia.* Shenzhen, China.

SAUER, J., SEIBEL, K. & RÜTTINGER, B. (2010) The influence of user expertise and prototype fidelity in usability tests. *Applied Ergonomics,* 41**,** 130-140.

SHACKEL, B. (2009) Usability - Context, framework, definition, design and evaluation. *Interacting with Computers,* 21**,** 339-346.

VABI SOFTWARE BV (2010) VA114. Delft, Vabi Software BV.