A Chatbot Solution for eGLU-Box Pro: The Usability Evaluation Platform for Italian Public Administrations

Stefano Federici[1][(✉)] , Maria Laura Mele[1][(✉)] , Marco Bracalenti[1][(✉)],
Maria Laura De Filippis[1][(✉)] , Rosa Lanzilotti[2][(✉)] , Giuseppe Desolda[2][(✉)] ,
Simone Borsci[3][(✉)] , Giancarlo Gaudino[4][(✉)], Antonello Cocco[4][(✉)],
Massimo Amendola[4][(✉)], and Emilio Simonetti[5][(✉)]

[1] Department of Philosophy, Social and Human Sciences and Education University of Perugia, Perugia, Italy
{stefano.federici,marialaura.mele,marco.bracalenti, marialaura.defilippis}@unipg.it
[2] Department of Computer Science, University of Bari Aldo Moro, Bari, Italy
{rosa.lanzilotti,giuseppe.desolda}@uniba.it
[3] Department of Learning, Data analysis, and Technology – Cognition, Data and Education – CODE group, Faculty of BMS, University of Twente, Enschede, The Netherlands
[4] DGTCSI-ISCTI – Directorate General for Management and Information and Communications Technology, Superior Institute of Communication and Information Technologies, Ministry of Economic Development, Rome, Italy
{giancarlo.gaudino,antonello.cocco,massimo.amendola}@mise.gov.it
[5] Department of Public Service, Prime Minister's Office, Rome, Italy
e.simonetti@funzionepubblica.it

Abstract. This paper shows a chatbot solution for eGLU-box Pro, a usability testing platform for Italian Public Administration (PA). eGLU-box Pro is a web-based tool designed to help PA practitioners in creating remote usability tests and analyzing participants' answers and interaction data after they complete the usability tasks. The impact of the chatbot solution on users' experience was assessed by bio-behavioral evaluation methods such as eye tracking, electroencephalography, and facial expression recognition. This work describes the platform and its integrated chatbot solution and shows the results of a preliminary laboratory study involving 20 end-users. The study is part of an ongoing design and development project based on a user-centered approach.

Keywords: Chatbots · Usability testing tools · User experience evaluation · Psychophysics

1 Introduction

Chatbots are intelligent conversational agents that interact with users by textual dialogues written in a natural language [1]. The use of chatbots aims at supporting interpersonal services, decision-making processes, and training in different contexts [2–5].

A recent review by Abd-Alrazaq and colleagues [6] highlighted that chatbots are usually assessed only in randomized controlled trials. The literature also shows that chatbot interaction efficiency is rarely assessed and, when assessments are performed, they are usually done on qualitative or non-standardized aspects of the interaction, thus providing data that cannot be statistically compared [1, 7, 8]. Because chatbots are interaction systems, their design should include users from the earliest stages by following a user-centered approach rather than a system-centered approach [1].

This study describes the user-centered design of a chatbot component of eGLU-box Pro, which is a digital web platform designed to support Italian Public Administration (PA) in conducting usability evaluation studies. The platform is based on the eGLU 2.1 protocol, which has been recently developed by the Working Group for the usability of the PA (https://www.agid.gov.it). The latest released version of the platform is eGLU-box PA v. 2018.1, which is now an integral part of the design guidelines for PA web services promoted by the Agency for Digital Italy (AGID) and by the Digital Team. The previous versions of eGLU-box have been validated in several studies [9–11] based on heuristic assessments and both user experience (UX) and psychophysical usability evaluations with PA practitioners in workplace conditions, web end-users in experimental laboratory conditions, and web end-users in remote online conditions.

This paper describes a new version of eGLU-box, called eGLU-box Pro, which allows for the improvement of the functionality of the platform, achieved through the development of an interface for the virtual management of remote tests (chatbot). In this study, the UX evaluation of the eGLU-box Pro chatbot was conducted by comparing the bio-behavioral components of the user experience of interaction derived from a usability test on a website. This study uses psychophysiological measures because they reflect top-down factors of human–computer interaction such as interest, attention, active participation, and mood. Eye-tracking technologies, facial expression recognition algorithms, and electroencephalography (EEG) can accurately measure the psychophysiological components of users' experience while interacting with a chatbot.

The paper is organized as follows. Section 2 describes the user-centered technology. Section 3 describes in detail the experimental methodology used for the assessment of the chatbot solution for eGLU-box Pro. Sections 4 and 5 present and discuss the results of the experimental assessment. Section 6 summarizes the findings.

2   eGLU-Box Pro

eGLU-box Pro is a web platform designed and developed during two previous research projects, funded and coordinated by the Superior Institute of Communication and Information Technologies (ISCTI) and the scientific and technical body of the Italian Ministry of Economic Development (MISE). The main goal is to improve the usability of Italian PA websites by involving the users of these websites in the evaluation, with a low-cost methodology (eGLU-LG 2.1 protocol) and an ad hoc technology (eGLU-box Pro). Another important goal is to increase the awareness of those who work in PA, from employees to managers, on the importance of usability, a fundamental element to be able to encourage citizens in using PA digital services.

2.1   The Platform

eGLU-BOX Pro supports the execution of remote asynchronous usability tests [12] according to the eGLU-LG 2.1 protocol. Two types of users are involved in remote usability tests with eGLU-box Pro, namely the evaluators and the participants. Different activities are supported by eGLU-box Pro according to such user types.

- *Evaluators* can create usability tests and visualize the analysis of the data automatically performed by the platform. When creating a test, evaluators have to define: the welcome script, the list of tasks to be performed by participants (specifying the URL where the task starts and the URL where the task is considered completed), possible questionnaires to administer to participants after the tasks have been executed (choosing between SUS [System Usability Scale], UMUX-Lite [Usability Metric for User Experience]; NPS [Net Promoter Score], and custom; [13, 14]), the data to be automatically collected during the user interaction with the website under evaluation (i.e., desktop recording and video and/or audio recording), and the participants to be invited.
  eGLU-box Pro automatically stores and analyzes all the collected data, showing in an aggregated form all the results of the performed test. Specifically, it shows the time and the success of each task and of each user as well as the task success of all tasks and all participants, the average score of each administered questionnaire, and a summary report exportable as a PDF. In addition, eGLU-box Pro allows the evaluator to analyze and annotate the video and audio of each user and each task.
- *Participants*, on the other hand, are guided in carrying out the usability test they are invited to complete. The platform, through a wizard process, first tests the peripherals that may be necessary to record the data selected by the evaluator (e.g., webcam and microphone), and then administers one after the other the tasks the participant has to carry out and, finally, the questionnaires.

This platform has been developed as a web app; thus, it does not require the installation of plug-ins or specific software because it has been realized through the use of solid web programming tools such as the PHP Laravel 5.6 framework, together with the use of HTML 5, Bootstrap 4, CSS 3, and jQuery. eGLU-box Pro has been certified according to ISO/IEC 25010:2005 "Software Engineering – Software product Quality Requirements and Evaluation (SQuaRE)".

2.2   The Chatbot

eGLU-box Pro has been used during more than 1,000 usability tests, allowing us to gather important feedback on different aspects. One of the most interesting results regards the absence of a real conductor that typically guides participants during an in-lab user test. We observed that participants, acting alone during a remote test, sometimes experience difficulties. For example, they need clarifications on aspects such as the use of audio/video registrations or need to solve some technical problems (e.g., peripherals not working). It emerged that, in a large number of cases, this negatively affects the user test results [paper under review]. For this reason, to compensate for the absence of a

real conductor during the remote user test with eGLU-box Pro, we designed a chatbot that emulates all possible help and support that a conductor typically provides during an in-lab synchronous user test as a response to participants' requests.

The chatbot was implemented using Google Dialogflow. The chatbot is always visible and active from the moment the participant logs into the platform homepage until the end of the user test. The chatbot assumes three different faces, one for each state it may be in: inactive (Fig. 1, left), meaning that the user has closed the chatbot; active (Fig. 1, center), meaning that the user is interacting with the chatbot; and alert (Fig. 1, right), when the chat window is closed but the chatbot wants to communicate something to the participant.

Fig. 1. Expressions of the chatbot.

Users can also interact with the chatbot by means of specific buttons: increase/decrease font size; reset the chat history; enable or disable voice reading because the audio may be annoying; send the message, avoiding pressing "Enter" on the keyboard every time; and minimize the chatbot.

The design of the chatbot conversation follows the design patterns proposed by the Natural Conversation Framework, inspired by the natural way people speak [15]. These design patterns are specific to the creation of conversational agents such as chatbots and help to structure the dialogue in naturalistic conversational sequences between a chatbot and a real user through four components:

1. an interaction model based on expandable sequences;
2. a format based on the interaction model;
3. reusable templates for common speaking activities; and
4. a method of navigation by conversation.

The resulting chatbot interacts with the test participant in the following way (see Fig. 2). After logging into the platform, the chatbot welcomes the participant with a short message. The chatbot also explains that to start a test the participant must click the "Start" button next to the name of the study. The welcome message ends by explaining that by saying or typing the word "Help," the test participant will be supported with additional information. Help can be requested during any time in the user test and allows the participant to choose from a menu displayed by the chatbot to have more support on:

- How to run the task
- How to start a test
- What is a user test

- How to move to the next task
- What is the purpose of eGLU-box Pro
- What the recordings will be used for
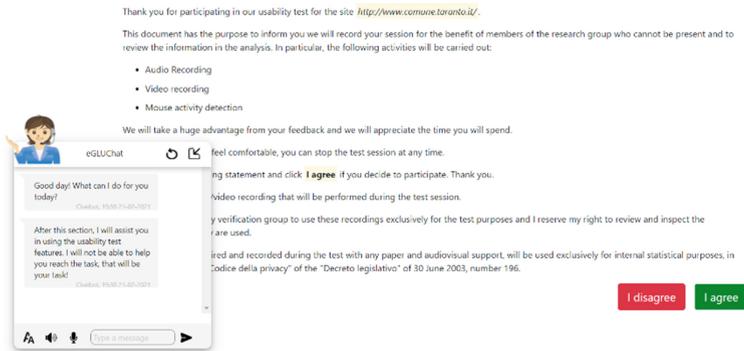- How to abandon a test
- How to fill out the questionnaires

Fig. 2. Chatbot integrated in eGLU-box Pro.

Once the participant has selected a specific user test, before starting the task execution he/she has to test the peripherals that the evaluator has selected for data capture, namely webcam, microphone, and desktop recording. For each peripheral, if the check is successful, the chatbot informs the participant that the test has been correctly executed, and vice versa it provides the participant with useful information to solve technical problems.

After the peripheral check, the test starts. In particular, all the tasks are administered in sequence. At the beginning of each task, the chatbot, simulating the behavior of an in-person conductor, tells the participant that he or she has to execute a task. In addition, the chatbot announces that the test is about to end with the evaluation by administering questionnaires that the user must complete. For each questionnaire administered by eGLU-box Pro, the chatbot briefly describes what the user is about to fill out, in terms of the name of the questionnaire (e.g., SUS), number of questions (10), and the type of answers they will be able to give (choice of 1 to 5). When the last questionnaire is completed, the chatbot thanks and dismisses the participant.

3   Method

This work is part of an ongoing study to assess eGLU-box Pro. The user-centered design of the chatbot integrated in eGLU-box Pro is here described as the first part of a multi-step process. This study evaluates the performance of the first version of the chatbot. A laboratory study has been conducted as described in Sect. 3.2. to investigate with psychophysics whether providing the automatic usability assessment tool with a chatbot affects the user experience.

Eye tracking methodology usually combines near-infrared technology with high-resolution optical sensors to measure where an observer is looking based on the pupil center corneal reflection (PCCR). The two main dynamics of eye movements, namely saccades and fixations, reveal the underlying cognitive processes of interaction, such as reasoning, problem solving, or attention [16].

Facial expression recognition uses patterns of involuntary facial muscular movements with algorithms to measure affective valence and seven universal basic emotions (joy, sadness, anger, contempt, disgust, fear, and surprise) as proposed by Paul Ekman and coded by the Facial Action Coding System (FACS) [17].

Frontal EEG alpha asymmetry is related to processes of approach/withdraw from affective or unexpected stimuli [18]. EEG measures the electrical field produced by the neural activity. EEG rhythms can reveal several components of cognition. In human–computer interaction research, one of the most studied rhythms is related to alpha activity. A decrease or desynchronization in alpha waves happens when cortical activity increases [19]. In particular, alpha frontal asymmetry may reveal approach or withdrawal cognitive processes [20]. An alpha lateralization to the left frontal lobe is related to approach, and a right lateralization appears as a withdrawal response from unexpected stimuli [21].

3.1 Apparatus and Measures

The version of eGLU-box Pro evaluated in this study runs on a PC with an Intel® Core™ i7-4710MQ CPU @ 2.50 GHz/2.49 GHz, RAM 8 GB, and 2880 × 1620 screen resolution. The experimental assessment was carried out using iMotions (www.imotions.com), a research platform that synchronizes biometric sensors in a single platform for psychophysical experiments. In this study, an eye tracker, a facial expression recognition software, and EGG technology were used.

The eye tracker model was a Tobii X2-30 Eye Tracker Compact Edition, with a 30 Hz gaze sample rate and 0.4° accuracy. In this study, the fixation number and duration (in milliseconds) were calculated.

The Affectiva Affdex technology was used to measure facial expressions to determine emotional valence (-100 = negative valence, 100 = positive valence) and seven emotions (0 = absent, 100 = present) with a 20% confidence threshold.

An EEG Emotiv Epoc+ was used. This wireless EEG headset records 16 channels based on the international 10–20 system at 2048 Hz internal, filtered, and downsampled to 128 or 256 per second per channel. The EEG device was used to calculate the frontal alpha asymmetry index by the difference between the alpha EEG power in the frontal right electrode and the alpha EEG power in the frontal left electrode [22].

3.2 Procedure

A usability test of an Italian PA website (http://www.sviluppoeconomico.gov.it) was performed. The test was administered through either the eGLU-box Pro tool with the integrated chatbot (experimental condition) or the eGLU-box tool without the integrated chatbot (control condition).

The experimental sessions were carried out in line with the COVID-19 prevention procedures recommended by the World Health Organization (https://www.who.int/eme

rgencies/diseases/novel-coronavirus-2019). Before entering the laboratory, each participant was asked to read and sign an informed consent document that asked epidemiological and clinical questions aimed at self-certifying that each participant had no symptoms and no risk of having come into contact with people that were COVID-19 positive within the last 14 days.

The test comprised four tasks showing four user scenarios asking participants to search for specific information, which is shown in a certain landing page that participants do not know before completing the task. Participants are free to follow the path they want to search for the required information. Each task lasts a maximum of 5 min, an average time beyond which users generally lose interest in the interaction. After 5 min from the beginning of each task, the eGLU-BOX platform automatically stops the current task and shows a new task or the ending pages of the test.

Both experimental and control conditions were monitored using the iMotions platform. The test was conducted under an artificial constant dim light. Participants were seated at about 600 mm from the screen. A 9-point eye-tracking calibration and EEG benchmark data acquisition after the EEG electrode impedance assessment were conducted.

3.3   Participants

An experimental test with 20 participants was conducted (10 control condition participants, 10 experimental condition participants). Participants (50% female, mean age 23.55 years old) declared they had normal or corrected-to-normal vision, no color blindness, and no professional experience in usability studies. Two outliers per condition were removed by the eye tracking dataset. One outlier was removed from the experimental condition of the EEG dataset.

4   Results

The results show the data on eye tracking, facial expression, and EEG measured in both conditions during the test.

4.1   Facial Expressions

The participants' emotions were computed by analyzing the webcam videos with the Affectiva software. For each frame, the analyzed emotions were joy, anger, disgust, surprise, fear, sadness, and contempt (from 0 to 100, where 100 indicates the maximum level of emotion), as well as valence (from -100 to $+100$, where -100 indicate negative emotion and $+100$ positive emotion). Affectiva produced 1,549,481 data points associated with emotions.

The mean value of basic emotions and valence measured during the four tasks was calculated for both control and experimental conditions (Table 1). All the emotion statistics have been calculated by removing frames whose values were <5, because emotions under this threshold were considered noise.

Table 1. The mean (M) and standard deviation (SD) and *t*-test comparisons for basic emotions and valence in the control (condition 0) and experimental (condition 1) groups.

	Condition	M	SD	Welch's *t*-test
Engagement	0	47.687	47.602	$t = 0.233, p > .05$
	1	47.602	34.191	
Anger	0	16.078	13.934	$t = 1.858, p = .001$
	1	18.007	24.641	
Sadness	0	25.939	26.977	$t = 2,725.321, p = .000$
	1	12.373	8.331	
Disgust	0	20.473	25.898	$t = 201.995, p = .000$
	1	11.212	13.007	
Joy	0	82.670	28.934	$t = 59.709, p = .000$
	1	79.260	31.152	
Surprise	0	21.849	17.921	$t = 0.965, p > .05$
	1	22.232	17.498	
Fear	0	8.857	6.088	$t = 276.526, p = .000$
	1	23.724	20.232	
Contempt	0	68.692	34.877	$t = 53.094, p = .000$
	1	23.724	20.232	
Valence	0	−1.445	13.453	$t = 3,425.220, p = .000$
	1	−.314	8.745	

Welch's *t*-test was performed to determine if there was a statistically significant difference on basic emotions between the control and the experimental group (Table 1). No significant difference between the two groups was found for engagement and surprise ($p > .05$). As Table 1 shows, the control group (condition 0) showed significantly higher scores of sadness, disgust, joy, contempt, and overall negative valence, whereas the experimental group had significantly higher mean values for anger and fear.

A one-way analysis of variance (ANOVA) was performed on affective valence and the mean time percent of each basic emotion, showing no significant difference among the four experimental tasks, $p > 0.05$, meaning that the content of each task did not emotionally affect the participants. Further investigation using Welch *t*-test did not reveal any additional significance.

4.2 EEG Alpha Asymmetry

The effect of test condition on participants' approach/avoidance states was studied. After a previous automatic decontamination process for artefact removal, the FAA mean values were calculated (Table 2).

Table 2. The mean (M) and standard deviation (SD) for frontal alpha asymmetry in the control (condition 0) and experimental (condition 1) groups.

	Condition	M	SD
Frontal alpha asymmetry	0	0.7165	5.926
	1	−0.1641	0.864

A one way analysis of variance (ANOVA) was performed on FAA mean values, showing no significant difference among the four experimental tasks, $F(3, 75) = 0.639$, $p > 0.05$, meaning that the content of each task did not affect the participants' approach/avoidance EEG correlates.

A one way ANOVA on FAA mean values also showed no significant difference between the two groups (control and experimental), $F(3, 75) = 0.639$, $p > .05$.

4.3 Eye Tracking

The mean values of both the duration and number of eye fixations measured during the four tasks were calculated for both control and experimental conditions (Table 3).

Table 3. The mean (M) and standard deviation (SD) for the duration and number of eye fixations in the control (condition 0) and experimental (condition 1) groups.

	Condition	M	SD
Duration of fixations (ms)	0	175.095	25.810
	1	161.998	26.820
Number of fixations	0	101.773	118.080
	1	117.136	84.500

No significant difference among the four experimental tasks was found by a one way analysis of variance (ANOVA) performed on mean values of duration, $F(3, 56) = 1.033$, $p > .05$, and number of eye fixations, $F(3, 56) = 1.444$, $p > .05$. Hence, the content of each task did not significantly affect the eye movements of participants.

A one way ANOVA between the two groups (control and experimental) showed no significant differences in the duration, $F(1, 63) = 3.061$, $p = .51$, or the number of fixations, $F(1, 63) = 0.358$, $p > .05$.

5   Discussion

The aim of this work was to introduce the user-centered design of a chatbot solution integrated in eGLU-box Pro, a usability assessment platform for Italian PA.

An experimental investigation was performed to analyze the impact on psychophysiological measures of experience when using a chatbot solution during a typical usability test of a website with eGLU-box PA Pro. Two groups were compared: One group performed the test with the chatbot version of the platform (experimental condition), while the other group used the platform without the chatbot solution enabled (control condition).

Statistical comparisons of the results of the experimental condition with the control condition were performed. Overall, there were no significant differences between the two groups for the number and duration of eye fixations and the approach/withdrawal motivation measured with EEG methods. The only significant differences were for facial expression recognition, with mean negative valence values higher for the control group performing the usability test without the aid of the chatbot compared with the experimental group using the chatbot solution. In particular, on the one hand, higher mean values of sadness, disgust, and contempt were found for the control group, but also higher values of joy. On the other hand, participants in the experimental condition showed higher mean values of anger and fear than the control condition. Finally, based on within-subject analyses, there was no difference for any of the three psychophysiological measures among the four tasks of the usability test meaning that the test difficulty was balanced across tasks.

The literature shows that eye fixations reflect cognitive processes in terms of mental load. Specifically, the more a visual stimulus demands analytic processes, the more fixation time and number of fixations is higher [23]. The average eye fixation in the experimental condition (about 161 ms per task) and number of fixations (about $N = 117$ per task) did not significantly differ from the control condition (with a mean fixation time of about 175 ms per task and a mean number of fixations of about $N = 101$ per task). This result means that the introduction of a conversational agent during a typical usability assessment test with eGLU-box did not increase cognitive workload in participants.

The findings on EEG measures also showed no differences in FAA mean values between the two groups, meaning that users' approach/withdrawal motivation processes did not significantly change in either condition. Introducing a conversational agent in an automatic usability assessment procedure did not significantly affect the quality of interaction in terms of user experience.

The main difference between the groups was in basic emotions. Overall, the control group had a significantly higher negative-valence experience, showing emotions such as sadness, disgust, and contempt that we did not find in the group using the chatbot as a source of help while performing the tasks. Those results might be due to the absence of a supervisor monitoring the user on the control group. Hence, the participant is free to complete the tasks without any tips or feedback, thus leading to more negative emotions during the test. The chatbot group experienced significantly higher emotional states related to anger and fear and lower emotional states related to joy. Those results might be due to usability issues related to the introduction of the chatbot, which might have prompted a decrease in enjoyment and an increase in negative emotions because of a reduction of perceived freedom if compared to the group that has not received any help or feedback during the test. Further studies will focus on the emotional aspects of the user

experience with the proposed chatbot solution and investigate which usability aspects of it negatively affect user's emotions.

6   Conclusions

In this paper, we presented the user-centered design of a chatbot solution integrated in a usability assessment platform called eGLU-box Pro. The platform aims at helping the Italian PA in the assessment of their digital web-based services. The user experience of the integrated chatbot was compared with a control condition without the chatbot through three bio-behavioral measures: eye tracking, EEG, and facial expression recognition. There was no noteworthy difference among the four tasks comprising the test between the conditions, meaning that the difficulty of the tasks was well balanced across the test. Participants in both conditions did not show differences in psychophysiological measures of subjective experience such as eye movements and approach/withdrawal processes measured by EEG. Only basic emotions and emotional valence highlighted differences between the two groups. On the one hand, the chatbot users experienced higher anger and fear and lower joy than the control users, probably due to usability issues strictly related to the chatbot presence during the tasks. On the other hand, the tests conducted with no chatbot providing them feedback and help in following some correct procedures increased general negative-valence emotions, in particular sadness, disgust, and contempt. This result might be related to the absence of any help during the test. Overall, the study shows that introducing the eGLU-box PA conversational agent in an automatic usability assessment procedure did not significantly affect the quality of interaction in terms of user experience except for emotions. Future phases of the ongoing project presented here will focus on the usability aspects of the chatbot solution related to emotions, satisfaction and pleasantness of use.

References

1. Radziwill, N.M., Benton, M.C.: Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint arXiv:1704.04579 (2017)
2. Ammari, T., Kaye, J., Tsai, J.Y., Bentley, F.: Music, search, and iot: how people (really) use voice assistants. ACM Trans. Comput.-Hum. Interact. **26** (2019). https://doi.org/10.1145/3311956
3. Beaudry, J., Consigli, A., Clark, C., Robinson, K.J.: Getting ready for adult healthcare: designing a chatbot to coach adolescents with special health needs through the transitions of care. J. Pediatr. Nurs. **49**, 85–91 (2019). https://doi.org/10.1016/j.pedn.2019.09.004
4. Costa, S., Brunete, A., Bae, B.C., Mavridis, N.: Emotional storytelling using virtual and robotic agents. Int. J. Hum. Robot. **15** (2018). https://doi.org/10.1142/S0219843618500068
5. Dmello, S., Graesser, A.: Autotutor and affective autotutor: learning by talking with cognitively and emotionally intelligent computers that talk back. ACM Trans. Interact. Intell. Syst. **2** (2012). https://doi.org/10.1145/2395123.2395128

6. Abd-Alrazaq, A.A., Alajlani, M., Alalwan, A.A., Bewick, B.M., Gardner, P., Househ, M.: An overview of the features of chatbots in mental health: a scoping review. Int. J. Med. Inf. **132**, 103978 (2019). https://doi.org/10.1016/j.ijmedinf.2019.103978

7. Balaji, D., Borsci, S.: Assessing user satisfaction with information chatbots: a preliminary investigation. University of Twente, University of Twente repository (2019). http://essay.utwente.nl/79785/

8. Tariverdiyeva, G., Borsci, S.: Chatbots' perceived usability in information retrieval tasks: an exploratory analysis. University of Twente, University of Twente repository (2019). http://essay.utwente.nl/77182/

9. Federici, S., et al.: Heuristic evaluation of eGLU-Box: a semi-automatic usability evaluation tool for public administrations. In: Kurosu, M. (ed.) HCII 2019. LNCS, vol. 11566, pp. 75–86. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22646-6_6

10. Catarci, T., et al.: Digital interaction: where are we going? In: Proceedings of the 2018 International Conference on Advanced Visual Interfaces: AVI 2018, pp. 1–5 (2018). https://doi.org/10.1145/3206505.3206606

11. Federici, S., Mele, M.L., Bracalenti, M., Buttafuoco, A., Lanzilotti, R., Desolda, G.: Biobehavioral and self-report user experience evaluation of a usability assessment platform (UTAssistant). In: VISIGRAPP 2019: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. Volume 2: HUCAPP, pp. 19–27 (2019)

12. Hammontree, M., Weiler, P., Nayak, N.: Remote usability testing. Interactions **1**, 21–25 (1994). https://doi.org/10.1145/182966.182969

13. Borsci, S., Federici, S., Bacci, S., Gnaldi, M., Bartolucci, F.: Assessing user satisfaction in the era of user experience: comparison of the Sus, Umux and Umux-lite as a function of product experience. Int. J. Hum.-Comput. Interact. **31**, 484–495 (2015). https://doi.org/10.1080/10447318.2015.1064648

14. Reichheld, F.F.: The one number you need to grow. Harv. Bus. Rev. **82**, 133 (2004)

15. Moore, R.J.: A natural conversation framework for conversational UX design. In: Moore, R.J., Szymanski, M.H., Arar, R., Ren, G.-J. (eds.) Studies in Conversational UX Design. HIS, pp. 181–204. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95579-7_9

16. Duchowski, A.T.: Eye Tracking Methodology: Theory and Practice. Springer, London (2007). https://doi.org/10.1007/978-1-84628-609-4

17. Ekman, P., Friesen, W.V.: The Facial Action Coding System (Facs): A Technique for the Measurement of Facial Action. Consulting Psychologists Press, Palo Alto (1978)

18. Schur, M., Ritvo, L.B.: A principle of evolutionary biology for psychoanalysis: schneirla's evolutionary and developmental theory of biphasic processes underlying approach and withdrawal and freud's unpleastire and pleaswe principles. J. Am. Psychoanal. Assoc. **18**, 422–439 (1970). https://doi.org/10.1177/000306517001800210

19. Pizzagalli, D.A.: Electroencephalography and high-density electrophysiological source localization. In: Caccioppo, J.T., Tassinary, L.G., Berntson, G. (eds.) Handbook of Psychophysiology, pp. 56–84. Cambridge University Press, Cambridge (2007)

20. Schneirla, T.C.: An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In: Jones, M.R. (ed.) Nebraska Symposium on Motivation, vol. 7, pp. 1–42. University of Nebraska Press, Lincoln (1959)

21. Coan, J.A., Allen, J.J.B.: Frontal EEG asymmetry as a moderator and mediator of emotion. Biol. Psychol. **67**, 7–49 (2004). https://doi.org/10.1016/j.biopsycho.2004.03.002

22. Allen, J.J.B., Coan, J.A., Nazarian, M.: Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. Biol. Psychol. **67**, 183–218 (2004). https://doi.org/10.1016/j.biopsycho.2004.03.007

23. Jacob, R.J.K.: The use of eye movements in human-computer interaction techniques: what you look at is what you get. TOIS **9**, 152–169 (1991). https://doi.org/10.1145/123078.128728