A Mixed Usability Evaluation on a Multi Criteria Group Decision Support System in Agriculture

Julian Grigera, Alejandra Garrido, Pascale Zaraté, Guy Camilleri, Alejandro Fernandez

A Mixed Usability Evaluation on a Multi Criteria Group Decision Support System in Agriculture

Julián Grigera[1,2], Alejandra Garrido[1,3], Pascale Zaraté[4], Guy Camilleri[5], Alejandro Fernández[1,2] [1]LIFIA, Facultad de Informática, UNLP, Argentina

[2]CIC, Buenos Aires, Argentina [3] CONICET, Argentina

[4]ADRIA group, IRIT, Université de Toulouse, 2 rue du Doyen Gabriel Marty, 31042 Toulouse Cedex 9, France

[5]SMAC group, IRIT, 118 route de Narbonne, 31062 Toulouse Cedex 9, France

{julian.grigera, garrido, alejandro.fernandez}@lifia.info.unlp.edu.ar, {zarate, camiller}@irit.fr

ABSTRACT

Decision Support Systems (DSS) help users make decisions by organizing the decision-making process in stages, dealing with several input sources. In the agricultural context, these tools are becoming more important given the variety of data that producers face, market regulations and interaction of several actors with different backgrounds. Different DSS have existed for years, but their adoption is reportedly very low, being poor usability one of the most relevant factors. While there are surveys indicating this, we have not found usability evaluations on DSS with concrete results. In this work, we evaluated decision-making in the agricultural domain using a group DSS called GRUS. We present 3 different usability evaluations: a heuristic evaluation, a user test, and an automated diagnosis.

KEYWORDS

Usability evaluation methods, Agriculture, DSS, GDSS

1 INTRODUCTION

Decision support systems help users make effective decisions by organizing the decision-making process in clear stages, dealing with several kinds of input sources. Specifically, in the context of agriculture, this kind of tools are becoming increasingly important given the variety of data considered by producers, new regulations or changing weather conditions.

https://doi.org/10.1145/3233824.3233852

The decision-making process in the agricultural context is especially complex since it involves interaction of several actors with different backgrounds, facing big amounts of data to respond faster to increasingly demanding clients. Decisions made by one or several stakeholders are generally managed on several criteria.

In this context, a tool for enabling Cooperative Decision Processes is essential to better communicate the information and arrive to a coordinated solution. GRUS is a highly customizable facilitation platform, focused on collaborative, multi-criteria decision making. It supports decision making in any domain (like agriculture) and using a variety of collaborative methods.

The complexity of Cooperative Decision Processes requires IT support where decision makers with different skills can have a good user experience. The literature on the field shows that, while there are many DSS available, their uptake in the agricultural field is extremely low [13], one of the main reasons being poor usability. However, we have not found usability evaluations in this context, rather surveys' reports with collective opinions.

The contribution of this work is the report of a mixed usability evaluation of GRUS in the agriculture value chain. This system is general enough to run decision-making sessions on any subject, and provides support for collaboration, thus, it was a representative artifact to evaluate. We ran three different usability evaluation methods: a heuristic evaluation, 3 user tests, and an automated test. For the user tests in particular, we focused on three specific agriculture scenarios for the decision-making sessions, using the tomato production in farm cooperatives as a specific context. To the best of our knowledge, this is the first usability report on this kind of tool applied to agriculture.

2 RELATED WORK

Most previous research on usability of decision support systems (DSS) is focused on explaining their low adoption. Usability is usually studied together with other factors considered to influence this phenomenon, such as trust issues or high maintainability.

Early research in DSS for agriculture [13] shows that the adoption was low at that time, blaming poor usability amongst other factors, in the overly complex interfaces, cluttered with irrelevant information, and confusing terminology. They also

point out the unnecessary amount of input data. Considering a much more recent work by Rose et al. [14], the authors still consider the uptake of evidence-based DSSs in agriculture to be disappointingly low. Based on surveys conducted on 244 farmers and advisors in the UK, the authors present a list of 15 influential factors affecting the adoption and use of DSSs, like usability, performance and relevance to user. They propose that these factors should be considered when designing new DSSs. In the paper, authors argue that *ease of use* is one of the most influential factors that farmers/advisors consider before adopting a DSS.

In a review from 2012 by Rossi et al. [15], the authors also recognize that one of the largest limitations of DSSs is poor usability ("not user-friendly interfaces"). In the agricultural context, the problem shows especially in cultivation decisions, which are particularly complex. The authors indicate that the amount of redundant information can be a decisive factor.

Other works in the area, also considering the low adoption of DSSs in agriculture, propose the involvement of the users in the development process [10–12]. Nurkka et al. popose user-centered design in the field of Precision Agriculture (PA), modelling farmers' activities with the Core-Task Analysis approach, using the production of malt barley as a PA case.

While the aforementioned works present many interesting insights on DSS usability, and even large data recollection from real farmers, none of them present actual results on usability evaluations, which is the focus of our study.

3 BACKGROUND ON GROUP DECISION MAKING

In large organizations, most decisions are taken after extensive consultations with many participants, rather than by individual decision makers [4]. According to Smoliar and Sprague [16], decision making in organizations usually involves interaction of several actors. This includes information communication, but its main aim is to enable decision makers to come to a shared understanding, assisting them at achieving a coordinated solution to the problem at hand. The process of group decision making has been analyzed from different perspectives. Recently, Zaraté [18] suggested that the increasing complexity of organizations, and the use of Information and Communication Technologies to support them, require decision processes to be modified.

To support a group engaged in decision making, Macharis et al. [8] introduced a methodology based on the Multiple Criteria paradigm through the PROMETHEE methodology. They propose that each decision maker create their own performance matrix by determining their own individual values. Then a global evaluation of each alternative is performed using a weighted sum aggregation technique. Decision makers' weights may differ. One benefit of this structure is the ability to conduct a stakeholder-level sensitivity analysis. Nevertheless, the proposed system doesn't allow decision makers to share their preferences, or to co-build a decision. In a collective decision framework, decision makers must balance their own attitudes and preferences with the goal of building common preferences and consensus within the group.

Generally, decisions made by one individual or by several stakeholders are managed on several criteria. For example, anyone who wants to buy a car will evaluate every alternative on many criteria, like price or motor. Decision makers then obtain a preferences matrix. The idea is to produce a ranking of alternatives using Multi-Criteria Decision Analysis (MCDA) methodologies. In a real-world setting, stakeholders apply both shared criteria (known to the group, even if not everyone agrees on its relevance), and private ones are (known only to the stakeholder). Using shared criteria is not always possible. A previous experiment [19] showed that users value having both shared and private criteria.

The GRoUp Support (GRUS) system developed at IRIT [2] is a multi-criteria Group Decision Support System. This system is conceived as a toolbox based on the Web. Decision making in GRUS is modeled as facilitated meetings (which can be as long lived as needed) that consist of a series of connected activities. Each meeting follows a **process**, i.e. a template that determines the activities' workflow. The process is dynamically adaptive and two kinds of roles are defined: the decision maker who is part of the decision process and the facilitator who has for objective to support the decision-making process. GRUS provides a selection of well tested processes that the facilitator of a meeting can choose. Moreover, the facilitator can design an ad-hoc process for a meeting. Figure 1 shows a screenshot of the process to decide on a car model, with the activities at the top, highlighting the current activity that consists in multi-criteria evaluation. With the help of GRUS, a facilitator guides the participants trough the activities.

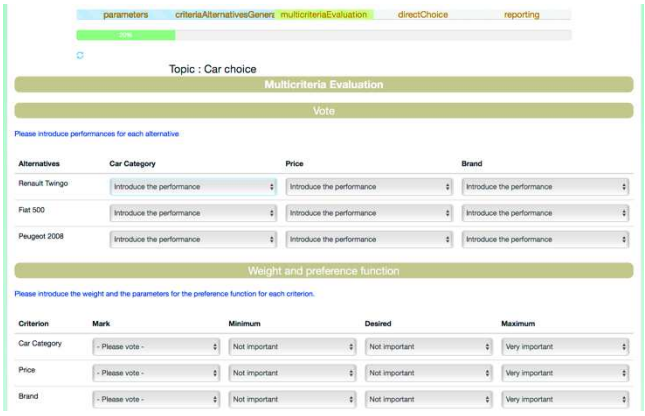

Figure 1. GRUS screenshot.

4 USABILITY EVALUATION

Developing usable websites is essential for their success [7]. If websites are not constructed considering usability requirements, as User Centered Design [1] proposes, usability may still be measured and improved afterwards. Fernandez et al. report on different usability evaluation methods (UEMs) for the web [3], many of which are applied after the system was deployed with the purpose of finding and fixing usability problems on existing web interfaces. According to these studies, in this work we have used 3 types of UEMs:

• **Heuristic evaluation:** this is an inspection method ran by a usability expert, who identifies heuristic violations in web artifacts

• **User testing:** the most common type of empirical method, where the evaluation is performed with volunteers working on predefined tasks. In this case, we ran a moderated test, with an expert guiding through the tasks and making observations.

• **Automated testing:** also called performance measurement by Fernandez et al., where a tool records usage data and analyzes it automatically to discover problems from usage statistics. We have constructed a tool called USF (Usability Smell Finder) [5] that captures user interaction events, analyzes the events from several users on the same interface element, and reports on possible usability problems (called usability smells) on the user interaction.

This selection was made for maximizing the output of problems with the available resources. User testing is a very effective method for finding usability problems, since is involves real users performing real tasks. The heuristic evaluation was planned as a complement since it goes beyond the planned tasks for the user tests. Finally, the automated testing, even if may contain false positives, was considered as the exhaustive alternative, potentially picking up problems that experts are not aware of. We next describe the results of applying these 3 UEMs on the GRUS system. Detected issues are listed in Table 1 (Heuristic, User Testing and Automated testing are marked as "H" "U" and "A" respectively).

4.1 Heuristic evaluation

The heuristic evaluation was performed by one expert, following Nielsen's template that defines 10 usability heuristics [9]. During the evaluation, the most relevant feature of the system, i.e. the shared meetings, was analyzed against the heuristics, creating test users and meeting, analyzing the registration process itself, etc.

As a result, a total of 11 issues were found: 6 considered as specific to GRUS, and the other 5 as more general issues that may also affect other DSS, especially for group decision-making. Amongst the GRUS-specific problems, there were form validation issues detected, especially during the sequential steps of the meetings (issue #12 in the table, mostly related to *Heuristic 5: Error prevention*) and confusing terminology in process steps and some validation buttons (issue #10, regarding *Heuristic 4: Consistency and standards*). Amongst the general DSS issues, most had to do with collaboration, like difficulties in showing the status of the other participants or failing to correctly cue the next action (issue #9 / *Heuristic 1: Visibility of system status*). There was also an issue indicating excessive detail of information on the final report (issue #8 / *Heuristic 8: Aesthetic and minimalist design*).

4.2 User Testing

User tests were run by using GRUS to make group decisions for in the context of a cooperative farming organization. In all cases there was a group of 4 to 5 participants acting as farmers. In one of them, all participants were together, and in the 2 remaining cases some were online, connected by a conference call. All sessions were run by a facilitator in charge of coordinating the process' steps.

Table 1: Detected Usability Issues

#	Issue description	H	U	A
1	Confusing Icons for Task Managing	✓		✓
2	Unresponsive "Next Task" Icon			✓
3	No alternatives edition	✓	✓	✓
4	Confusing Titles Hierarchy			✓
5	Static Report			✓
6	Time as a Progress Bar	✓		✓
7	Confusing "Join Meeting" functionality			✓
8	Overloaded Report	✓	✓	✓
9	Unclear current meeting's status	✓	✓	
10	Redundant controls for "next step"	✓	✓	
11	Complex GUI for Multi-Criteria features	✓	✓	
12	No return from edition to current process	✓		
13	Lack of inline validation at some steps	✓		
14	Missing inline help features	✓		
15	Confusing controls' terminology	✓		

4.2.1 Decision-making scenarios. In this work we use term *scenario* to describe a day-in-the-life-of situation in the use of a software system (in our case a group DSS). Scenarios are widely used in during software analysis to provide descriptions of processes (current or future) including actions and interactions between the users and the system [17].

From the domain perspective, the three scenarios will focus on one decision that we identified as difficult, of large impact (large gain or large loss), highly dependent on collaboration and information sharing, and still unsolved. Collaboratively decide on the probability of occurrence among various alternative futures.

The green belt of La Plata city is a farming area of approximately 6.000 hectares. According multiple meetings with farmers, agronomists and experts, a recurring problem was the difficulty to assess the plans of the farming community as a whole and consequently adapt one's own farming plan, given the lack of information. Consequently, when farmers make a decision that depends on the projected production of a given crop, they resort to intuition and talks with colleagues.

The tests consisted on a decision on a scenario regarding the tomato production in the green belt of La Plata city. At the onset of a tomato production season there are multiple important decisions the farmer faces. Some reflect the market strategy, e.g. when to start, which variety to plant, and how much. There are also decisions on the style of work, like planting density, training system, conduction style and pruning of stems and trusses.

The first scenario was about deciding the most appropriate variety of tomato seed to plant for the organization. The decision process started as one with medium complexity (no clustering) but was edited during the session to include clustering, turning into a high complexity process. There were 4 alternatives proposed with 3 different criteria. The second scenario focused on defining the best date to plant regarding weather conditions. The decision process used was a complete one. There were 3 alternatives and 3 criteria clusters. The third scenario focused on deciding about conduction style (how many main stems each tomato plant will have) and pruning of stems and trusses. This was set in a context of 4 greenhouses, each led by a greenhouse leader. Options were letting the plant grow freely, leaving a single branch or many, using the same strategy for all greenhouses, or not. The possible criteria were quality, yield, maintenance labor, required resources.

4.2.2 Usability evaluation results. During all sessions, we observed the participants' behavior and obtained usability issues.

We next show a consolidated list of the most significant problems, mainly related to general Multi-Criteria Group Decision-Making situations. The most relevant problem detected had to do with informing the status of the multi-step process. Users were often confused about whether they should wait for the rest of the participants or move ahead. The system does have controls for indicating one's finished but they were not enough for reassuring the decision makers, so they have to resort to verbal communication. There were also problems inherent to multi-criteria decision theory, like understanding private criteria influence on the final decision, or criteria weight functions, which are difficult to present in general to stakeholders that are not experts in multi-criteria group decision making. At the end of each session, the report included details of the process but it was difficult for the participants to find the decision's outcome, which was the most relevant part. There were also problems related with lack of editing possibilities (for criteria and alternative proposal). While updating the decision-making process steps, the facilitator run into some trouble, inherent to the complexity of process itself. Other issues specific to the GRUS GUI were detected, similar to those reported in the first part of section 4.1.

4.3 Automated Testing

The third usability method consisted is using an automated critique agent that processes interaction logs to find usability problems, called Kobold [6]. Based on the interaction it tracked from the user tests, the tool reported 8 usability issues, expressed as usability *smells* in the report but adapted to the current context, out of which the 6 most relevant ones where overlapped with the user test and heuristic analysis (2 and 4 respectively). For instance, it detected unresponsive elements (issue #3), by logging repeated clicks with no consequent action, and by logging erratic mouse movement it signaled some of the elements like the progress bar and the task managing icons (issues #1 and #6). Amongst the issues exclusively found by Kobold were confusing titles hierarchies (issue #4), unresponsive buttons (#2) and confusing controls for joining a meeting (#7), specific to GRUS.

5 CONCLUSIONS

Analyzing the full list of detected usability issues, we found out that most problems are related with the participants' lack of awareness on the ongoing multi-stepped process, an inherently difficult problem to tackle from the GUI perspective, but very relevant, being one of the main features of this kind of decision-making sessions. The second most repeated issue, detected in all 3 UEMs, was the overloaded report, which is also consistent with the findings of other authors on the field. This is also very relevant, since the decision-makers need a clear response on the decision taken. All details on the meeting and calculations may be important, but secondary to the main output, i.e. the final decision.

After running the tests, we consider that combining the 3 selected methods was a favorable approach. On one hand, finding the same issue across different UEMs was reassuring, but also the complement in the variety of errors allowed us to get a big picture of the general inconveniences of these systems.

According to the existing literature, the lack of adoption of DSSs in the agricultural field is still a pressing issue and, being usability one of the most relevant problems reported, improvements in this aspect with respect to the state of the art could benefit their uptake. The findings in this study intend to give a concrete report of usability problems that, even if they were found on a specific software (GRUS), they could be helpful for other tools, especially for collaborative decision making in the agricultural field.

We are planning two different extensions for this work: on one hand, we intend to solve the found issues on GRUS and run a follow-up test to verify the improvements. On the other hand, we plan to run similar usability tests on a different group DSS, to assess whether the usability issues repeat themselves.

ACKNOWLEDGMENTS

Authors of this publication acknowledge the contribution of the Project 691249, RUC-APS: Enhancing and implementing Knowledge based ICT solutions within high Risk and Uncertain Conditions for Agriculture Production Systems (www.ruc-aps.eu), funded by the European Union under their funding scheme H2020-MSCA-RISE-2015.

REFERENCES

[1] Abras, C., Maloney-krichmar, D. and Preece, J. 2004. User-Centered Design. *Berkshire Encyclopedia of Human Computer Interaction*. 1–14.

[2] Camilleri, G. and Zaraté, P. 2014. *EasyMeeting: a group decision support system (Release 1)*.

[3] Fernandez, A., Insfran, E., Abrahão, S. and Abrahao, S. 2011. Usability evaluation methods for the web: {A} systematic mapping study. *Information and Software Technology*. 53, 8 (2011), 789–817. DOI:https://doi.org/10.1016/j.infsof.2011.02.007.

[4] Gorry, G. and Morton, M.S. 1971. A framework for management information systems. (1971).

[5] Grigera, J., Garrido, A., Rivero, J.M. and Rossi, G. 2017. Automatic detection of usability smells in web applications. *International Journal of Human Computer Studies*. 97, (2017). DOI:https://doi.org/10.1016/j.ijhcs.2016.09.009.

[6] Grigera, J., Garrido, A. and Rossi, G. 2017. Kobold: Web Usability as a Service. *Automated Software Engineering - Tool Demonstrations*. (2017), to appear.

[7] Lee, Y. and Kozar, K.A. 2012. Understanding of website usability: Specifying and measuring constructs and their relationships. *Decision Support Systems*. 52, 2 (2012), 450–463. DOI:https://doi.org/10.1016/j.dss.2011.10.004.

[8] Macharis, C. and Brans, J.-P. 1998. The GDSS PROMETHEE procedure. *Journal of Decision Systems*.

[9] Nielsen, J. 1995. 10 Usability Heuristics for User Interface Design. *Nielsen Norman Group*.

[10] Nurkka, P., Norros, L. and Pesonen, L. 2007. Improving usability of and user acceptance of ICT systems in farming. *Paper presented at the EFITA/WCCA Joint Conngress in IT in Agriculture, Glasgow*. (2007).

[11] Oliver, D.M., Bartie, P.J., Heathwaite, A.L., Pschetz, L. and Quilliam, R.S. 2017. Design of a decision support tool for visualising E-coil risk on agricultural land using a stakeholder-driven approach. *LAND USE POLICY*. 66, April (2017), 227–234. DOI:https://doi.org/10.1016/j.landusepol.2017.05.005.

[12] Parker, C. 1999. A user-centred design method for agricultural DSS. *EFITA-99: Proceedings of the Second* (1999).

[13] Parker, C.G. and Campion, S. Improving the uptake of decision support systems in agriculture. *Proceedings First European Conference for Information Technology in Agriculture* 129–34.

[14] Rose, D.C., Sutherland, W.J., Parker, C., Lobley, M., Winter, M., Morris, C., Twining, S., Ffoulkes, C., Amano, T. and Dicks, L. V. 2016. Decision support tools for agriculture: Towards effective design and delivery. *Agricultural Systems*. 149, (2016), 165–174. DOI:https://doi.org/10.1016/j.agsy.2016.09.009.

[15] Rossi, V., Caffi, T. and Salinari, F. 2012. Helping farmers face the increasing complexity of decision-making for crop protection. *Phytopathologia Mediterranea*.

[16] Smoliar, S. and Sprague, R. 2002. Communication and understanding for decision support. *Proceedings of the International Conference IFIP TC8/WG8* (2002).

[17] Wohlin, C. 2005. *Engineering and managing software requirements*.

[18] Zaraté, P. (Pascale) 2013. *Tools for collaborative decision-making*. ISTE.

[19] Zaraté, P., Kilgour, M. and Hippel, K. 2016. Private or Common Criteria in a Multi-criteria Group Decision Support System: An Experiment. *International Conference on Collaboration Technologies (CRIWG 2016), Kanazawa, Japan, Vol. 9848* (Kanazawa, Japan, 2016), 1–12.