Chapter 18
Usability of Planning Support Systems: An Evaluation Framework

Patrizia Russo, Maria Francesca Costabile, Rosa Lanzilotti and Christopher J. Pettit

Abstract Previous research on Planning Support Systems (PSS) showed that low usability of these computer-based tools is one of the reasons why they are not widely used by planning professionals. Few studies for evaluating PSS usability are performed, possibly because developers do not regard it as their task, do not have enough skills to conduct them, and have not been stimulated so far to appreciate their value. In this chapter, a framework is described that aims at guiding usability evaluation of PSS; it is developed on the basis of a more general usability evaluation framework. The current version of the framework has been applied to evaluate three PSS by performing a test with a small group of land use planners. Results of this user test are discussed, also providing some recommendations for the design of PSS, specifically those addressing Land Suitability Analysis (LSA), capable to generate a positive user experience.

1 Introduction and Motivation

Planning Support Systems (PSS) are computer-based tools designed to support the activities of planning professionals. Both literature and field experts indicate that their adoption is very limited so far. In particular, low usability of PSS has been

P. Russo (✉) · C.J. Pettit
Faculty of Architecture Building and Planning, University of Melbourne, Parkville, VIC 3010, Australia
e-mail: p.russo@student.unimelb.edu.au

C.J. Pettit
e-mail: cpettit@unimelb.edu.au

M.F. Costabile · R. Lanzilotti
Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro, Bari 70125, Italy
e-mail: maria.costabile@uniba.it

R. Lanzilotti
e-mail: rosa.lanzilotti@uniba.it

S. Geertman et al. (eds.), *Planning Support Systems and Smart Cities*,
Lecture Notes in Geoinformation and Cartography,
DOI 10.1007/978-3-319-18368-8_18

identified as one of the most important factors for this (Te Brömmelstroet 2010; Vonk et al. 2005). Based on this finding, it is time to carry out further research on PSS usability.

Geertman and Stillwell (2003) reported that most PSS are not subjected to well-considered design processes and evaluations. Our experience indicates that PSS design does not sufficiently take into account the user interface, which is the most important component from the point of view of users. User-centred design and usability evaluation of PSS are rather rare (e.g. Vonk and Ligtenberg 2010; Arciniegas et al. 2013). During design and development of PSS, very little attention has been devoted to user-oriented aspects that might improve the overall experience of planning professionals, such as the effectiveness and efficiency of the interaction with such systems and the user engagement and satisfaction. Instead, PSS development has been very much technology-oriented (Vonk and Geertman 2008). Few studies for evaluating PSS usability have been performed, possibly because developers do not regard it as their task and have not so far been stimulated to conduct them. Allen (2008) suggested to rigorously conduct evaluations of PSS in order to encourage developers to increasingly consider user-oriented aspects and to facilitate the selection of appropriate PSS by potential users. It is worth remarking that conducting usability evaluation requires specific skills that PSS designers usually do not possess.

A framework that might serve as a guide for evaluating the usability of PSS is the focus of this research. It has been inspired by DECIDE, a more general usability evaluation framework provided in Rogers et al. (2011). The presented framework represents a first attempt towards more systematic PSS usability evaluation. It is expected to increase and improve PSS evaluation, by providing appropriate guidance to those people that should evaluate a PSS but have not yet enough expertise on usability. In fact, this framework aims to be used by different stakeholders, namely: (i) designers to conceive innovative and more effective PSS, (ii) developers to evaluate new PSS during their development and before their release, (iii) potential users to identify the PSS to adopt in their planning organisations, and (iv) teachers to select PSS suitable for didactic purposes. Thus, this evaluation framework could constitute an important step towards advancing the state-of-the-art in PSS usability.

After an introduction and definition of the concepts of usability and user experience in Sects. 2 and 3 describes the methodological approach and the current status of this framework. As an example of its application, a user test was conducted with a small group of land use planners; they engaged with three PSS that address Land Suitability Analysis (LSA) tasks, namely CommunityViz (http://placeways.com/communityviz), Envision (Newton and Glackin 2013) and Online What if? (Pettit et al. 2013). The participants were not familiar with the three PSS and tested them in a within-subjects design (Graziano and Raulin 2012). The user test was performed with a thinking-aloud approach, complemented with questionnaires. In addition, other methods of predominantly qualitative nature, including video and screen recording were triangulated. Section 4 briefly reports the user test and discusses some results. This section also provides recommendations for the design of PSS, specifically those addressing LSA, capable of supporting a good user experience. Section 5 provides conclusions and outlines next steps in the research.

2 Usability and User Experience

There are different quality factors which interactive systems can be evaluated against. This research focuses on usability evaluation of the user interface, to determine how well the functionality can be used for specific circumstances and environments, also depending on the tasks to be performed and users' characteristics. Usability is a complex concept, which has been defined in different ways in the literature, emphasising specific attributes. The definitions provided by ISO 9241-210 (ISO 9241 2010) and ISO 9126-1 (ISO 9126 1998) are widely acknowledged by the Human–Computer Interaction (HCI) and the Software Engineering communities; they focus on attributes such as effectiveness, efficiency, understandability, learnability and satisfaction. For more on usability definitions and attributes, refer to e.g., Rogers et al. (2011) and Ardito et al. (2014).

Nowadays, the importance of people's overall experience when interacting with software has been recognised by the HCI community. This increasing interest in user experience (UX) has occurred along with some developments in the field of Information and Communication Technology (ICT) over the years (Costabile and Buono 2013; Rogers et al. 2011). In the past, ICT aimed at providing useful and usable functionality, today it attempts to involve users in positive and engaging experiences. The occurrence of mobile systems and internet, which accompany people everywhere, have generated great attention on UX, going beyond the common usability attributes. Depending on the software, positive experiences can be determined, for example, by the level of pleasure and fun (e.g. video games) or satisfaction (e.g. business software that allows accomplishing tasks). Designers of ICT put increasingly emphasis on aesthetics and attractiveness that, according to the literature, generally help improving users' experiences. With respect to the user interface, a good layout with appropriate fonts and colours and engaging interaction techniques are likely to contribute to a better user experience. The research presented in this chapter primarily addresses usability; however, it also focuses on more subjective attributes typical of UX, such as pleasure and user satisfaction, which are also important aspects of usability.

3 A Usability Evaluation Framework for PSS

Evaluation is a central activity in the software design process that aims to assist in developing a product that satisfies users' requirements. Understanding user requirements is fundamental: it is a process of negotiation between designers and users over a certain period of time. Evaluation facilitates the understanding between designers and users, because involving users in evaluating a system prototype is instrumental in highlighting to designers what are the users' needs and expectations. Deciding when and how to evaluate a prototype (or a complete system) requires some expertise and may differ among products. While there are many books and websites that describe evaluation methods, novice evaluators actually need more guidance on how to plan and perform the overall evaluation. The DECIDE framework proposed by Rogers et al. (2011) addresses this need by providing a structure for performing evaluation studies, focusing on the issues that have to be taken into account. Its name derives from the initial of the name of each of the six main activities proposed by the framework (Determine, Explore, Choose, Identify, Decide, Evaluate). The main contribution of the research presented in this chapter is the definition of a framework that, even if it is inspired by DECIDE, is specific for PSS evaluation. It aims at supporting various stakeholders, e.g. researchers and PSS developers, in performing usability and UX evaluation of such systems, without being expert evaluators. In the PSS evaluation framework, the six activities indicated by DECIDE are specialised for the evaluation of PSS, as described in the following subsections.

3.1 Determine the Evaluation Goals

The first activity when planning a usability evaluation is to determine what the goals of the evaluation are. This helps to clarify what is the scope of the evaluation and what should be achieved once the evaluation is carried out. For example: are we going to evaluate the ease of learning of a PSS because we want to use it for a planning course at University? Are we interested in the ease of use of a PSS because it should be adopted by a planning organisation? Are we comparing the usability of similar functionality of two or more PSS? Are we evaluating a PSS to inform the design of its next version? These goals are some of the typical goals of PSS evaluations provided by the framework.

3.2 Explore the Questions

This activity includes the formulation of more specific questions that underpin the goals and should be answered through the evaluation. Indeed, the evaluation is performed just because there are questions about the system usability which should be answered. The creation of a hierarchy of questions and sub-questions allows focussing on issues that need to be addressed, in order to plan the evaluation more in detail. For example, the question "is the PSS usable?" can be decomposed in sub-questions like: Is the user interface easy to navigate? Is the terminology confusing because it is inconsistent? Is the feedback provided to users sufficient? Is the response time too slow?

In relation to the goals presented in 3.1, possible questions are: which steps are difficult to understand by students of planning courses? Can students efficiently use the PSS within a time period of two weeks? What are the skills of the planning professionals who should adopt the PSS? Which PSS provides the most efficient functionality? What could potentially be improved in the current version? Should the next version be web-based or implemented as a plugin of a Geographic Information System (GIS)?

3.3 Choose the Evaluation and Data Collection Methods

Methods for achieving the goals and answering the questions have to be chosen. Many evaluation methods are proposed in the literature (see for example, Rogers et al. 2011; Lanzilotti et al. 2011). The most commonly adopted are inspection methods and user-based methods. Inspection methods involve experts who analyse a system and provide judgments based on their expertise on usability and UX. Their great advantage is that they are cost-effective, since they do not require users nor special equipment or lab facilities, and experts can detect a wide range of problems of complex systems in a limited amount of time. A well-known inspection method is heuristic evaluation (Nielsen 1993). It involves a small set of experts who inspect the system and evaluate the interface against a list of heuristics. The main drawback of such a technique is the dependency on the inspectors' skills and experience, as heuristics are often generic and underspecified (Lanzilotti et al. 2011).

Among user-based methods, user tests allow the analysis of users' performance on the tasks which the system is designed for. They are generally considered the most complete form of evaluation, because usability is assessed through samples of real users. However, reproducing realistic situations of usage in a laboratory is difficult, for example, the selection of adequate user samples and the reproduction of ecological settings for usage in a limited amount of time (Rogers et al. 2011). A cost-effective technique in usability testing is thinking-aloud, which requires users to speak out loud their thoughts while performing tasks. This technique offers a window over the users' mental models, allowing evaluators to detect possible misconceptions about the system and the interface elements which cause them. It provides useful results even with a small number of users (Nielsen 1993).

Studies have outlined that inspection methods and user tests are complementary and can be effectively coupled to obtain a reliable evaluation process (Lanzilotti et al. 2011). Some evaluation methodologies have been proposed that rely on these methods. For example, the e-Learning Systematic Evaluation (eLSE) methodology prescribes a structured flow of activities for evaluating e-learning systems (Lanzilotti et al. 2006). Specifically, eLSE suggests coupling inspection activities and user testing, and precisely indicates how to combine them to make evaluation more reliable and still cost-effective. The PSS evaluation framework proposes a similar approach, suggesting to perform inspection first, while user testing will be performed only in specific cases, when the evaluator feels the need of a more objective evaluation that can be obtained through user involvement. In particular, the PSS framework suggests to use the thinking-aloud technique; it can be complemented with other methods involving users, namely interviews and questionnaires, which are especially valuable to assess user satisfaction and other hedonic qualities of UX.

When performing a user test, methods for collecting data have to be chosen in advance. Screen recording and measures of user performance are often adopted. Multiple data collection methods are recommended as (i) they provide rich datasets that are most valuable in rather unexplored research fields such as PSS usability evaluation, and (ii) the convergence of information from different sources can be investigated and reliability examined (Hilbert and Redmiles 2001).

3.4 Identify the Practical Issues

Many practical issues have to be considered when conducting an evaluation. The expertise of evaluators to conduct the study and the needed resources are very important. Indeed, one of the reasons why most developers neglect usability evaluation is because they feel that they do not have the right expertise and/or that too many resources are required (Ardito et al. 2014). As stated above, one of the goals of the PSS evaluation framework is to provide enough guidance to PSS stakeholders, suggesting methods appropriate for specific situations and indicating the required resources. For example, usability inspection should be performed by at least four usability experts (Nielsen 1993). If only one expert is available, more novice evaluators should be involved.

In user-based methods, an important issue is the choice of the participants, who are ideally representatives of the target user group. At least 4–5 participants should engage with the software (Nielsen 1993). PSS is considered specialised software to be used by planning professionals, with limited application to other user groups. As such, planning professionals are recommended as participants of user-based PSS evaluations. In order to find appropriate participants, a set of requirements that planning professionals have to meet should be defined. Examples of requirements might be planning professionals' level of experience in planning practice, their specialisation (for example, statutory or strategic planning), their familiarity with computer applications.

If two or more products are evaluated with user testing, a decision has to be made whether a between- or within-subjects design is adopted. In a between-subjects design, each participant is randomly assigned to each of the various experimental conditions, i.e. he/she is going to use only one product; in the within-subjects design, each participant is tested under each experimental conditions, i.e. he/she is using all products (Graziano and Raulin 2012). Other decisions about the selection of tasks that participants have to perform during the test, about facilities and equipment to be used, etc., have to be made. The PSS framework provides a description of such issues.

3.5 Decide How to Deal with the Ethical Issues

When people are involved in evaluations, ethical issues have to be addressed, in order to protect participants and their privacy. Participants have to be informed about the data that is gathered during the evaluation and how this is used. This is often done through a plain language statement. Participants are commonly also provided with a consent form that they have to sign, in order to state that they know about the procedure and to agree to participate. An example of a plain language statement and consent form is provided in the PSS framework. To warrant that the evaluation is carried out in alignment with ethical codes, the evaluation study has usually to be presented to the responsible authority where the evaluation is carried out and which has to approve it. However, how ethical issues are treated differs among institutions. As the evaluation cannot be conducted until the study has been approved, enough time should be allowed for getting ethics clearance.

3.6 Evaluate, Analyse, Interpret and Present the Data

Before actually running the evaluation, decisions have to be made about what data are needed, how they are collected, how they are analysed, how they are presented. Several books address those issues. For more details, the reader may refer to Graziano and Raulin (2012). Another delicate choice is whether qualitative data versus quantitative data have to be considered and the identification of proper methods/metrics to evaluate them. Several examples for measuring UX attributes are in Albert and Tullis (2013). It is part of future work on the framework to add specific examples of methods for collecting, analysing and presenting data.

Most importantly, it has to be defined how the quality of the collected data can be demonstrated. To this aim, important factors, such as reliability and validity, have to be considered. *Reliability* is an index of the consistency of the measures applied to the data. Good measures give consistent results, regardless of who does the measuring. A measure is not wholly reliable or unreliable, but varies in its degree of reliability. A correlation index can be used to quantify the degree of reliability. An example of correlation index is the coefficient *alpha*, although much more complicated ones exist (Graziano and Raulin 2012). *Validity* of the measure is its effectiveness in reflecting the characteristic measured. Validity, like reliability, varies in its degree. Once again, a correlation index is typically used to quantify the degree of validity.

The data collected during the evaluation have to be analysed, interpreted and properly presented. Statistics are powerful tools for analysing and understanding data. Specifically, descriptive statistics summarise, simplify and describe a large number of data. Examples of descriptive statistics are: mode, median, mean, average deviation, variance, standard deviation. They are the basis for later analyses in which inferential statistics could be used. Inferential statistics help researchers to interpret what the data mean. Examples of inferential statistics are: T-Test (it applies on independent groups, i.e., when the samples are different, as in the between-subjects design), Correlated T-Test (used in within-subjects design), Analysis of Variance or ANOVA (when more than two groups are involved), Repeated measures ANOVA (used in within-subjects design involving more than two groups).

Besides video recording, which provides further qualitative data, it is suggested to take pictures during the evaluation (e.g. of participants during an interview or a user test), because they are also useful for documentation purposes. Data are better presented if organised in tables, histograms, pie diagrams, etc., whenever it is possible.

4 Applying the PSS Evaluation Framework: A User Test

A user test was conducted with land use planners working in the metropolitan area of Melbourne (Australia) to evaluate the usability of three PSS. The PSS framework guided this usability evaluation: as a first step, it helped to make decisions about the goals of the user test, which was conducted with the thinking-aloud protocol. The goals were: (i) the analysis of PSS characteristics, in order to identify whether they are appropriate for the planners and for LSA tasks, and (ii) the identification of possible usability problems that the planners encountered while using each PSS.

4.1 PSS Evaluated

Three PSS were selected after a comprehensive review of available PSS which has resulted in the formulation of an online resource of PSS (http://docs.aurin.org.au/projects/planning-support-systems/). The following criteria were used to select the PSS:

- perform LSA through *Spatial Multi-Criteria Evaluation*;
- are transferrable to other case studies;
- offer a reasonable level of guidance, in order to not require step-by-step instruction during the user test;
- are either open-source or not too expensive.

LSA determines the suitability of each land unit (usually land parcel) for a specific purpose, based on a set of parameters that are taken into account for calculating the output. The PSS modules devoted to LSA were evaluated: they allow users to give different importance to parameters by assigning weights. This method, also known as *Spatial Multi-Criteria Evaluation*, is well-supported in decision-making processes (Arciniegas et al. 2013). The outcome includes a map that displays the suitability score for each land parcel through colour-coding. The map is also referred to as land suitability layer. The three PSS implement LSA differently, providing different weighting scales or parameter processing, as illustrated in the following.

CommunityViz is a commercial and well-known PSS (http://placeways.com/communityviz, Vonk et al. 2005). It is implemented as a desktop extension in ArcGIS (GIS software). The *suitability wizard* of Scenario 360™—one of its two main components—was evaluated. Datasets for performing LSA are added as layers (shapefile format) to ArcMap (ArcGIS component). The user has to establish the LSA parameters that he/she wants to consider. In the so-called *assumptions* window, the user assigns to each parameter a weight in the range of 0–10 through *sliderbars* (see Fig. 1). The outcome (suitability layer) is provided in ArcMap. Spatial analysis functionality available in ArcMap can be used for analysing the obtained suitability layer. Detailed help documentation is provided by CommunityViz through hyperlinks in the user interface.

Envision is a PSS implemented currently as a plugin in QuantumGIS (GIS software), primarily developed to support sustainable redevelopment of precincts (Newton and Glackin 2013). The module that performs LSA was evaluated. Datasets and parameters for effectuating LSA are stored in a database. 34 parameters, related to property, demographics and location, which the LSA can be based on, are provided. Users select the desired parameters and assign weights in the range of 0–20 to each parameter through *sliderbars* (see Fig. 2). Very few actions

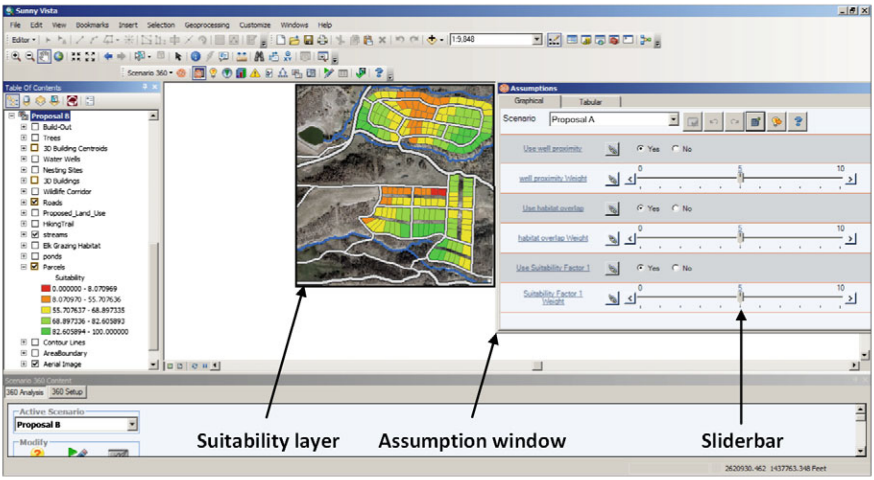

Fig. 1 CommunityViz: the assumption window including sliderbars and the suitability layer in ArcMap

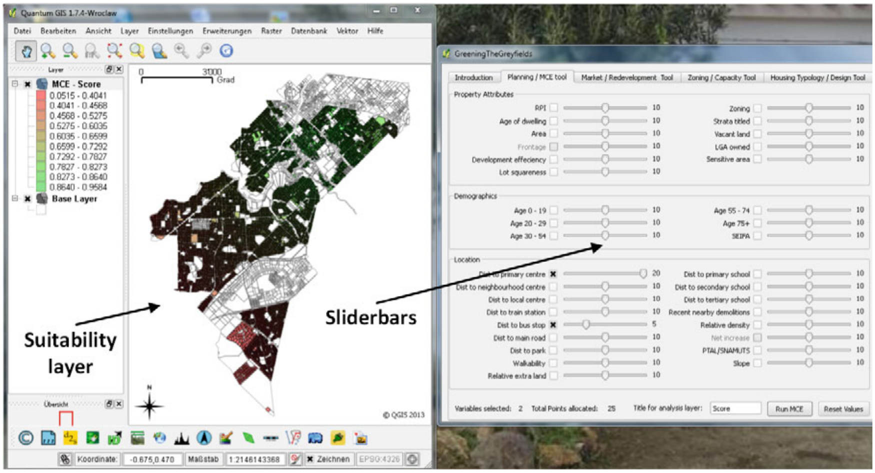

Fig. 2 Envision: the window to input parameters through sliderbars (*right*) and the suitability layer in QuantumGIS (*left*)

precede the process of selecting and weighting parameters, thus facilitating user interaction. The outcome (suitability layer) is provided in QuantumGIS. Similar to CommunityViz, spatial analysis functionality is available in QuantumGIS and can be used to analyse the suitability layer.

The *Online What if?* (OWI) (Pettit et al. 2013) is a web-based PSS version of the well-known What if? (Klosterman 1999; Vonk et al. 2005), slightly modified. The suitability analysis module was evaluated. The datasets and parameters for effectuating LSA are stored in an attribute table (shapefile format). The user has to select the desired parameters and assign weights in the range of 0–100 to each parameter through *spin boxes* (see Fig. 3). Various actions precede the process of selecting and weighting parameters. In contrast to the other two PSS, in OWI land use parcels are classified, in the resulting suitability layer, as not developable, not convertible, not suitable, or suitability as indicated in a 5-point scale from low to high. Basic navigation functions are provided such as panning and zooming for exploring the suitability layer.

4.2 Participant, Design and Procedure

The user test was carried out in a University usability lab, where proper equipment was available. It involved 6 professional land use planners (age between 25 and 45 years old, three female). Participants were chosen to meet the following criteria: (i) are strategic land use planners, (ii) are familiar with LSA and GIS operations such as layer (de)activation, map zooming and panning, and (iii) have never used the three PSS before.

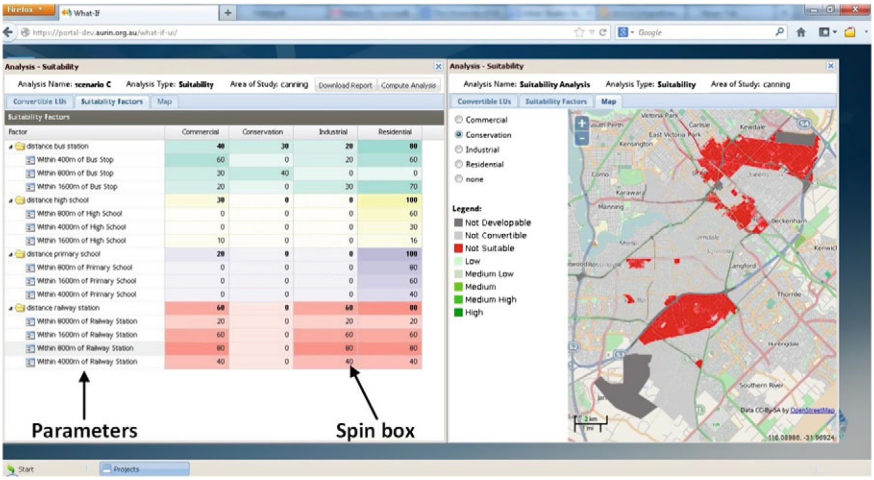

Fig. 3 Online what if?: parameters and spin boxes to assign their values (*left*) and the suitability layer (*right*)

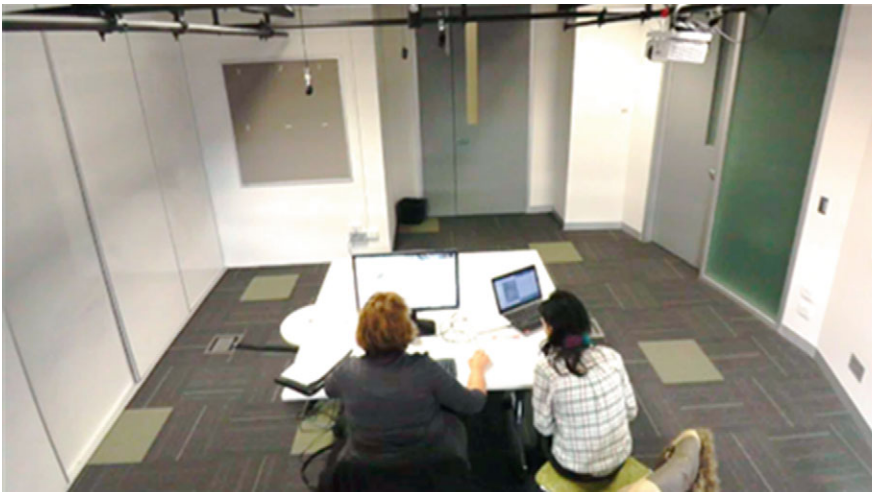

Fig. 4 A land use planner interacting with a PSS during the user test (*left*) and the experimenter (*right*)

During the test, each participant was assigned the task of identifying land most suitable for residential redevelopment by using the PSS on a desktop (see Fig. 4). For providing the same conditions, all three PSS used datasets of a case study area, namely Canning in Western Australia. The participants engaged with the three PSS in sequence, but in different order to avoid potential learning effects (Graziano and

Raulin 2012). As the participants had never used the three PSS, a short introduction was given to each tool before using it, providing information required for performing the task.

Table 1 shows the sequence of steps for each participant P1, …, P6, indicating the order in which the participant used the PSS and providing the time for each step. After engaging with each PSS, the participants were asked to complete an intermediate questionnaire, composed of 5 open questions, in which participants had to write down: (1) their experiences while engaging with the PSS; (2) the context in which these experiences occur; (3) to what extent the PSS's features and functions were satisfactory for identifying land suitable for redevelopment; (4) the three most positive aspects of the PSS used and, finally, (5) the three most negative aspects of the PSS used. At the end, a final questionnaire was completed. It was composed of 2 questions: the former was a close question that required participants to rate the diversity of the three PSS on a 5-point scale (1 = strongly disagree—5 = strongly agree); the latter was an open question which asked participants to indicate to what extent learning how to use the PSS become easier through engaging with additional PSS in the testing.

A technical problem prevented recording the first part of the interaction of P5 and made it not possible to begin with CommunityViz and go on with Envision and OWI, which was the planned design that considered all permutations of the three PSS.

4.3 Some Results and Recommendations

As suggested by the PSS evaluation framework, the user test was carried out with thinking-aloud protocol. This method was triangulated with video recording of user interaction, screen capturing including measurements of some specific actions performed by the participants, and questionnaires. The detailed results of this study are currently described in a technical report (Lanzilotti and Russo 2014) and will be published in a forthcoming paper. In the following, some findings are discussed with respect to the two evaluation goals, which are the analysis of the appropriateness of PSS characteristics for LSA tasks and the identification of usability problems encountered by planners during the test. Based on these findings, some recommendations for the design of PSS, specifically for those addressing LSA, are provided.

User performance was analysed by measuring the time spent to perform the task and by counting those actions, performed by the user, which revealed usability problems. The scheme used to code user's actions followed the one developed by (Vermeeren et al. 2002) to detect usability problems in task-based products for adults. Specifically, the actions were operationalized with three values: (1) *random actions*, referring to those actions not belonging to the correct sequence of actions to accomplish the task; (2) *puzzled actions*, concerning actions for which the user indicated either (i) not knowing how to perform the task or what action is needed

Table 1 User test procedure, indicating for each participant P1, …, P6 the order in which each PSS (CommunityViz (CViz), Envision (Env), OWI) was used and reporting the time in minutes (m) for each step

	General intro & intro to 1st PSS	1st PSS	Questionnaire on 1st PSS & intro to 2nd PSS	2nd PSS	Questionnaire on 2nd PSS & intro to 3rd PSS	3rd PSS	Questionnaire on 3rd PSS & final questionnaire	Total time
P1	(3 m)	Env (13 m)	(9 m)	CViz (13 m)	(6 m)	OWI (17 m)	(11 m)	(1 h 12 m)
P2	(4 m)	CViz (26 m)	(14 m)	OWI (23 m)	(9 m)	Env (20 m)	(16 m)	(1 h 52 m)
P3	(4 m)	OWI (27 m)	(6 m)	Env (19 m)	(6 m)	CViz (14 m)	(7 m)	(1 h 23 m)
P4	(3 m)	Env (19 m)	(9 m)	OWI (23 m)	(7 m)	CViz (15 m)	(10 m)	(1 h 26 m)
P5	n/a	Env (14 m)	(10 m)	CViz (13 m)	(7 m)	OWI (17 m)	(10 m)	(1 h 11 m)
P6	(4 m)	OWI (34 m)	(8 m)	CViz (13 m)	(8 m)	Env (14 m)	(8 m)	(1 h 29 m)

The last column shows the test total time for each Pi. For P5, the time of the first step is not available (n/a) due to a technical problem

Table 2 Performance in accomplishing the task with CommunityViz, Envision and OWI

	CommunityViz	Envision	OWI
Random actions (*N*)	3	1	14
Puzzled actions (*N*)	4	0	15
Uncomfortable actions (*N*)	2	3	9
Time (*min*)	M = 15 SD = 5	M = 16 SD = 3	M = 23 SD = 6

for it, or (ii) not being sure whether a specific action is needed or not; (3) *uncomfortable actions*, referring to actions for which the user indicated that executing the action was difficult or uncomfortable. Table 2 reports the number of the identified users' actions alongside the mean (M) and the standard deviation (SD) of task execution time for the three PSS.

During the interaction with OWI, participants carried out a total 14 random actions: they randomly clicked on interface objects and moved among different windows. Three random actions were observed during the interaction with CommunityViz and only one with Envision. Regarding to the analysis of puzzled actions, during the interaction with OWI participants appeared to be more confused than with the other two PSS, i.e. 15 puzzled actions were observed. With CommunityViz and Envision, 4 and 0 puzzled actions were identified, respectively. Examples of utterances highlighting puzzled actions were: *"I don't really understand!"* or *"How do you actually run it?"*. In many situations in which puzzled actions arose, users could not proceed without help and either asked for it or the experimenter had to intervene. Finally, during the interaction with OWI participants indicated that they are not able to execute an action, as highlighted by the 9 uncomfortable actions coded. A less number of such actions were registered for the other two PSS, i.e. 2 for CommunityViz and 3 for Envision.

The following reasons might explain these results. First of all, the three investigated PSS are different in the way they allow users to specify parameters for LSA and to assign weights to them, which also represent the most significant activities for accomplishing the task. While Envision provides in a window a predefined set of parameters among which the user can select, CommunityViz and OWI require the user to perform more steps, in order to indicate the parameters of interest. Both solutions have strengths and weaknesses. The interface should facilitate user interaction by limiting the number of actions the user has to perform. Thus, the solution proposed by Envision is easier to learn and use, because the user has only to tick off the desired parameters from the provided set. Indeed, the total number of actions highlighting usability problems in Envision is the lowest (see Table 2). However, some participants complained that this set was lacking some parameters, one participant explicitly mentioned the lack of parameters related to heritage, ecological and site contamination. The other two PSS give more freedom to the user, at the expense of more work to do, in order to indicate the parameters of interest. The user does not get proper guidance from the interface, as indicated by the higher number of random, puzzled and uncomfortable actions.

It also emerged that the layout of the assumption window in CommunityViz was not clear. Participants had difficulties in understanding the meaning of the different widgets due to their misleading position, as well as the used fonts and colours.

It was observed that the sliderbars provided by both CommunityViz and Envision was an easier mechanism for selecting the weight to be assigned to a parameter than the spin boxes used in OWI. The latter also requires more time for selecting the weight. Furthermore, users appreciated a smaller weighting range, indicating 0–10 as most preferable.

The calculation of the outcome (suitability layer and scores) is much shorter in Envision than in CommunityViz and OWI. Nonetheless, even for Envision a participant said: “it’s a bit slow” and explained that, in order to better support planning practice, efficient software is desired.

Because the participants were familiar with GIS software, they expected similar functionality when analysing the suitability layer. Indeed, three participants were surprised that, even if Envision is a plugin of QuantumGIS, this functionality was not available. Another participant appreciated the web-based PSS (i.e. OWI) because it is ready to be used and does not require the installation of any software.

Some participants used the help documentation in CommunityViz accessible through hyperlinks; however, they did not appear to be able to assimilate information in that short period of time during the user test. The help button was not selected in Envision (it was actually not working), while OWI does not currently provide any.

For all three PSS, participants remarked the need for more explanation on what the weighting actually mean, how the computation is performed and how to interpret the results. For example, participant P4 said in relation to the suitability scores in the legend: “this doesn’t mean anything to me, to be honest”, “I don’t know what the numbers mean” and “it’s not very useful, not helpful, without knowing more information”.

From the above discussion, it is evident that much more attention has to be devoted to PSS user interfaces and to the whole efficiency of the systems. In particular, the following recommendations can be derived for the design of PSS, specifically the modules addressing LSA, in order to create systems than can satisfy their users, generating a positive user experience.

- Layout, colours and fonts for the presentation of the parameters have to be carefully chosen to make it easier to understand for the user.
- Explanations about the weighting system, the calculation and interpretation of suitability scores should be provided.
- Help documentation, including examples and short demo of system use, should be available.
- The speed of PSS operations, through the selection of appropriate methods and technology optimisation, should be maximised.

5 Conclusions

This chapter presents a framework that serves as a guide for performing usability evaluation of PSS. Several stakeholders should evaluate PSS for various reasons, but they do not have the required expertise. The PSS evaluation framework has been defined to support such stakeholders, so it is expected to increase and improve PSS evaluation.

As a first application of this framework, a user test was performed for evaluating the usability of three PSS modules devoted to LSA. Even with a small number of participants (6 professional land use planners), the test highlighted some strengths and weaknesses of the three PSS with respect to user needs and expectations. Further studies with more participants and/or other methods have to be conducted.

The test results allowed the derivation of some recommendations for PSS designers. The more general recommendation is to devote more attention to the design of the system-user interaction, involving target users, i.e. planning professionals, according to user-centred approaches.

Currently, the framework is rather concerned with having PSS evaluated by participants individually. However, some PSS are supposed to be used in group settings. Future works could, therefore, involve an extension of the framework in order to adapt, where necessary, the various steps to group evaluation. As another future development, the framework could take into account the pattern-based inspection proposed in Lanzilotti et al. (2011). This technique uses evaluation patterns that indicate which are the critical aspects of the application to look at, and which actions to perform during the inspection in order to analyse such aspects. Evaluation patterns are defined to provide support primarily to novice and not professional evaluators; the rationale is that they capture the expertise of skilled evaluators (i.e., their behaviour in conducting an inspection), and express it in a precise and understandable form, so that this expertise can be reproduced, communicated, and exploited. In order to apply the pattern-based inspection to the evaluation of specific systems, a proper set of evaluation patterns has to be provided. Once patterns specific for PSS will be defined, the pattern-based inspection will be the inspection method to be used.

Acknowledgments The support of CRC for Spatial Information is acknowledged. The authors are grateful to the land use planners who participated in the user test.

References

Albert, W., & Tullis, T. (2013). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. Newnes: Morgan Kaufmann.

Allen, E. (2008). Clicking toward better outcomes: Experience with INDEX. In R. K. Brail (Ed.), *Planning support systems for cities and regions* (pp. 139–166). Cambridge, Massachusetts: Lincoln Institute of Land Policy.

Arciniegas, G., Janssen, R., & Rietveld, P. (2013). Effectiveness of collaborative map-based decision support tools: Results of an experiment. *Environmental Modelling and Software, 39*, 159–175.

Ardito, C., Buono, P., Caivano, D., Costabile, M. F., & Lanzilotti, R. (2014). Investigating and promoting UX practice in industry: An experimental study. *International Journal of Human-Computer Studies, 72*(6), 542–551.

Costabile, M. F., & Buono, P. (2013). Principles for human-centred design of IR interfaces. In A. M. Ferro, N. Former, P. Mueller, & G. Santucci (Eds.), *Information retrieval meets information visualization* (pp. 28–47). Berlin: Springer.

Geertman, S., & Stillwell, J. (2003). *Planning support systems in practice*. Berlin: Springer.

Graziano, A. M., & Raulin, M. L. (2012). *Research methods, a process of inquiry* (8th ed.). New York, USA: Pearson.

Hilbert, D. M., & Redmiles, D. F. (2001). Extracting usability information from user interface events. *ACM Computing Surveys, 32*(4), 384–421.

ISO 9126. (1998). ISO/IEC 9126-1: Information technology—Software product quality.

ISO 9241. (2010). ISO/IEC 9241-210: Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems.

Klosterman, R. E. (1999). The what if? Collaborative planning support system. *Environment and Planning B: Planning and Design, 1*(26), 393–408.

Lanzilotti, R., Costabile, M. F., Ardito, C., & De Angeli, A. (2006). eLSE methodology: A systematic approach to the e-learning systems evaluation. *Educational Technology and Society, 9*(4), 42–53.

Lanzilotti, R., De Angeli, A., Ardito, C., & Costabile, M. F. (2011). Do patterns help novice evaluators? A comparative study. *International Journal of Human-Computer Studies, 69*(1–2), 52–69.

Lanzilotti, R., & Russo, P. (2014). Applying a usability framework to evaluate planning support systems. IVU Technical Report nr. 4.2014. Dipartimento di Informatica, Universita' degli Studi di Bari Aldo Moro, Bari, Italy.

Newton, P., & Glackin, S. (2013). Using geo-spatial technologies as stakeholder engagement tools in urban planning and development. *Built Environment, 39*(4), 473–501.

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

Pettit, C. J., Klosterman, R. E., Nino-ruiz, M., Widjaja, I., Russo, P., Tomko, M., & Sinnott, R. (2013). The online what if? Planning support system. In S. Geertman, F. Toppen, & J. Stillwell (Eds.), *Planning support systems for sustainable urban development* (Vol. 195, pp. 349–362). Berlin: Springer.

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction design: Beyond human-computer*. West Sussex: Wiley.

Te Brömmelstroet, M. (2010). Equip the warrior instead of manning the equipment: Land use and transport planning support in the Netherlands. *The Journal of Transport and Land Use, 3*(1), 1–17.

Vermeeren, A., den Bouwmeester, K., Aasman, J., & de Ridder, H. (2002). DEVAN: A tool for detailed video analysis of user test data. *Behaviour and Information Technology, 21*(6), 403–423.

Vonk, G., & Geertman, S. (2008). Improving the adoption and use of planning support systems in practice. *Applied Spatial Analysis and Policy, 1*(3), 153–173.

Vonk, G., Geertman, S., & Schot, P. (2005). Bottlenecks blocking widespread usage of planning support systems. *Environment and Planning A, 37*(5), 909–924.

Vonk, G., & Ligtenberg, A. (2010). Socio-technical PSS development to improve functionality and usability—Sketch planning using a Maptable. *Landscape and Urban Planning, 94*, 166–174.