JOURNAL OF BUILDING PERFORMANCE SIMULATION
https://doi.org/10.1080/19401493.2024.2341089

Usability and fitness testing for building performance simulation tools

Shady Attia [a], Aly El-Degwy [b] and Maha Attia [c]

[a]Sustainable Building Design Lab, Dept. UEE, Faculty of Applied Sciences, University of Liège, Belgium; [b]Methods Chair, Faculty of Business Administration, Radboud University, the Netherlands; [c]Linux Foundation, San Francisco, CA, USA

ABSTRACT

Architects and engineers who seek design decision aid for sustainable building and city design frequently experience decisional conflict and require support across the design process. Building performance simulations have become central to supporting the design process, but little is known about the fit and usability of simulation tools and the factors influencing the implementation of simulation-based design in practice. This paper presents a novel framework for simulation tool usability called USER-FIT. The framework defines usability based on ISO definitions, a review of evidence in the literature, and our experience supporting the development of the framework. USER-FIT provides six measures for modelers and software developers to test how a simulation tool is useful, usable, and satisfactory for the intended users to inform sustainable design decision-making. Further research and testing are needed to support better the acquisition and implantation of usability testing for building performance simulation tools and applications.

ARTICLE HISTORY
Received 19 February 2023
Accepted 25 March 2024

KEYWORDS
User experience; decision support; energy modeling; design task; performance-based design; workflow

1. Introduction

BPS tools are computational decision support tools capable of generating quantitative evidence and insights supporting decision-making about possible performance-based design alternatives and solutions (Hensen and Lamberts 2012). Thus, the potential of BPS tool integration in architectural and urban design workflow is high for robust decision-making. However, the uptake of those BPS tools in practice is low. A set of conditions impedes architects and urban planners from enabling BPS tools in the design process and relying on computation to yield better decisions under conditions of uncertainty.

Building performance simulation and analysis tools (BPS) have real potential to decrease the environmental impact of buildings and cities and improve energy efficiency and indoor environmental quality through their wide adoption and meaningful use in architectural, engineering, and construction education and practice (Hensen and Lamberts 2012; Lam 2020). This is the rationale behind the American Institute of Architects (AIA), started in 2008 by the Committee on the Environment (COTE), an AIA Knowledge Community, and strengthened by the AIA Commitment to the 2030 challenge to design all new buildings, developments, and major renovations to be carbon-neutral by 2030 (AIA 2019). Similarly, the Architects' Council of Europe (ACE) has actively campaigned to open up the sustainability discourse to a more holistic approach, integrating performance simulation in the design process and architectural curricula beyond energy efficiency and across the whole lifecycle of buildings (ACE 2019; ACE 2021). However, there are huge gaps between the design practice and the potential of BPS to inform design decision-making, primarily due to cognitive, technological, and social challenges (Alsaadani and De Souza 2016; Attia, Hensen, et al. 2012; Fernandez-Antolin et al. 2020; Hopfe et al. 2017). The cognitive challenge mainly concerns usability issues, which did receive significant attention in the BPS community (Attia et al. 2013; Cetin and Mahdavi 2010; Cozza, Jusselme, and Andersen 2018; Hensen et al. 2004; Ouf, O'Brien, and Gunay 2018). Unlike many other industries (e.g. aviation, automobile, IT, consumer software, and consumer electronics), where usability is the norm in product design, the practice of usability testing in BPS software and programming interface development has been sporadic, unsystematic casual (Fernandez-Antolin et al. 2020) and shallow (Weytjens et al. 2011). The lack of systematic user experience measurement for BPS tools is partly due to insufficient attention to usability frameworks and methods in performance-based building design and project delivery workflows.

Therefore, there is a need for a BPS-specific usability framework to facilitate the adoption (Son, Lee, and Kim 2015) and meaningful use of BPS. A framework that can

CONTACT Shady Attia shady.attia@uliege.be

be used to increase efficiency and productivity, ease of use and learning, increase user retention and satisfaction, decrease human errors, decrease development time and cost, and decrease support and training costs. This research aims to extend the knowledge of usability testing for BPS tools. This paper presents the initial form of a unified framework of BPS usability called USER-FIT. The scope of the research paper is mainly building designers, BPS modelers and developers during early design stages under high uncertainty of design concepts and solutions. Specifically, the study seeks to cover the concept design, developed design and technical design stages based on the RIBA workplan (RIBA 2021). This qualitative research aims to produce and describe a framework for BPS software usability testing and user experience trials. This non-empirical article aims to build a theory and framework for user experience evaluation during BPS tools use. This conceptual article will serve as a seed for future usability tests. The paper covers usability topics and proposes procedures found in literature and ISO standards – and is limited to developing a theoretical framework without controlled study trials (case study testing). Instead, a triangulation – through the use of multiple data sources to reach convergence – of the framework and its six working steps has been performed to validate the comprehensiveness and context richness of the study.

This paper provides a valuable contribution to the new body of knowledge on usability testing of BPS used during early design stages, considering existing ISO standards. The detailed criteria, usability testing indicators, and properties describe the critical usability testing measures. This paper includes instructions to apply usability testing to evaluate and measure user experience in a structured and objective way. The suggested framework, USER-FIT, could also be used as a principle for developing BPS-specific guidelines and standards. Contrary to most usability approaches, which concentrate solely on software usability testing (Lin and Gerber 2014), we took other adoption and fitness criteria of usability testing into account during the design process. For software developers and architects, this paper sheds light on the importance of software and Graphical User Interfaces (GUI) development and criteria to be embedded during their evaluations. Defining usability testing criteria for the BPS tool can increase the BPS simulation uptake in practice and help the construction sector towards climate change-proof building designs.

2. Literature review

The International Building Performance Simulation Association (IBPSA) has developed a body of knowledge on user experience since its establishment in the 1980s (Oh and Haberl 2016). Since its establishment, there has been a proliferation of Building Performance Simulation (BPS) tools used worldwide. Until 2017, the US Department of Energy (DOE) maintained an up-to-date listing of 417 BPS tools on the Building Energy Software Tools Directory (BESTD) website, ranging from research software to commercial products with thousands of users (see Figure 1). In 2015, the list was revised to reach 260 tools. In 2018, IBPSA USA started hosting the newly revised list, reaching 214 tools in 2020.

Based on Figure 1, there is a general question about the underlying premise of the uptake and usability of BPS tools in the industry. Evidence suggests an increasing

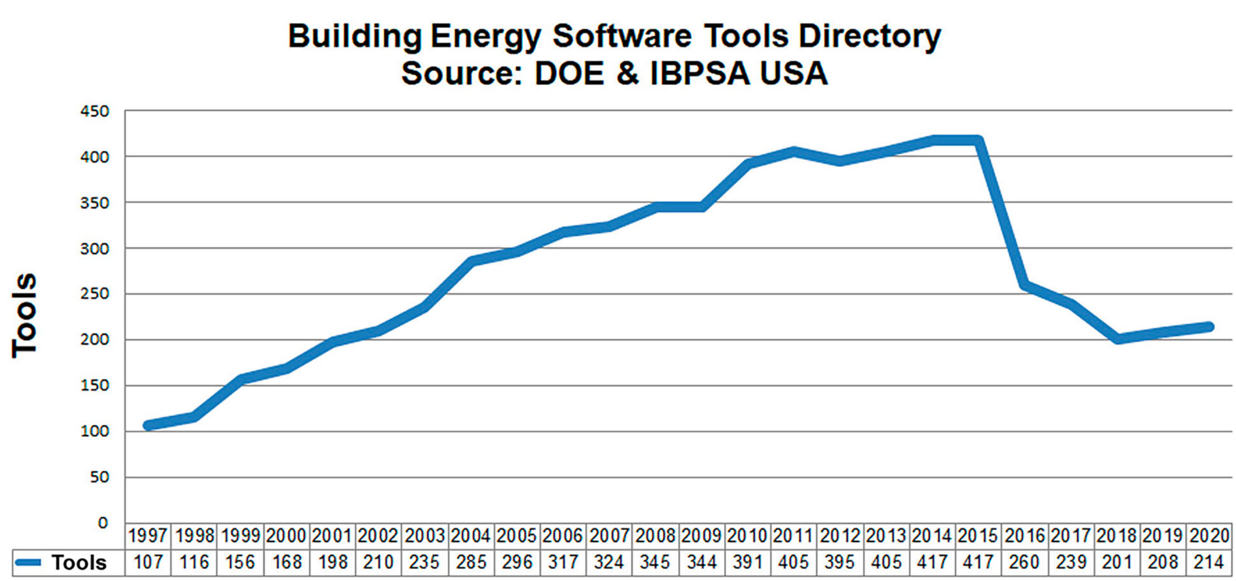

Figure 1. BPS tools developed between 1997 and 2020 (Attia, Hensen, et al. 2012; IBPSA USA 2020).

use rate and engagement of BPS among building professionals (Wang et al. 2020). The penetration of BPS tools used for energy-efficient and green building design has been documented (Han et al. 2018) in many studies (Gan et al. 2020). The industry has embraced simulation when it comes to things like rating (e.g. LEED, BREEAM, DGNB, NABERS) and ensuring designs meet Energy Performance Regulations. So it is likely that existing tools such as IES, DesignBuilder, OpenStudio, Sefaira and similar are fully adapted for compliance activities. At the same time, the use of BPS for code compliance is facing challenges to move away from a prescription of minimum physical parameters to an increasingly relevant but complex world of performance-based modeling (K. Augenbroe 2019; Negendahl 2015). Also, for compliance checking, far more sophisticated performance criteria are needed than the current practice of simplistic single-number performance targets (Donn 2023).

Since 2006, there has been a proliferation of parametric simulation tools such as Grasshopper, Ladybug, Honeybee and similar plugins and Application Programming Interfaces (API) for existing simulation engines, such as EnergyPlus and TRNSYS. The plugins and APIs are used to automatically optimize the energy (Gan et al. 2020), daylight and carbon performance of buildings. However, they are not effectively used, and their uptake remains low in informing green building design(Abdelmegid et al. 2020; Hand 1998; Papamichael and Pal 2002). Moreover, the persistent gap between predicted and actual performance, often referred to as the 'Performance Gap', raises questions about the trustworthiness and effective utilization of BPS in design decision support. There is also the issue of the modelers' experience. BPS tools require engineering knowledge and expertise. Finally, the reason for the low uptake of BPS tools beyond compliance purposes is multifactorial. The challenge of complex building design, a large number of undecided variables (Wang, Teigland, and Hollberg 2024) and wide ranges and complicated parameters affect the energy performance of buildings, many of which are out of the scope and expertise of architects and building designers (Rezaee et al. 2015). The unreliable performance prediction is related to many uncertainties, including the multi-criteria nature of the design. This would explain the sharp decline in Figure 1 and explain why we do not see many new tools emerging over the last 10–15 years.

Therefore, this study is focused on the usability of BPS tools and user experience as one of the factors that impede the building simulation in practice. There is a well-established body of knowledge about the fit and usability of building simulation tools. Several studies addressed the effective use of building performance simulation in design and how to measure usability issues and simulation software limitations of the implemented prototypes (Verheij and Augenbroe 2006). The work of Lomas on the empirical validation of building energy simulation programs is one of the earliest comparative studies to evaluate BPS under interfaces (Lomas et al. 1997). Also, Pati and Augenbroe called for integrating formalized user experience within building design models (Pati and Augenbroe 2007). Despite the significant progress in advancing building performance simulation applications and the array of tools posted at the directory of the Building Energy Software Tools Directory (IBPSA-USA 2023), it is evident that the BPS market uptake in building and city design, beyond compliance, is low (Hensen and Lamberts 2012). According to the previously identified 'Seven Deadly Sins' of software development by Maver (Maver 1995) in 1995, in computer-aided architectural design, failure to evaluate BPS tools is a hindering factor for BPS evolution and uptake. As mentioned by Clark (Clarke 2020), '*There is no independent investigation of tool ease of use and applicability to real problems*'. By investigating the user experience problem in the existing literature, we found very important studies that focused on providing credible user feedback from practitioners.

Among those studies that addressed the effective use of BPS in design with Human Computation Interaction (HCI), several studies were published. One of the earliest studies on building performance simulation usability is the work of Newton et al. (Newton, James, and Bartholomew 1988), who discussed the application of BPS from a user perspective. His work described how designers understand and use the simulation output, explore design options, interact with the software interface and explore the potential of parametric analysis. Also, Morbitzer et al. (2001) evaluated the graphical user interface of simulation tools and classified BPS tools based on the levels of ease of use. His work aimed to develop a tool that enables non-simulation experts to create a detailed simulation model and monitor the use of simulation at an early building design stage in an architectural design practice. More formally, Hopfe and Hensen (2009) measured the experience of user groups of BPS prototypes, aiming to improve the software design and observe the reactions of simulation users. Also, Mahdavi (2011) investigated the building simulation tools and environment usability. He reported that usability features in general and user interfaces in particular lag behind computational tools. Early simulation tool developers were mostly engineers and physicists, not experts in HCI (Human Computer Interaction). He confirmed the significant potential for enhancement of user experience.

A series of PhD studies also addressed the usability of BPS tools and conducted several user experience tests. For example, Peterson developed a simulation program

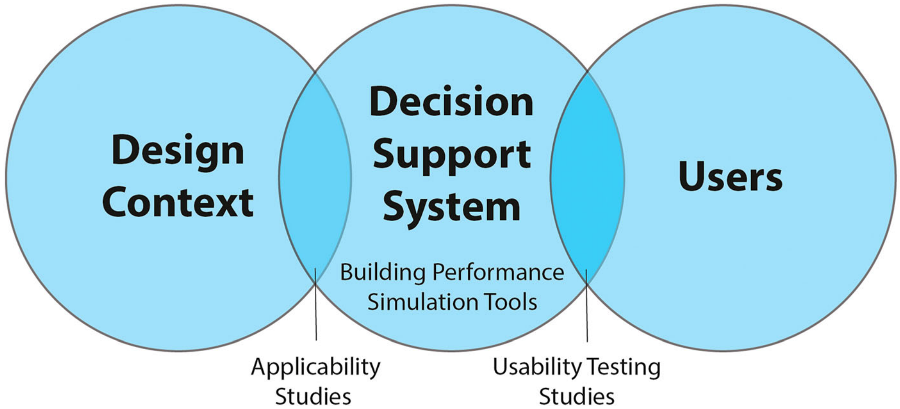

Figure 2. Decision support systems need attention to the user, design, and planning context.

for informed decisions of architects during early design (Petersen 2011). Following Peterson's program, students and researchers conducted usability testing within a master course to assess the operability of the developed tool. Attia (2012) developed a tool for designing decision-making zero-energy residential buildings in hot, humid climates. Extensive usability testing took place, investigating the robustness of the final design outcomes and reliability and using building performance simulation tools to inform architects' design decision-making during early design stages. Furthermore, Negendahl (2016) focused on integrating computational collaborative simulation models in the building design phase. Following a case studies approach, he investigated the ease of integrating BPS tools in the design process and collaboration between building designers and simulation modelers. The study targeted improving the speed of multiple and parallel performance evaluations, reducing working hours for the simulation modelers, and informing the design of high-performance buildings. MEEX (2018) developed early design support for material-related environmental impact assessment of dwellings in Flanders. Usability testing focused on the graphical user interface (GUI), data input and output, and real-time feedback. They suggested three usability testing measures based on the work of Weytjens et al. (2011), namely: (1) time of use, (2) adaptability and flexibility, and (3) comparison and feedback loop.

Other studies, like the work of Bazafkan (2017), assessed the usability and usefulness of parametric BPS tools in the architectural process. He used some usability testing criteria, including the learning curve, data entry ease and error notification features. Jayapalan-Nair et al. (2024) conducted a usability evaluation of a building simulation web-based design guide using the ISO 9241 standard. Purup (2021) developed a practice-oriented approach using building performance simulation tools to fit the design workflow in the early stages. Purup suggested involving design practice in the BPS tools development without dictating a tool-centric workflow.

According to the examined studies, user experience tests are designed and implemented differently using contradicting measures and criteria. The authors are not aware of any frameworks that systematically evaluate the usability of BPS tools. The reliance on case studies to test a tool is insufficient and ineffective in addressing workflow and cognitive discrepancies. Designing and implementing BPS tools and Application Programming Interfaces (API) without assessing the users' total experience and tools fitness (Hamilton et al. 2022) through usability testing is considered incomplete. BPS methods, tools, and front-end interfaces risk remaining instrumental and academic applications rather than truly influencing architectural design and urban planning practices. As shown in Figure 2, the critical lesson from the BPS usability debate in the last two decades is that more attention should be paid to the user and the design context rather than developing ever more advanced tools (Clarke 2020). Despite BPS-based decision support being a relatively mature field with more than 40 years (Oh and Haberl 2016) of academic and applied history, it still lacks systematic and rigorous methodologies for assessing the usability of BPS methods and tools, executing usability testing sessions, and managing the knowledge derived.

3. Methodology

This study follows a qualitative research methodology based on a three-step approach to develop and validate the literature-introduced framework. The study is designed as a conceptual article to create new knowledge by building on carefully selected sources of information combined and tested through empirical insights (Jaakkola 2020). A detailed report has been published, including the details of the research methodology (Attia and Maha 2023).

First, the systematic literature review focused on recent studies published in the last 20 years and performed usability testing during the development of BPS tools and methods for design decision support during early design. The systematic review protocol was developed with input from systematic review methodologies in the literature. The eligibility criteria and information sources are described in Appendix 1 of the study report (Attia and Maha 2023).

Second, students at Liege University were asked to answer research questions during their attendance at a building performance simulation course. The question was integrated into the course evaluation of a simulation class taught for second-year architectural engineering master's students by the first author. Architecture students were asked to rank the most common design activities and where building performance simulation can be helpful during early design stages. The survey question was answered between 2015 and 2023 and was answered by 165 respondents (See Section 4.1).

The third and final research stage involved interviewing practicing architects. Despite the non-empirical nature of conceptual articles (Jaakkola 2020), researchers in the engineering fields are not used to this type of study without conducting empirical research for validation. Therefore, in-depth interviews took place with BPS tool users (Boyce and Neale 2006) to test the developed framework. The criteria for selecting interviewees was that the person should be native Belgian, educated in a Belgian school of architecture and work in a design firm in Belgium. The interviewee had to share a case study of a building simulation exercise, of a project during early design stages, where he/she is using a BPS tool for design decision support. Similar to the approach of Purup, we only interviewed 12 architects (Purup and Petersen 2020a) in French and Flemish in their offices. The interviewees were recruited through the professional network of the first author of this paper and were conducted randomly depending on their availability. The first author, who has experience with interviews (Attia, Lioure, and Declaude 2020; Liege 2021), used a semi-structured interview approach. The interview questions were revised by the second author, who is a software engineer specializing in usability testing and can be found in Appendix II of the detailed study report (Attia and Maha 2023).

Moreover, the researchers created memo logs during interviews, allowing learning from the subjects and reflecting on several specific ideas (Connelly 2016). Memo logs were handwritten during interviews with users. They were useful for improving the quality of the developed framework (USER-FIT) and its inclusiveness. More importantly, the *memo log* was consulted several times during the barrier definition and framework validation stages. The memo logs helped define usability as how a simulation tool is useful, usable, and satisfying for the intended users to inform the decision-making to achieve sustainable building design. The framework was validated between 2021 and 2023 through interviews that aimed to uncover the point of view of interviewees gradually.

4. Performance simulation problems

In this section, we present the summary of the results of the systematic review and memo logs regarding the problems that hinder using BPS in the architectural and engineering practice during the early design stages. A detailed description of the performance simulation problems can be found in the detailed study report (Attia and Maha 2023).

4.1. Decisional needs and the user

Based on a sample of 165 early-career architects and building designers, Figure 3 shows the most common way of modeling applied as part of the design activity during early stages involving free-hand drawing, computer-aided design, and building performance simulations. The respondents were asked to indicate the design ways where BPS can play a major role in informing their design. The result confirms the variety of tasks and activities BPS can use during the early design stages. The modeling ways colored in blue in Figure 3, represent the tasks where BPD simulation was found useful for architects and building engineers including tasks of 3D form creation, massing and orientation, spatial 3D zoning and simulation of solar path. The designers identified those blue bars based on an analysis of the design iteration that take place with help with geometrical and CAD modeling in program like Rhino and SketchUp. The approach used to generate the findings in Figure 3 are explained in methodology section.

In summary, the critical lesson from the integration of building performance simulation in the architectural and planning design process debate in the last decade is that rather than developing ever more advanced tools, more attention should be paid to the user and their way of organizing form and function (Attia, Gratia, et al. 2012; Gerber and Lin 2013; Goldstein and Khan 2017; Natanian 2021; Turrin, Von Buelow, and Stouffs 2011). Designers suffer from a very low rate of BPS modeling productivity (Roth et al. 2018) and the interoperability between BIM and BEM (Li et al. 2022; Samuelson, Lantz, and Reinhart 2012). Combining this user perspective with acknowledging the design's multi-collaborative iterative process nature and the 3D geometrical representation (Dogan and Reinhart

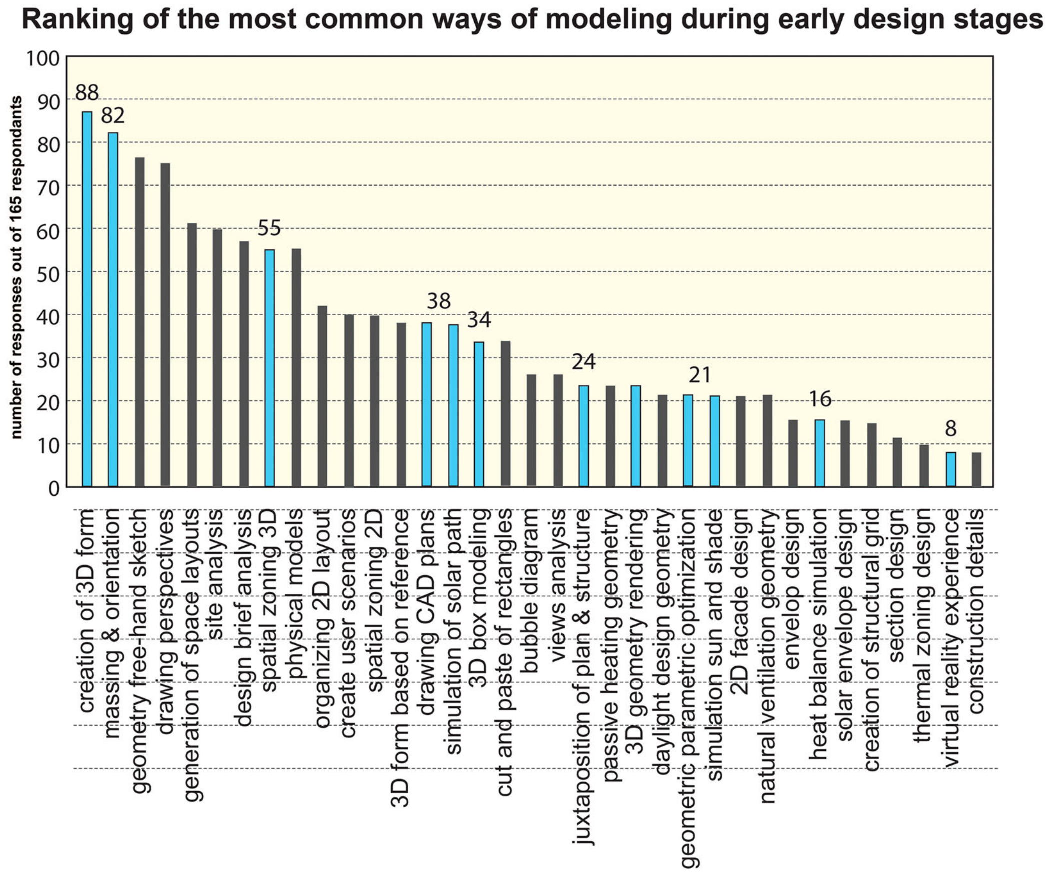

Figure 3. a sample of architects and building engineers ranking their most common way of modeling during early design stages and indicating (in blue) the potential of BPS use.

2013) logically results in a focus on seamless exchange during the decision support process.

4.2. Decision support management

The clarity of decisions and the ability to guide and communicate is another challenge that hinders the integration of performance simulation in the design process. While trial-and-error processes may be an option to integrate performance simulation in the design process, architects and urban planners usually do not consider it a decision-support option because they leads to decision fatigue. Managing the design complexity under time constraints requires timely and rich decision support regarding diversity and adaptability.

Performance simulation with embedded intelligence is another missing feature in many existing tools. By embedded intelligence, we mean the ability of the BPS tool to reflect on its operational performance by providing advice to users in the form of pre-set building templates, warning or auto-filling features for building code compliance (Attia, Hensen, et al. 2012). Comparative approaches to decision support systems are more effective in evaluating a set of well-defined competing options (Hopfe 2009). The ability of many BPS tools to compare different design alternatives, benchmark, or create reference cases cannot be found in most existing BPS tools.

The third issue with BPS decision support systems relates to visualization and the clarity and ability to understand the results and graphs (Korkmaz et al. 2010). Very few tools allow interactive visualizations such as parallel, coordinated graphs (Amer et al. 2020) or interactive Pareto front scatter plots, bar charts, or energy performance rating graphs for compliance with the EPBD or Energy Star requirements. The complexity of post-processing and BPS visualization information overload does not make it easy to understand simulation outcomes or grasp the results (Miyamoto, Allacker, and De Troyer 2022).

4.3. Decision quality

The quality of decision support of performance simulation is based on the robustness of decision-making to yield informed decisions under uncertainty. The robustness of design decision support entails meeting the desired performance and most righteous design solutions in a specific context of environment, climate and occupant behavior (Attia et al. 2013). A robust design entails nearly optimal solutions and informs the designer to narrow the performance gap. As a decision support system, BPS can enable practitioners to efficiently explore a space of design alternatives and compare them in terms of performance (Nault et al. 2018). Three major obstacles impact the design decision quality and reduce the uptake of BPS in the project delivery process.

Firstly, understanding the architectural design and urban planning context is one of the missing aspects of value-based decision-making. The critical lesson from the BPS debate in the last decade is that rather than developing new advanced tools, more attention should be paid to the design and planning context (Alsaadani and De Souza 2016; Purup and Petersen 2020a; Shiel, Tarantino, and Fischer 2018).

Secondly, the variety of key performance indicators (KPI) and the lack of structured approaches to assessing various factors' impact on building performance are reported as a burden in the building simulation literature (Attia, Hensen, et al. 2012). Architects and urban planners use indicators to perform decision analysis and scenarios-based modeling to test strategies over myriad plausible paths. Research reports the decision-making stress of designers associated with abundant performance indicators and the lack of decision trees to understand better the associations and correlations that can inform design decision-making (Li, Wang, and Hong 2021).

Thirdly, more attention should be paid to the decision quality from the user perspective of BPS tools. The decision quality requires well-prepared architects with a good knowledge of building physics (Beausoleil-Morrison 2019) and can interpret and use buildings' performance simulation results. Several studies reported the lack of knowledge (Charles and Thomas 2009) or low preparedness of many designers (Alsaadani and De Souza 2012) and modelers (Imam, Coley, and Walker 2017) to use BPS tools (Fernandez-Antolin et al. 2020).

5. Definition and stages of usability

A usability study is a research method that assesses how easy it is for participants to complete core tasks in a design (Albert and Tullis 2013). Our systematic literature review on usability testing definition and techniques resulted in three ISO standards, namely ISO 9241-210 (ISO 2018), ISO 25010 (ISO/IEC 25010 2011), and ISO 25066 (ISO/IEC 25066 2016). As shown in Figure 4 and Table 1, the three standards evaluate user interaction with front-end development or design in creating, testing, and refining the user's side of a website or application.

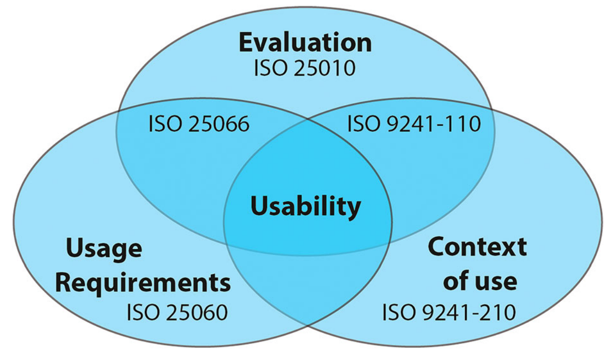

Figure 4. The key usability testing standards and their domain of operation for software and human computational interactions

During a usability study researcher, follow participants as they interact with the product. Therefore, users' feedback helps the BPS software design team improve the user experience.

Usability studies can take place at various points in the product design process. Concept and interface testing are applied at the beginning of the product prototype development with few interactions. A usability study should be conducted when the prototype is fully interactive. This is the most common time of usability testing because it gives the design team insights on what needs to be revised or added before the product launches. Also, usability testing should be conducted when the application of the software tool is complete. Study participants navigate the software from the opening page to check out as part of any usability. As a researcher, the developer collects feedback while watching the users interact with the prototype. In some usability studies, users get interviewed after interacting with the prototype to get more feedback. The following section describes the key stages of usability testing based on the reviewed ISO standards.

5.1. Planning

One of the key elements of conducting usability testing is to define the study goals and key study questions. By establishing research goals by the developers' team before the usability study begins, designer biases can be avoided. The phrasing of interview questions must be done thoughtfully. Open-ended questions and not leading questions are recommended. Also, ready-validated tests and metrics are encouraged to be used. Also, the research criteria must be defined beforehand. Based on those elements, a step plan must be established to conduct the study, including the participants' selection process. The planning of the usability testing is the key step for a proper plan to get successful research results and honest feedback.

Table 1. Summary of relevant ISO standards used as a basis to develop a framework USER-FIT for building performance simulation usability

Standard	Title	Relevance to USER-FIT
ISO 9241-210: 2019	Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems (ISO 9241-210 2010)	• Define the approach to centered design for the human operator • Specify requirements for users based on principles of usability
ISO 9241-110: 2020	Ergonomics of human-system interaction – Part 110: Interaction principles (ISO 9241-210 2010)	• Specify the needs of users in the interaction with the system
ISO/IEC 25010: 2011	Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models (ISO/IEC 25010 2011)	• Specify the features for assessing the quality of a software
ISO/IEC 25060: 2017	Information technology – User interface component accessibility – Part 25: Guidance on the audio presentation of the text in videos, including captions, subtitles, and other on-screen text (ISO 25060:2010 2010)	• Define terminology information related to the central design of the human • Specify information related to usability
ISO/IEC 25066: 2019	Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Common Industry Format (CIF) for Usability – Evaluation Report (ISO/IEC 25066 2016)	• Understand how to document an evaluation of usability

5.2. Participants recruitment

It is recommended to recruit a handful of participants for the usability study. Five participants are the minimum to start with in a usability study. This sample size is large enough to uncover major user issues but small enough to keep costs and time down. Participants must be found from a representative sample. The small group of testers should represent the key user groups and user groups that are often marginalized. The recruitment can continue beyond five until the research team reaches saturation, and the moderator must understand the feelings and emotions behind their answers and the reason behind each answer. Follow-up questions for more clarity are very encouraged.

5.3. Types of usability testing

Several usability techniques have arisen over the years, with the most popular methods being usability testing and heuristic evaluations. According to Nielsen, '*heuristic evaluation is a process where experts use rules of thumb to evaluate the usability of a user interface independent reviews and reported issues*' (Nielsen 1995). Evaluators use established heuristics (e.g. Nielsen-Molich's) and reveal insights that can help design teams enhance product usability from early in development (Rosenbaum 1989). However, heuristic evaluations are not recommended for BPS tools because they do not involve the user's opinion in the testing. Therefore, usability testing is the most effective method to evaluate user interface with real tasks through potential users. The problems found with usability testing are true problems in that at least one user encountered each problem.

Among the usability testing methods, there are two types of usability studies: moderated and unmoderated. A moderator guides participants through the study in real-time in a moderated usability study. The moderator aims to help participants interact with the product and collect feedback. On the other hand, an un-moderated usability study does not have a designated moderator. In the unmoderated usability study, participants test the prototypes without human guidance. Usually, the study is recorded on video, and the user experience team reviews the video footage after the study. Moderated and unmoderated usability studies have benefits and limitations depending on the scope and goal of the study. As shown in Table 2, moderated usability testing is more advantageous than unmoderated usability studies; however, both types suffer from subjectivity and bias.

Usability testing requires practice and experience so that moderators interact well with participants, which is critical to the success of any usability test and the validity of insights. The book of Jospeh, Dumas, Beth, and Loring (2008) presents the golden rules of moderating a usability test (Dumas and Loring 2008).

5.4. Conducting the usability and validation

Conducting the usability testing will be mainly based on gathering and collecting data. Major pain points of users must be identified based on specific criteria and usability tests with the product. Chapter 5 presents a comprehensive usability testing framework with several criteria and steps to conduct a usability test for BPS software. After collecting the data based on the proposed criteria and steps, researchers should synthesize the results, looking for the actual meaning in the data. Patterns in quantitative data should be identified, and similar trends in qualitative data from participants' answers to interview questions. The study must be validated before sharing or promoting any insights with developers.

A usability test must be unbiased and neutral regarding the product. As a qualitative method, a usability study

Table 2. Comparison of the advantages and disadvantages of moderated and moderated usability studies (UX Design 2022).

	Benefits	Limitations
Moderated usability studies	• Guide the participant through the study • Ask specific questions and follow up in real-time • Rapport building between the moderator and participant	• The moderator could influence or bias the participants • Less flexible • Participants may not identify with the moderator
Unmoderated usability studies	• The participant uses the tool in the real world • Participant completes a task on their own time and in their own space • Participants may feel more comfortable giving feedback without others around	• No human guidance if issues arise • No real-time follow-up questions • Little to no control over the environment

is a subjective approach. Therefore, usability researchers must be aware of a series of biases. A bias is a favoring or having a prejudice based on limited information. All humans are susceptible to those biases. The key is to identify those biases and to become aware of them to guard against them. Many biases are likely to come up during the moderation of usability studies, including:

- **Implicit bias:** The collection of attitudes and stereotypes we associate with people without knowledge.
- **Serial position effect:** When given a list of items, people are more likely to remember the first few and the last few, while the items in the middle tend to blur.
- **Friendly Bias:** The tendency of people to agree with those they like to maintain a non-confrontational conversation. Honest feedback is lost.
- **Social Desirability Bias:** People tend to answer questions in a way that others will view favorably.

Identifying the biases and minimizing their effects is part of any usability testing study. Testing and revising the usability testing results allows for more objective evaluations. As a qualitative method, usability studies is a subjective approach that provides many rewarding insights with a shallow risk (Connelly 2016; Tenner 2015). But to avoid bias, it is recommended to triangulate the usability tests and feedback forms through various types of tests (see Section 4.3) to validate the findings (Byrne 2001). Triangulation confirms and validates the quality results using quantitative studies. Using multiple sources and methods can minimize inadequacies in one approach or process.

6. Usability testing framework for building performance simulation

The usability testing framework USER-FIT is proposed based on the systematic literature review, memo log, and prolonged engagement. As shown in Figure 5, the framework is developed for usability testing of building performance simulation tools. This framework has divided the process of BPS tools usability testing into six working steps in line with ISO standards (Beausoleil-Morrison 2019; Charles and Thomas 2009; Li, Wang, and Hong 2021) and collecting systematic data on decision needs, decision support, and decision quality. Each step identifies the responsible persons, activity, input, and output.

The first step is to specify the (1) Design Decisional Need to set a goal of developing usability planning and assign actors. The second step is to assess the user need to define the design decision support needs, design constraints, and relevant parties with the testing process. The third step is to identify test cases and design tasks during the workflow to propose and evaluate the utility of decision support. The fourth step is to study the execution of the usability testing system to support decision-making and set system requirements. The fifth step is to visualize information and test the interaction and interpretation of results. The final sixth step is to evaluate the decision support quality and the design decisions according to the design brief of design decisional need and building performance requirements. The results from usability could be used to revise the tool or software and help detect some errors in interface functionality. USER-FIT framework comprises six-stage usability testing and should be used for communication and documentation. In Sections 5.1–5.6, each stage will be described in detail.

6.1. Design decisional need

Assessing the design decision-support needs is key in managing users' expectations and eliciting the key decision determinants. This stage involves screening and identifying the expected knowledge concerning the design intent and decisional conflicts that the BPS should

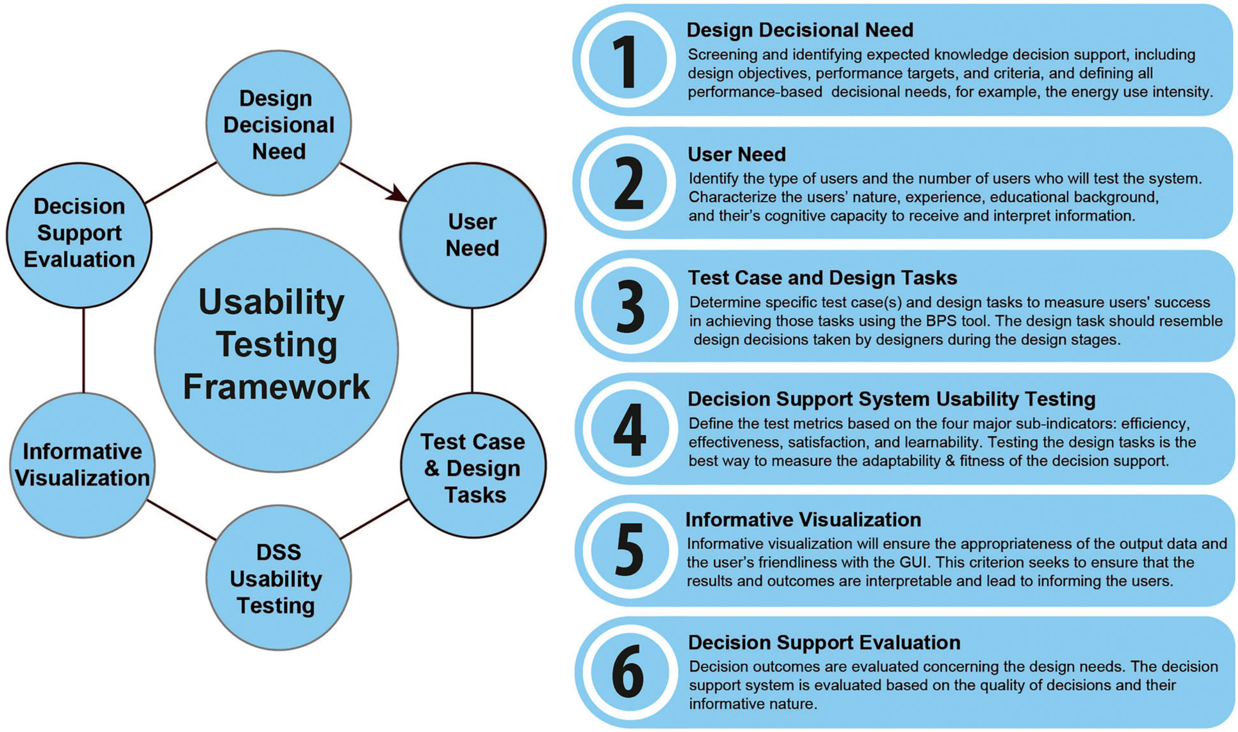

Figure 5. The USER-FIT framework of building performance simulation usability

support. The characteristic of the design choice or compliance requirements that need to be made (e.g. geometry, envelope, HVAC systems, renewables), number of design parameters (their range and intervals) (Wang, Teigland, and Hollberg 2024), design alternatives, performance targets (energy use intensity, greenhouse gas emissions, cost, discomfort) (Afroz, Gunay, and O'Brien 2020), degree of uncertainty, the robustness of outcomes, whether it is irrevocable must be assessed (AIA 2019). This stage requires defining benchmarks and key performance indicators, such as Energy Use Intensity EUI or GHG emissions. Therefore, cognizance of the architectural design problem or design situation, alternatives, and outcomes is essential (Augenbroe 2019). Evoking the perceived likelihood or probability of design outcomes of each alternative can allow software developers to better understand users' decisional needs. The phase of decision-making in the context of project design work stages (pre-design, schematic design, or design development) must also be determined. The AIA (AIA 2019) and RIBA (Sinclair 2019) defined phases of designing and building a project. Building performance simulation tools' expected role in proactively informing decision-making must be pre-defined. Assessing the design decision determinants can provide software developers and modelers with important cues regarding areas of decisional conflict. Accurate and complete decisional needs assessment sets the BPS tools for effective decision-support interventions. A designer's unique perception of the decision based on their knowledge of parametric variations or design alternatives, possible project outcomes, and the likelihood of their occurrence constitutes an example of one determinant.

6.2. User need

The second stage of usability testing is to identify the users, their ways of communication, their ability of data interpretation, and the number of users (Faulkner 2003) for testing the system or BPS software (Süsser et al. 2022). It involves assessing the users' requirements for the design and evaluation of the usability of BPS tools. User needs analysis is an important criterion that provides information necessary for the simulation and design decision analysis (Østergård, Jensen, and Maagaard 2016). User analysis is the process of identifying the types of users, including their needs, ontology, ways of communication, and characteristics of each type of user. For BPS, the type of users includes designers at various levels (e.g. modelers, architects, mechanical engineers, building physicists, and urban planners) and in various specialty areas (e.g. envelope design, HVAC design, acoustic design, lighting design, urban design) and students at various specializations. User characteristics for each type of modeler include knowledge and experience of BPS, education background, knowledge of computers including programming, cognitive capacities and limitations, perceptual variations, age-related skills and personality (Østergård 2017). User analysis can help software developers create supportive BPS environments targeted toward specific users and have the proper knowledge and information structure that match the users' literacy level (Imam, Coley, and Walker 2017). Several studies investigated the customization of BPS tools based on user needs (Alsaadani and De Souza 2016), e.g. architects or engineers, to perform building simulations and inform the design decision (Alsaadani and Bleil De Souza 2019; Alsaadani and De Souza 2016).

6.3. Test case and design tasks

Determining the test case is identifying the steps of a usability testing operation using a specific design case study. As part of the BPS usability, a set of design tasks must be defined and tested concerning the case study (Häkkinen, Ala-Juusela, and Shemeikka 2016). The set of tasks is then analyzed to identify what makes a task easy or difficult. A critical objective of the test case design tasks analysis is to find out why a task's user interface or simulation result representation (e.g. a bar chart vs. a scatter plot to find the primary energy use of a building over a year) is better than another. Modelers can compare user performance associated with different user interfaces by defining a testing task battery and performing a task analysis for the same design decision implemented in different user interfaces (Zhang and Hong 2017). Moder can perform the comparison in terms of time on task, the number of steps, and mental effort, all metrics of usability efficiency (see 5.4).

Design tasks need to be selected to increase the exposure of architects, engineers, and urban planners to decision support and decrease identified barriers interfering with enhancing designers' knowledge and skills to establish environmental design competencies. Problem-based test case scenarios and specific design tasks, such as determining the nearly optimal window-to-wall ratio for different building orientations or sizing of the HVAC systems, should be developed to build knowledge and skills to use them for performance simulation tools better.

6.4. Decision support system usability testing

The fourth criterion of the USER-FIT framework is to identify the test metrics such as success rate, error rate, efficiency in time on task, user satisfaction, etc. These metrics are selected based on the goal and objective of system testing. The result from usability could be used to revise the system and help detect some errors in system functionality. The usability testing of BPS tools should be user-centered and based on decision-making indicators and metrics. These indicators and metrics should be selected based on the goal, design stage, and decision support system testing objective. In this step, the quality of the BPS tool functionality is assessed to detect some errors and revise the decision support system. Knowledge of designers' backgrounds, contextual design realities, and identified design decisional needs will help determine which design tasks are amenable to decision support. Decision support should be adapted to the designer's characteristics and preferred role in decision-making. Consideration should be given to the designer's experience, specialization, and background.

Four major sub-indicators, efficiency, effectiveness, satisfaction, and learnability, should be used for testing. The selection of the decision-support indicators should be guided by implementing the ISO 9241-210 and ISO 25010 requirements.

6.4.1. Efficiency

Efficiency is a useful way to measure the resources expended concerning the accuracy and completeness with which users achieve their goals. This characteristic can be timing the user on a task during the usability testing, comparing users' time with peers, or counting the error rates. The following paragraphs explain the three most common efficiency measures based on ISO 25010 and (Albert and Tullis 2013). Those three metrics, mouse heat maps and eye-tracking cameras, can also be used for more quantitative insights.

a. *Mean time per task:* The average task completion time of all the users participating in the usability testing becomes a measure of efficiency. Calculating the time on task requires coming up with time-based efficiency. The formula for calculating time-based efficiency is indicated in Equation 1, where the end time is subtracted from the start time (Albert and Tullis 2013).

$$Time\ Based\ Efficiency = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N}\frac{n_{ij}}{t_{ij}}}{NR} \quad (1)$$

Where: N = Number of tasks; R = Number of users; N_{ij} = Result of task i by user j; if the user successfully completes the task, then $N_{ij} = 1$, if not, then $N_{ij} = 0$

T_{ij} = The time spent by user j to complete task i. If the task is not completed, time is measured until the moment the user quits the task.

b. *Overall relative efficiency* is the ratio of the time users take to complete a task to the total time taken by all other users. It can be calculated in Equation 2 (Albert and Tullis 2013).

$$\textit{Overall Relative Efficiency} = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N} n_{ij}t_{ij}}{\sum_{j=1}^{R}\sum_{i=1}^{N} t_{ij}} \times 100\% \quad (2)$$

c. *Error rate and frequency of occurrence:* This measure counts the number of errors participants made when attempting a given task. Counting the number of errors made by users can be daunting for many researchers, but this metric provides excellent information about the usability of a system. Learnability considers how easy it is for a user to perform a task the first time they encounter an interface and how many repetitions are required to perform it effectively. To help measure this metric and obtain valuable insights from this metric, a short description needs to be set where details about how to score those errors and the severity of certain errors are given. This metric can show how simple and intuitive the system is.

6.4.2. Effectiveness

Effectiveness is a metric that measures users' task completion and accuracy. The sub-indicator includes wider concepts related to decision-making robustness. Effectiveness can be measured using two main usability metrics: the completion rate, also called the success rate and the accuracy of task results. The following paragraphs explain the two most common effectiveness measures based on ISO 25010 and (Albert and Tullis 2013).

a. *Completion rate:* The percentage of users who successfully completed the tasks. The metric can be collected during any stage of development. Effectiveness can thus be represented as a percentage by using Equation 3. The completion rate is highly dependent on the context of the task being evaluated.

$$\textit{Effectivness} = \frac{\textit{number of tasks completed sucessfully}}{\textit{total number of tasks undertaken}} + 100\%$$

b. *Accuracy of task result:* The accuracy with which users achieve specified goals.

6.4.3. Satisfaction

Satisfaction measures the capacity of the decision support system to satisfy users in a specified context of use. The After Scenario Questionnaire (ASQ) and the System Usability Scale (SUS) provide a quick and reliable way of measuring users' satisfaction (Bangor, Kortum, and Miller 2008). The following paragraphs explain the two most common satisfaction measures based on ISO 9241 and (Albert and Tullis 2013).

a. *Task Level Satisfaction:* After users attempt a task, they answer a questionnaire to measure its difficulty. Typically ASQ consists of up to five questions. These post-task surveys are usually in the form of Likert-scale ratings and aim to gain insight into the task's difficulty from the participants' perspective.
b. *Test Level Satisfaction*: Test Level Satisfaction is measured by giving a formalized questionnaire to each test participant at the end of the test session. This estimates their impression of the overall ease of use of the tested system. The SUS consists of a 10-item questionnaire in which respondents have five response options ranging from 'strongly agree' to 'strongly disagree'. Scores should be added, and the total multiplied by 2.5. The mean score of the chosen responses should range between 0 and 100. The higher the score, the more users are satisfied. Any value above 75 is considered a good satisfaction rate.

6.4.4. Learnability

Learnability considers how easy it is for a user to perform a task the first time they encounter an interface and how many repetitions are required to perform it effectively. Learnability can be measured using three main usability metrics: first-use learnability, learning curve steepness, and efficiency of the interactions.

a. *First-use learnability:* How easy is the interface for first-time users to use? This is important for users who plan to use the software or tool only once.
b. *Learning curve steepness:* How quickly do users improve with repeated interface use? This is important for users who want to feel a sense of progression with the interaction and memorability for getting better at using software or tools.
c. *Efficiency of the interactions:* How high can users reach productivity once they fully learn how to use the software or tool? This aspect is vital for users who frequently use the software or tool.

The three types of metrics can be measured through a five-scale Likert question or recorded video observation, which timed and recorded the screen to identify hot spots and assume the time taken for a design task. The study of Baxter and Oatley (1991) provides an example of software learning ability measurement.

6.5. Informative visualization

Informative visualization of the simulation result for a given task performed by a specific type of user allows for extracting meaning. The appropriateness of the simulation data and its representation must be evaluated for any BPS graphical interface or dashboard (Pilgrim 2003). Communicating data to BPS users and clients is very difficult. Therefore, it is important to visualize building performance (Donn, Selkowitz, and Bordass 2012). Different visualizations of a common environmental performance indicator can generate different representational meanings, design decision difficulties, and cognitive outcomes. One of the most influential design decision support means to compare a design alternative or parametric variation of the same building and determine whether it is understandable for the task and the user. An analysis of the informative nature of data visualizations and users' ability to interpret is key. Expert review of graphical usability and the effectiveness of various types of representation is a major step of BPS software evaluations (Li et al. 2022). Communicating data-driven analysis requires high-quality data visualization. This can be only achieved through intensive usability testing of user interface properties and their ability to inform users.

6.6. Decision support evaluation

The decision support provided by BPS suits should be evaluated from the perspective of the designers and design teams. Decision outcomes can be evaluated from multiple perspectives. An architect may consider a good decision to be one that best reflects their preferences or causes the least negative consequences. Decision support systems should be able to support high-quality design decisions as one that is informed, consistent, and one in which BPS users express satisfaction with how the decision was arrived at (Attia 2012). Moreover, the satisfaction with the design outcome must be evaluated separately from the satisfaction of the decision-making quality discussed in criterion four (see section 5.4). The decision-making process can be evaluated in terms of whether the designer participated and developed more confidence and skills. Skills and capacity building can be assessed by the degree to which designers and design teams can transfer the decision-making skills learned to future decisions.

The decision support evaluation does not only involve users' satisfaction or perception, but it can also involve post-occupancy assessment. The decision outcomes, such as improved environmental building performance, reduced occupancy complaints, reduced operation cost or embodied energy, and appropriate facility management, can also be examined through post-occupancy evaluation (Göçer, Hua, and Göçer 2015). Thus, a designer might have felt empowered and satisfied with the decision support tool, while the design outcome may not have been what they expected. Therefore, it is crucial to determine how valuable and meaningful was the BPS tools or software in the decision-making (Bleil De Souza and Tucker 2015). The evaluation might include visiting the building after occupancy and reflecting on the energy performance gap (De Wilde 2014).

7. Interview results

The following sections summarize the feedback from interviews conducted with 12 practicing architects. Three females and nine male persons with ages ranging from 35 to 55 years were interviewed. The interviewees were all architects. The interviewees come from six Belgian architectural and engineering firms.

7.1. Validation of the framework

During the early interview stages, the final version of the USER-FIT framework, shown in Figure 5, was not articulated. The questions helped the authors to group and classify the interview responses into six categories. At the end of the interview stage and after analyzing the responses of interviews number seven the architects shared their strong agreement to categorize the usability testing process under six main categories. Saturation was reached through the continuation of the interviewee recruitments until 12 participants. The recruitment processes ended in summer 2023 when insights and answers started to replicate and overlap. Data saturation was used as a factor to judge the repetition of answers by expert that indicated similar relevant insights.

7.2. Quotations

The interview analysis and memo logs allowed to highlight several key responses from interviewees, called quotations. A list of quotation was created and is shown in Table 3. The most relevant and representative quotes to validate the framework are presented. All quotes were addressed by one of the steps of the framework. These verbatim quotations played a key role in clarifying the data and framework steps needed to conduct a usability testing for a BPS tool.

Table 3. Some relevant interview quotations

Quotations	Interview #	Step #
'I can only integrate the modeling result during one or two design iterations, the tool has to blend into the workflow of the busy design process'	2	1
'Time is scarce, the tool should be efficient in term of time requirments'	5	2, 4
'I don't care about accuracy but I care about comparisons'. I need to improve my awareness through the BPS tool use and not find absolute performance values'	9	3
'I need the tool to warn me from what might went wrong with modeling assumptions'	1	4
'The user interface needs to be more intuitive and simple to learn: Short steps, clean UI and easy understanding of results. The tool has be simple to learn'	6	4
'The tool is too complex and does not fit in the design process allowing to compare design alternatives. The tool should report the results in a comparative way through common visual graphs'	11	5
'I use the tool for insights and relevant findings, I need the tool to help me reflect on the consequences of my design choices'	7	6
'The simulation results must be accurate to ensure a reliability of results and to help me convince the client'	4	6

8. Discussion

This paper presented USER-FIT, a usability-testing framework to evaluate building performance simulation tools used to support environmental design decision-making for buildings and cities. The framework has the following working steps: (1) for describing decisional support needs; (2) for defining user needs; (3) for testing the usability of test cases and design tasks; (4) for testing the usability of the decision support system; (5) for providing informative visualization and (6) evaluation of decision support, once fully developed, for developing BPS Usability guidelines and standards. This section discusses the study findings and positions them regarding the state-of-the-art.

8.1. Findings and recommendations

In summary, the following strategies for further developing and synergizing usability testing are required to facilitate the use of performance simulation tools to enhance the quality of designers' involvement in environmental design decision-making:

1. A capacity-building strategy aimed at informing, influencing, and assisting BPS modelers and other building design professionals to gain competence in providing quality design decision support, learnability and memorability through pre- and post-licensure education.
2. A regulatory strategy aimed at embedding indicators of quality building design decision-support in legislation and policy, for compliance and performance-based design (i.e. accreditations standards, the scope of practice regulation, environmental and sustainability design certification, reflective practice tools);
3. A research strategy designed to evaluate further hypotheses underlying the framework for performance simulation tools in the built environment; interventions for sustainable implementation of building design decision support in routine architectural practice and designers' performance across AEC design firms and with diverse building functions; and
4. An architectural and engineering strategy to balance cost, time, accessibility and performance (efficiency, effectivness and satisfaction) of simulation tools, including supporting corrective (error detection) environmental design decision-making.

Facing environmental design decisions can be exceedingly complex for architects, mechanical engineers, and design professionals. Environmental design decisions have evolved from concrete and problem-bound to a constellation of shared, environmentally informed decisions across the AEC design industry continuum. To fully implement environmental design decision-making, architects need to be supported to participate in weighing potential outcomes. BPS software developers are well-positioned to provide the necessary usability testing, but many lack the awareness, knowledge, and skills. Modelers need opportunities to develop and use new skills to be design-centered and responsive to architects' needs.

Furthermore, design practice environments, regulatory bodies, and educational systems can facilitate the integration of usability testing skills in practice. Promoting policies that endorse architects' decision-support interventions that are practical, visible, accessible, evidence-based, and equitably enforced is essential. Professional accreditation bodies can influence overcoming barriers to usability testing and creating supportive environments, exemplary workflows and software tools. By enhancing modelers' and architects' decision-support skills as individuals and collectively as self-regulating professionals, models can make essential differences in the quality of environmental building design decisions.

8.2. Strengths and limitations

Although this research has primarily been focused on usability testing for architectural performance simulation tools, emerging evidence supports the relevance of the simulation-based design decision support approach for other types of design professionals, such as energy modelers, mechanical engineers, and urban designers. Theoretically, USER-FIT is transferrable across architectural and engineering design environments. Studies conducted in different countries indicate that architects want to participate in making environmental design decisions and an abundance of decisions across AEC design settings that are innately challenging, given the need to weigh the benefits and risks across design alternatives and parametric options (McLennan 2004). This novel framework will help software developers as well as BPS users. In practice, each design decision can potentially cause uncertainty about the best course of action. Therefore, the framework can be used for tailored coaching to the unique needs of designers and clients to perform BPS. The study successfully triangulated the findings and presented key performance criteria that could be used to redesign BPS products to improve usability.

A usability testing framework like USER-FIT and ISO standards (ISO 25010, ISO 25010 and ISO 9241-210) can be used by independent researchers and modelers of IBPSA to ensure the ease of use of BPS in the architectural and urban design practices. To the best of our knowledge, USER-FIT is the first framework that seeks to be implemented and translated into usability tests by BPS developers in the AEC industry based on Human Computation Interaction (HCI) principles. Despite the work of Hopfe (Hopfe and Hensen 2009), Mahdavi (Bazafkan, Pont, and Mahdavi 2019), Bleil De Souza (Bleil De Souza and Tucker 2015) and Petersen (Purup and Petersen 2020a; Purup and Petersen 2021) many other studies listed in the literature review (Section 2), there has been no study that developed a steped and systematic approach to evaluate the total user experience of BPS tools users. The paper provides the foundation and essential information to understand usability testing, reporting, and user experience reality in BPS-based and performance-based design. The paper will raise the bar for future usability testing of BPS software for what steps and criteria need to be followed and implemented. The paper provides a simple and first-hand insight into usability tests theory and the theoretical boundaries to develop and perform usability tests by and for users and third-party evaluations.

Although we did not perform a case study or experimental trial, the framework criteria were validated through the interviews. We are aware that the paper would have benefited from a demonstration or experimental trial on the applicability of USER-FIT through user experience testing. However, the embodiment of this work is to present a theoretical framework based on the state of the art. As mentioned in the introduction and methodlogy this is a non-empirical conceptual article that integrate an extensive set of literature theories under a novel theoretical umbrella (Jaakkola 2020). The empricial research validation of 12 simulation-practicing architects played the roled of peer examination and their responses reached convergence and saturation. Also homogeneity could have improved if the sample of interviewwes included feedback from complete desin teams rather individual designers. Therefore, remains relevant when explored in the context of the existing BPS literature (Mahdavi 2011 ; Bazafkan 2017; Attia et al. 2013; de Klerk et al. 2019; Purup and Petersen 2020b).

8.3. Implications on practice and future research

In this study, we presented a consistent and structured framework for usability testing. No matter how dedicated the developers of BPS programs are to providing data-driven simulation results, the ability of BPS tools to explain the building physics phenomena and provide building performance interpretations to the designers remains weak. Most existing BPS tools fail to help users interpret simulation results to understand the buildings' heat balance in the form of design parameters that influence the total heat gains and losses. At the same time, The designer's background and experience are essential determinants of performance-based design. Environmental design problems such as high-embodied carbon materials, large glazing surfaces, undersized HVAC systems, thermal bridges, and discomfort may constrain a building designer's ability to think clearly or participate effectively in decision-making. Involving substitute decision-makers and essential others in decision-making requires architects and designers to develop various decision-support strategies and communication styles. Tailored decision-support for design tasks may include providing focused information, facilitating access to appropriate resources, helping designers re-align project outcome expectations, and assisting in clarification of design preference. In addition, support can involve helping designers strengthen their repertoire of decision-making skills through guidance, coaching, and rehearsal.

Usability testing effectively integrates BPS tools in designer practices during early design stages and throughout the project delivery process (AIA 2019). So far, few design firms have developed in-house usability testing protocols and case study testing practices to evaluate the implementation of simulation-based design decision support. Integrating new BPS programs with performance-informed workflows and tools in design firms is a challenge from a business perspective (Hong, Langevin, and Sun 2018). Specific organizational barriers impacting the implementation of computational design and performance simulation-based design decision support in urban, architectural, and engineering design practice include lack of administrative direction to use building performance simulation to provide design decision support, lack of usability testing of simulation tools during the different stages of projects delivery, limited or no mention of performance simulation-based decision support in the final design, and time pressures that do not integrate simulation results as part of the design process and design reviews iterations, and lack of an interdisciplinary design approach (Alsaadani and De Souza 2016).

Future research should focus on the pilot studies and empirical validation of the framework with design professionals. The validation can be achieved during performance-informed design workshops, workflow observation sessions and embedding usability testing in the professional design practice to reveal users' expectations of BPS tools (Cozza, Jusselme, and Andersen 2018). Usability testing must be part of any computational or performance-based design approach. At the same time, usability testing, best practice guidelines, and skills are required in design firms to adopt the BPS practice. Usability testing can facilitate integrating the BPS use in the design context to address design problems to design a sustainable built environment. The evaluation of the total user experience of the BPS application during the project delivery process of performance-based designs is vital and can be done with the help of USER-FIT. To this end, many of them should be used as an organizational-level intervention for testing the usability of BPS in design practice environments.

9. Conclusion

This study conducted a systematic review of recent (2013–2023) Ph.D. studies that performed usability testing during the development of BPS tools and methods for design decision support during the early design stages. Challenges hindering the widespread use of BPS tools among the IBPSA community were highlighted. Next, we proposed a novel usability testing framework called USER-FIT that modelers and designers can adopt in professional practice through simple, applicable evaluation criteria and usability tests within organizational policies and structures that optimize the potential of BPS uptake. The framework is based on six criteria, namely: (1) for describing decisional support needs; (2) for defining user needs; (3) for testing the usability of test cases and design tasks; (4) for testing the usability of the decision support system; (5) for providing informative visualization and (6) evaluation of decision support, once fully developed, for developing BPS Usability guidelines and standards. The study describes the foundation and theoretical background behind the development of the framework. How USER-FIT can be fitted to the project delivery process to improve the usability of BPS in the design process, and workflows is presented. This paper states that usability is scientifically and objectively defined through a systematic and unified framework. The paper is a foundation to formalize user experience studies and usability testing for software tools, building performance models, simulation educators and building performance software developers. Directions for future BPS software development and user experience testing were recommended.

Acknowledgments

This research was funded by the Liege University pedagogical support grant for architectural-centered design decision support under the Digital Design Research project. We thank our colleagues, Michael Donn, Steffen Peterson, Ayu Miyamoto, Pil Purup, Andrew Corney and David Goldwasser, for many valuable discussions. We also would like to thank our former and current trainees, Kevin Tabsoba and many others, for conducting various analyses and illustrations cited in this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Shady Attia http://orcid.org/0000-0002-9477-5098
Maha Attia http://orcid.org/0000-0002-1222-3792

References

Abdelmegid, M. A., V. A. González, M. Poshdar, M. O'Sullivan, C. G. Walker, and F. Ying. 2020. "Barriers to Adopting Simulation Modelling in Construction Industry." *Automation in Construction* 111: 103046. https://doi.org/10.1016/j.autcon.2019.103046

ACE. 2019. "Assessing the Role of Architecture in Building Performance as Defined by the EU Level(s) Scheme," Architects' Council of Europe, Brussels, Belgium, 2019. Accessed August 07, 2022. [Online]. https://bit.ly/3Y68w2f.

ACE. 2021. "The Architectural Profession in Europe 2020." Mirza & Nacey Research Ltd, United Kingdom, 2021. Accessed July 08, 2022. [Online]. https://www.ace-cae.eu/fileadmin/user_upload/2020ACESECTORSTUDY.pdf.

Afroz, Z., H. B. Gunay, and W. O'Brien. 2020. "A Review of Data Collection and Analysis Requirements for Certified Green Buildings." *Energy and Buildings* 226: 110367. https://doi.org/10.1016/j.enbuild.2020.110367

AIA. 2019. "Architect's Guide to Building Performance: Integrating Performance Simulation in the Design Process." The American Institute of Architects, Washington DC, USA. [Online]. https://content.aia.org/sites/default/files/2019-06/Energy_Design_Modeling_Guide_v4.pdf.

Albert, W., and T. Tullis. 2013. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Burlington, MA: Newnes.

Alsaadani, S., and C. Bleil De Souza. 2019. "Performer, Consumer or Expert? A Critical Review of Building Performance Simulation Training Paradigms for Building Design Decision-Making." *Journal of Building Performance Simulation* 12 (3): 289–307. https://doi.org/10.1080/19401493.2018.1447602

Alsaadani, S., and C. De Souza. 2012. "The Social Component of Building Performance Simulation: Understanding Architects." Proceedings of the 2012 Building Simulation and Optimization Conference, Loughborough University, Loughborough, UK, September, 2012, 10–11.

Alsaadani, S., and C. B. De Souza. 2016. "Of Collaboration or Condemnation? Exploring the Promise and Pitfalls of Architect-Consultant Collaborations for Building Performance Simulation." *Energy Research & Social Science* 19: 21–36. https://doi.org/10.1016/j.erss.2016.04.016

Amer, M., M. Hamdy, T. Wortmann, A. Mustafa, and S. Attia. 2020. "Methodology for Design Decision Support of Cost-Optimal Zero-Energy Lightweight Construction." *Energy and Buildings* 223: 110170. https://doi.org/10.1016/j.enbuild.2020.110170

Attia, S. 2012. *A Tool for Design Decision Making: Zero Energy Residential Buildings in hot Humid Climates*. Belgium: Presses univ. de Louvain.

Attia, S., A. De Herde, E. Gratia, and J. L. Hensen. 2013. "Achieving Informed Decision-Making for net Zero Energy Buildings Design Using Building Performance Simulation Tools." In *Building Simulation*, 3–21. Shanghai, China: Springer.

Attia, S., E. Gratia, A. De Herde, and J. L. Hensen. 2012. "Simulation-based Decision Support Tool for Early Stages of Zero-Energy Building Design." *Energy and Buildings* 49: 2–15. https://doi.org/10.1016/j.enbuild.2012.01.028

Attia, S., J. L. Hensen, L. Beltrán, and A. De Herde. 2012. "Selection Criteria for Building Performance Simulation Tools: Contrasting Architects' and Engineers' Needs." *Journal of Building Performance Simulation* 5 (3): 155–169. https://doi.org/10.1080/19401493.2010.549573

Attia, S., R. Lioure, and Q. Declaude. 2020. "Future Trends and Main Concepts of Adaptive Facade Systems." *Energy Science & Engineering* 8 (9): 3255–3272. https://doi.org/10.1002/ese3.725.

Attia S., and A. Maha. 2023. "USER-FIT: A Framework to Evaluate the Usability of Building Performance Simulation Tools in the Architectural Design Process." Report, ISBN: 9782930909240, Sustainable Building Design Lab, Liege University, Liege, Belgium. Accessed February 15, 2024. https://orbi.uliege.be/handle/2268/310019.

Augenbroe, G. 2019. "Chapter 10: The Role of Simulation in Performance-Based Building." In *Building Performance for Design and Operation*, edited by Jan Hensen, and Roberto Lamberts. 2nd ed, 123–154. London: Taylor & Francis.

Bangor, A., P. T. Kortum, and J. T. Miller. 2008. "An Empirical Evaluation of the System Usability Scale." *International Journal of Human-Computer Interaction* 24 (6): 574–594. https://doi.org/10.1080/10447310802205776.

Baxter, I., and K. Oatley. 1991. "Measuring the Learnability of Spreadsheets in Inexperienced Users and Those with Previous Spreadsheet Experience." *Behaviour & Information Technology* 10 (6): 475–490. https://doi.org/10.1080/01449291089243 05

Bazafkan, E. 2017. "Assessment of Usability and Usefulness of New Building Performance Simulation Tools in the Architectural Design Process." PhD diss., Wien. Accessed February 13, 2024. https://www.researchgate.net/profile/Ehsan-Bazafkan/publication/320931455_Assessment_of_Usability_and_Usefulness_of_New_Building_Performance_Simulation_Tools_in_the_Architectural_Design_Process/links/5b58616daca272a2d6677f60/Assessment-of-Usability-and-Usefulness-of-New-Building-Performance-Simulation-Tools-in-the-Architectural-Design-Process.pdf.

Bazafkan, E., U. Pont, and A. Mahdavi. 2019. "Usability and Usefulness of Non-Conventional Building Performance Simulation Tools in Architectural Design Processes." *Applied Mechanics and Materials* 887: 219–226. https://doi.org/10.4028/www.scientific.net/AMM.887.219

Beausoleil-Morrison, I. 2019. "Learning the Fundamentals of Building Performance Simulation Through an Experiential Teaching Approach." *Journal of Building Performance Simulation* 12 (3): 308–325. https://doi.org/10.1080/19401493.2018.1479773

Bleil De Souza, C., and S. Tucker. 2015. "Thermal Simulation Software Outputs: A Framework to Produce Meaningful Information for Design Decision-Making." *Journal of Building Performance Simulation* 8 (2): 57–78. https://doi.org/10.1080/19401493.2013.872191.

Boyce, C., and P. Neale. 2006. *Conducting in-Depth Interviews: A Guide for Designing and Conducting in-Depth Interviews for Evaluation Input*, Vol. 2. Watertown, MA: Pathfinder International.

Byrne, M. M. 2001. "Evaluating the Findings of Qualitative Research." *AORN Journal* 73 (3): 703–706. https://doi.org/10.1016/S0001-2092(06)61966-2

Cetin, R., and A. Mahdavi. 2010. "Exploring the Availability and Usability of Web-Based Building Performance Simulation Tools." Bausim 2010-Building Performance Simulation in a changing Environment.

Charles, P. P., and C. R. Thomas. 2009. "Four Approaches to Teaching with Building Performance Simulation Tools in Undergraduate Architecture and Engineering Education." *Journal of Building Performance Simulation* 2 (2): 95–114. https://doi.org/10.1080/19401490802592798

Clarke, J. A. 2020. "Why Tools for Buildings and Cities Performance Simulation Need to Evolve." *Buildings and Cities* 1–4.

Connelly, L. M. 2016. "Trustworthiness in Qualitative Research." *Medsurg Nursing Journal.* 25 (6): 435.

Cozza, S., T. Jusselme, and M. Andersen. 2018. "Usability Assessment of Building Performance Simulation Tools: A Pilot Study." Second International Conference for Sustainable Design of the Built Environment-SDBE 2018.

de Klerk, R., A. M. Duarte, D. P. Medeiros, J. P. Duarte, J. Jorge, and D. S. Lopes. 2019. "Usability Studies on Building Early Stage Architectural Models in Virtual Reality." *Automation in Construction* 103: 104–116. https://doi.org/10.1016/j.autcon.2019.03.009

De Wilde, P. 2014. "The gap Between Predicted and Measured Energy Performance of Buildings: A Framework for Investigation." *Automation in Construction* 41: 40–49. https://doi.org/10.1016/j.autcon.2014.02.009

Dogan, T., and C. Reinhart. 2013. "Automated Conversion of Architectural Massing Models Into Thermal 'Shoebox'-models." Proc. BS2013.

Donn, M. 2023. "Misuse of Building Performance Simulation." Accessed February 15, 2024. [Online]. https://www.buildingsandcities.org/insights/commentaries/misuse-building-performance-simulation.html.

Donn, M., S. Selkowitz, and B. Bordass. 2012. "The Building Performance Sketch." *Building Research & Information* 40 (2): 186–208. https://doi.org/10.1080/09613218.2012.655070

Dumas, J. S., and B. A. Loring. 2008. *Moderating Usability Tests: Principles and Practices for Interacting*. Burlington, MA: Elsevier.

Faulkner, L. 2003. "Beyond the Five-User Assumption: Benefits of Increased Sample Sizes in Usability Testing." *Behavior Research Methods, Instruments, & Computers* 35 (3): 379–383. https://doi.org/10.3758/BF03195514

Fernandez-Antolin, M.-M., J.-M. del-Río, F. del Ama Gonzalo, and R.-A. Gonzalez-Lezcano. 2020. "The Relationship Between the use of Building Performance Simulation Tools by Recent Graduate Architects and the Deficiencies in Architectural Education." *Energies* 13 (5): 1134. https://doi.org/10.3390/en13051134

Gan, V. J., I. M. Lo, J. Ma, K. T. Tse, J. C. Cheng, and C. M. Chan. 2020. "Simulation Optimisation Towards Energy Efficient Green Buildings: Current Status and Future Trends." *Journal of Cleaner Production* 254: 120012. https://doi.org/10.1016/j.jclepro.2020.120012

Gerber, D. J., and S.-H. E. Lin. 2013. "Geometric Complexity and Energy Simulation–Evolving Performance Driven Architectural Form."

Göçer, Ö, Y. Hua, and K. Göçer. 2015. "Completing the Missing Link in Building Design Process: Enhancing Post-Occupancy Evaluation Method for Effective Feedback for Building Performance." *Building and Environment* 89: 14–27. https://doi.org/10.1016/j.buildenv.2015.02.011

Goldstein, R., and A. Khan. 2017. "Simulation-based Architectural Design." In *Guide to Simulation-Based Disciplines*, 167–182. Berlin, Germany: Springer.

Häkkinen, T., M. Ala-Juusela, and J. Shemeikka. 2016. "Usability of Energy Performance Assessment Tools for Different use Purposes with the Focus on Refurbishment Projects." *Energy and Buildings* 127: 217–228. https://doi.org/10.1016/j.enbuild.2016.05.062

Hamilton, S. H., C. A. Pollino, D. S. Stratford, B. Fu, and A. J. Jakeman. 2022. "Fit-for-purpose Environmental Modeling: Targeting the Intersection of Usability, Reliability and Feasibility." *Environmental Modelling & Software* 148: 105278. https://doi.org/10.1016/j.envsoft.2021.105278

Han, T., Q. Huang, A. Zhang, and Q. Zhang. 2018. "Simulation-Based Decision Support Tools in the Early Design Stages of a Green Building—A Review." *Sustainability* 10 (10): 3696. https://doi.org/10.3390/su10103696

Hand, J. W. 1998. "Removing Barriers to the Use of Simulation in the Building Design Professions." PhD diss., University of Strathclyde UK. Accessed February 13, 2024. [Online]. https://www.researchgate.net/profile/Jon-Hand/publication/2438187_Removing_Barriers_To_The_Use_Of_Simulation_In_The_Building_Design_Professions/links/0a85e5301c5d944695000000/Removing-Barriers-To-The-Use-Of-Simulation-In-The-Building-Design-Professions.pdf.

Hensen, J., E. Djunaedy, M. Radošević, and A. Yahiaoui. 2004. "Building Performance Simulation for Better Design: Some Issues and Solutions." Proceedings of the 21st Conference on Passive and Low Energy Architecture (PLEA), Citeseer, 1185–1190.

Hensen, J. L., and R. Lamberts. 2012. *Building Performance Simulation for Design and Operation*. Oxfordshire: Routledge.

Hong, T., J. Langevin, and K. Sun. 2018. "Building Simulation: Ten Challenges." In *Building Simulation*, 871–898. Germany: Springer.

Hopfe, C. J. 2009. "Uncertainty and Sensitivity Analysis in Building Performance Simulation for Decision Support and Design Optimization." PhD diss., Eindh. Univ.

Hopfe, C. J., and J. L. M. Hensen. 2009. "Experiences Testing Enhanced Building Performance Simulation Prototypes on Potential User Group." 11th IBPSA Building Simulation Conference, 27–30 July, International Building Performance Simulation Association (IBPSA), 1976–1982. Accessed February 13, 2024. [Online]. https://research.tue.nl/en/publications/experiences-testing-enhanced-building-performance-simulation-prot.

Hopfe, C., V. I. Soebarto, D. Crawley, and R. Rawal. 2017. "Understanding the Differences of Integrating Building Performance Simulation in the Architectural Education System."

IBPSA USA. 2020. "Building Energy Software Tools Directory." Accessed December 2, 2020. [Online]. https://www.ibpsa.us/best-directory-list/.

IBPSA-USA. 2023. "Building Energy Software Tools." Accessed May 18, 2023. [Online]. https://www.buildingenergysoftwaretools.com/.

Imam, S., D. A. Coley, and I. Walker. 2017. "The Building Performance Gap: Are Modellers Literate?" *Building Services Engineering Research and Technology* 38 (3): 351–375. https://doi.org/10.1177/0143624416684641

ISO. 2018. "ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs): Part 11: Guidance on Usability." ISO.

ISO 25060:2010. 2010. *Systems and Software Engineering*. Geneva: International Organization for Standardization. Accessed January 1, 2023. [Online]. https://www.iso.org/standard/35786.html

ISO 9241-210. 2010. *Ergonomics of Human-System Interaction – Part 210: Human-Centred Design for Interactive Systems*. Geneva: International Organization for Standardization. Accessed January 1, 2023. [Online]. https://www.iso.org/standard/52075.html.

ISO/IEC 25010. 2011. *System and Software Quality Requirements and Evaluation (SQuaRE)*. Geneva: International Organization for Standardization. Accessed January 1, 2023. [Online]. https://www.iso.org/standard/52075.html.

ISO/IEC 25066. 2016. *Systems and Software Engineering - Systems and Software Quality Requirements and Evaluation (SQuaRE) - Commom Industry Format (CIF) for Usability - Evaluation Report*. Geneva: International Organization for Standardization. Accessed January 1, 2023. [Online]. https://www.iso.org/standard/63831.html.

Jaakkola, E. 2020. "Designing Conceptual Articles: Four Approaches." *AMS Review* 10 (1-2): 18–26. https://doi.org/10.1007/s13162-020-00161-0.

Jayapalan-Nair, R., E. Brembilla, C. Hopfe, and J. Mardaljevic. 2024. "Usability Evaluation of a Building Simulation Web-Based Design Guide on Potential User Group." Accessed February 13, 2024. [https://repository.lboro.ac.uk/articles/conference_contribution/Usability_evaluation_of_a_building_simulation_web-based_design_guide_on_potential_user_group/12621941

Korkmaz, S., J. I. Messner, D. R. Riley, and C. Magent. 2010. "High-performance Green Building Design Process Modeling and Integrated use of Visualization Tools." *Journal of Architectural*

Engineering 16 (1): 37–45. https://doi.org/10.1061/(ASCE)1076-0431(2010)16:1(37)

Lam, K. P. 2020. "Sustainability Performance Simulation Tools for Building Design." In *Sustainable Built Environment*, edited by V. Loftness and D. Haase, 589–655. New York: Springer. https://doi.org/10.1007/978-1-4614-5828-9_349

Li, H., et al. 2022. "Issues in Bi-Directional Interoperability Between BIM and BEM." Construction Research Congress 2022, 1355–1364.

Li, H., Z. Wang, and T. Hong. 2021. "Occupant-Centric Key Performance Indicators to Inform Building Design and Operations." *Journal of Building Performance Simulation* 14 (6): 814–842. https://doi.org/10.1080/19401493.2021.1876771

Liege, Belgium. 2021. "In-Depth Interviews." Accessed February 15, 2024. [Online Video]. https://www.youtube.com/watch?v=z8lc1tUctDI.

Lin, S.-H. E., and D. J. Gerber. 2014. "Designing-in Performance: A Framework for Evolutionary Energy Performance Feedback in Early Stage Design." *Automation in Construction* 38: 59–73. https://doi.org/10.1016/j.autcon.2013.10.007

Lomas, K. J., H. Eppel, C. J. Martin, and D. P. Bloomfield. 1997. "Empirical Validation of Building Energy Simulation Programs." *Energy and Buildings* 26 (3): 253–275. https://doi.org/10.1016/S0378-7788(97)00007-8

Mahdavi, A. 2011. "The Human Dimension of Building Performance Simulation," Accessed February 13, 2024. [Online]. https://www.aivc.org/resource/human-dimension-building-performance-simulation.

Maver, T. W. 1995. "CAAD's Seven Deadly Sins." In *Sixth International Conference on Computer-Aided Architectural Design Futures*, edited by M. Tan and R. Teh, 21–22. Singapore: CASA.

McLennan, J. F. 2004. *The Philosophy of Sustainable Design: The Future of Architecture*. Portland, Oregon: Ecotone Publishing.

MEEX, E. 2018. *"Early Design Support for Material Related Environmental Impact Assessment of Dwellings,"*. Hasselt: Hasselt University. https://researchportal.be/en/publication/early-design-support-material-related-environmental-impact-assessment-dwellings.

Miyamoto, A., K. Allacker, and F. De Troyer. 2022. "Visual Tool for Sustainable Buildings: A Design Approach with Various Data Visualisation Techniques." *Journal of Building Engineering* 56: 104741. https://doi.org/10.1016/j.jobe.2022.104741

Morbitzer, C., P. A. Strachan, B. Spires, D. Cafferty, and J. Webster. 2001. Integration of building simulation into the design process of an architectural practice.

Natanian, J. 2021. "Beyond Zero Energy Districts: A Holistic Energy and Environmental Quality Evaluation Workflow for Dense Urban Contexts in Hot Climates." PhD diss., Technische Universität München.

Nault, E., C. Waibel, J. Carmeliet, and M. Andersen. 2018. "Development and Test Application of the UrbanSOLve Decision-Support Prototype for Early-Stage Neighborhood Design." *Building and Environment* 137: 58–72. https://doi.org/10.1016/j.buildenv.2018.03.033

Negendahl, K. 2015. "Building Performance Simulation in the Early Design Stage: An Introduction to Integrated Dynamic Models." *Automation in Construction* 54: 39–53. https://doi.org/10.1016/j.autcon.2015.03.002

Negendahl, K. 2016. *Consequence Based Design. An Approach for Integrating Computational Collaborative Models (Integrated Dynamic Models) in the Building Design Phase*. Copenhagen: Technical University of Denmark, Department of Civil Engineering. https://orbit.dtu.dk/en/publications/consequence-based-design-an-approach-for-integrating-computationa.

Newton, D., R. James, and D. Bartholomew. 1988. "Building Energy Simulation - A User's Perspective." *Energy and Buildings* 10 (3): 241–247. https://doi.org/10.1016/0378-7788(88)90009-6

Nielsen, J. 1995. "How to Conduct a Heuristic Evaluation." *Nielsen Norman Group* 1 (1): 8.

Oh, S., and J. S. Haberl. 2016. "Origins of Analysis Methods Used to Design High-Performance Commercial Buildings: Whole-Building Energy Simulation." *Science and Technology for the Built Environment* 22 (1): 118–137. https://doi.org/10.1080/23744731.2015.1063958.

Østergård, T. 2017. *Proactive Building Simulations for Early Design Support: Multi-Actor Decision-Making Based on Monte Carlo Simulations and Global Sensitivity Analysis*. Copenhagen: Aalborg. https://vbn.aau.dk/en/publications/proactive-building-simulations-for-early-design-support-multi-act.

Østergård, T., R. L. Jensen, and S. E. Maagaard. 2016. "Building Simulations Supporting Decision Making in Early Design – A Review." *Renewable and Sustainable Energy Reviews* 61: 187–201. https://doi.org/10.1016/j.rser.2016.03.045

Ouf, M. M., W. O'Brien, and H. B. Gunay. 2018. "Improving occupant-related features in building performance simulation tools." *Building Simulation* 11: 803–817. https://doi.org/10.1007/s12273-018-0443-y

Papamichael, K., and V. Pal. 2002. *Barriers in Developing and Using Simulation-Based Decision-Support Software*. Berkeley, CA: Lawrence Berkeley National Lab.(LBNL).

Pati, D., and G. Augenbroe. 2007. "Integrating Formalized User Experience Within Building Design Models." *Computer-Aided Civil and Infrastructure Engineering* 22 (2): 117–132. https://doi.org/10.1111/j.1467-8667.2006.00474.x

Petersen, S. 2011. *Simulation-Based Support for Integrated Design of New Low-Energy Office Buildings*. Copenhagen: DTU Civil Engineering, Technical University of Denmark.

Pilgrim, M. J. 2003. "The Application of Visualisation Techniques to the Process of Building Performance Analysis." PhD diss., Loughborough University.

Purup, P. B. 2021. "A Practice-Oriented Approach to the Development of Building Performance Simulation Tools Fit for the Early Stages of Building Design."

Purup, P. B., and S. Petersen. 2020a. "Requirement Analysis for Building Performance Simulation Tools Conformed to fit Design Practice." *Automation in Construction* 116: 103226. https://doi.org/10.1016/j.autcon.2020.103226

Purup, P. B., and S. Petersen. 2020b. "Research Framework for Development of Building Performance Simulation Tools for Early Design Stages." *Automation in Construction* 109: 102966. https://doi.org/10.1016/j.autcon.2019.102966

Purup, P. B., and S. Petersen. 2021. "Characteristic Traits of Visualizations for Decision-Making in the Early Stages of Building Design." *Journal of Building Performance Simulation* 14 (4): 403. https://doi.org/10.1080/19401493.2021.1961864

Rezaee, R., J. Brown, G. Augenbroe, and J. Haymaker. 2015. "The Application of Inverse Approach to the Early Stage of Performance-Based Building Design." 14th Conference of International Building Performance Simulation Association, Hyderabad, India. Accessed February 13, 2024. [Online]. https://scholar.google.com/scholar?cluster=17371927719978030828&hl=en&oi=scholarr.

RIBA. 2021. "RIBA Plan of Work." Accessed May 18, 2023. [Online]. https://www.architecture.com/knowledge-and-resources/resources-landing-page/riba-plan-of-work.

Rosenbaum, S. 1989. "Usability Evaluations Versus Usability Testing: When and Why?" *IEEE Transactions on Professional Communication* 32 (4): 210–216. https://doi.org/10.1109/47.44533

Roth, A., et al. 2018. "Scripting Frameworks For Enhancing EnergyPlus Modeling Productivity." Building performance analysis conference and SimBuild, Chicago IL, September, 2018, 26–28.

Samuelson, H. W., A. Lantz, and C. F. Reinhart. 2012. "Non-technical Barriers to Energy Model Sharing and Reuse." *Building and Environment* 54: 71–76. https://doi.org/10.1016/j.buildenv.2012.02.001

Shiel, P., S. Tarantino, and M. Fischer. 2018. "Parametric Analysis of Design Stage Building Energy Performance Simulation Models." *Energy and Buildings* 172: 78–93. https://doi.org/10.1016/j.enbuild.2018.04.045

Sinclair, D. 2019. *Guide to Using the RIBA Plan of Work 2013*. Oxfordshire: Routledge.

Son, H., S. Lee, and C. Kim. 2015. "What Drives the Adoption of Building Information Modeling in Design Organizations? An Empirical Investigation of the Antecedents Affecting Architects' Behavioral Intentions." *Automation in Construction* 49: 92–99. https://doi.org/10.1016/j.autcon.2014.10.012

Süsser, D., N. Martin, V. Stavrakas, H. Gaschnig, L. Talens-Peiró, A. Flamos, C. Madrid-López, and J. Lilliestam. 2022. "Why Energy Models Should Integrate Social and Environmental Factors: Assessing User Needs, Omission Impacts, and Real-Word Accuracy in the European Union." *Energy Research & Social Science* 92: 102775. https://doi.org/10.1016/j.erss.2022.102775.

Tenner, E. 2015. "The Design of Everyday Things by Donald Norman (Review)." *Technology and Culture* 56 (3): 785–787. https://doi.org/10.1353/tech.2015.0104

Turrin, M., P. Von Buelow, and R. Stouffs. 2011. "Design Explorations of Performance Driven Geometry in Architectural Design Using Parametric Modeling and Genetic Algorithms." *Advanced Engineering Informatics* 25 (4): 656–675. https://doi.org/10.1016/j.aei.2011.07.009

UX Design. 2022. "Usability Testing Tips and Examples." Google UX Design Certificate (Sep. 06, 2021). Accessed August 22, 2022. [Online Video]. https://www.youtube.com/watch?v=nYCJTea1AUQ.

Verheij, H., and G. Augenbroe. 2006. "Collaborative Planning of AEC Projects and Partnerships." *Automation in Construction* 15 (4): 428–437. https://doi.org/10.1016/j.autcon.2005.06.011

Wang, D., X. Pang, W. Wang, Z. Qi, Y. Ji, and R. Yin. 2020. "Evaluation of the dynamic energy performance gap of green buildings: Case studies in China." *Building Simulation* 13: 1191–1204. https://doi.org/10.1007/s12273-020-0653-y

Wang, X., R. Teigland, and A. Hollberg. 2024. "Identifying Influential Architectural Design Variables for Early-Stage Building Sustainability Optimization." *Building and Environment*, 111295. https://doi.org/10.1016/j.buildenv.2024.111295

Weytjens, L., S. Attia, G. Verbeeck, and A. De Herde. 2011. "The 'Architect-Friendliness' of six Building Performance Simulation Tools: A Comparative Study." *International Journal of Sustainable Building Technology and Urban Development* 2 (3): Art. no. 3.

Xiao, Y., and M. Watson. 2019. "Guidance on Conducting a Systematic Literature Review." *Journal of Planning Education and Research* 39 (1): 93–112. https://doi.org/10.1177/0739456X17723971

Zhang, R., and T. Hong. 2017. "Modeling of HVAC Operational Faults in Building Performance Simulation." *Applied Energy* 202: 178–188. https://doi.org/10.1016/j.apenergy.2017.05.153

Appendices

Appendix I

The systematic review protocol was developed with input from systematic review methodologists found in the literature (Xiao and Watson 2019). The eligibility criteria and information sources are described in Table A1.

Table A1. Summary of inclusion and exclusion criteria of the systematic review protocol

Purpose: Increase the uptake of building performance simulation for design decision support in practice.	
Research Questions: How to evaluate the ease of use of building performance simulation by building designers, modelers and developers?	
Keywords: user experience, usability, testing, trials, modeling, building, design,	Synonyms: building energy modeling, simulation, design process
Data Sources: Web of Science, Scopus, IBPSA conference proceedings, CAAD conference proceedings, cumniAD database, International Standardization Organisation (ISO)	https://clarivate.com/ http://www.scopus.com http://www.ibpsa.org/ http://papers.cumincad.org/ https://www.buildingenergysoftwaretools.com/
Search Strings: 'experience' AND 'modeling' AND ('user' OR 'usability') (('simulation' OR 'evaluation') AND 'design' AND 'buildings' AND ('architecture' OR 'urban')) AND 'systems' AND ('workflow' OR 'process') AND 'performance' AND 'energy' AND 'environment'.	
Inclusion Criteria: Research field: AEC, HCI; Language: English; Publication date: 2010 to 2023; Type of work: scientific articles; Availability: full text; Subject: usability testing, user experience.	Exclusion Criteria: Not belonging to AEC Field; Not in English; Publication before 2010; Not a scientific work; Not related to the subject of interest; Full text unavailable.

Appendix II

Table A2. Interview guide translated from French and Flemish.

Research Questions	Topic	Usability Concepts
Q1: How do you use BPS tools during the early design stages?	Integration S1, S2, S3	design process, workflow
Q2: How to you improve your BPS skills?	user expertise S2	experience, building physics, building engineering
Q3: How do you evaluate the usability of a BPS tool in the design process?	user experience S3, S4, S5	learning curve, total user experience, user interface, speed, steps, simplicity, intuitiveness, performance issues
Q4: How can a BPS tool inform your design decision?	decision support S5, S6	performance requirements, value, quality, advice, assistance