Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 52-19 (2019) 25–30

Survey on usability assessment for industrial user interfaces

Valeria Villani, Giulia Lotti, Nicola Battilani, Cesare Fantuzzi

Department of Sciences and Methods for Engineering (DISMI), University of Modena and Reggio Emilia, Reggio Emilia, Italy (e-mail: name.surname@unimore.it).

Abstract: Modern production systems have become highly complex, given technological progress and competitive market needs. However, the role of human operators keeps being fundamental since they are in charge of supervising machines and take proper action in the presence of alarms and faults. As a consequence, the design of effective and easy to use human-machine interfaces (HMIs) is a key component of overall production efficiency.
In this paper we provide an extensive overview of different methods to assess the usability of HMIs for industrial operators. In particular, the methods covered by the survey range from subjective assessment to objective quantitative analysis that takes into account user's mental fatigue, user's interaction strategies and production efficiency.

Keywords: Human-machine interface, User interfaces, Man/machine systems

1. INTRODUCTION

In recent years, the complexity of modern production systems has dramatically increased, to comply with diversified needs of market, flexible production and competitiveness. Despite technological progress, the presence of human operators is still fundamental to operate machines and supervise plants.Machine complexity reflects on operators, who are compelled to deal with it. In this scenario, being the point of contact between the operator and the machine, human-machine interfaces (HMIs) have become a critical part of the system, since they greatly affect the efficiency of interaction and, ultimately, production. Designing usable user interfaces is, hence, an important and challenging part of the design of efficient production systems. According to the definition of usability provided in (ISO, 2018), an interaction system is usable if it supports the user in the execution of a task, making it more effective, easier and more pleasant.

Moving along these lines, the assessment of the usability of HMIs is a fundamental part of the design and validation of complex interaction system. In particular, the main goals of usability assessment are: i) verifying that the functionalities of the system are complete, accessible and easy to find by the user; ii) verifying the experience of interaction of users, in terms of learnability, usability and satisfaction that the system entails and identifying the areas of design that overload the user; iii) finding specific interaction problems or design features that, when used in the real context, cause unforeseen results or confusion among users.

This paper presents an extensive overview of the tools to assess the usability of user interfaces. The presented methods are general and can be applied also to domain different from the industrial one. In particular, we discuss approaches for either direct or indirect assessment of usability: methods to investigate precise features of usability are presented, together with quantitative parameters related to interaction that are influenced by the usability of a system. An example in this regard is operator's workload during the interaction, which represents an indirect measure of HMI usability.

2. SUBJECTIVE USABILITY ASSESSMENT

Subjective usability assessment consists in eliciting feedback from users of the HMI. Enrolled subjects do not necessarily have to be expert in the HMI or in the task. Indeed, while, of course, feedback from end users is fundamental, preliminary assessment by usability expert provides useful insights.

2.1 Predictive evaluation by usability experts

The assessment of the usability of a user interface by usability experts represents a kind of predictive evaluation, since it is carried out before involving real users, to address usability issues unrelated to system functionalities. These methods allow to predict, rather than observe, usability problems that will arise when the system is used or tested by end users. Thus they allow to take proper intervention with predictive corrective strategies. Specifically, predictive usability assessment is useful: i) before tests with users, since it allows not to involve users too early, on minor problems that can be easily found and solved; ii) before redesign, since it allows to understand what is fine and what should changed; iii) when evidence on existing problems has to be collected and presented systematically: for example, when users complain about something, and the other stakeholders (e.g., developers, management) need to be convinced. The most common approaches to predictive usability assessment are heuristic evaluation and cognitive walkthrough.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2019.12.078

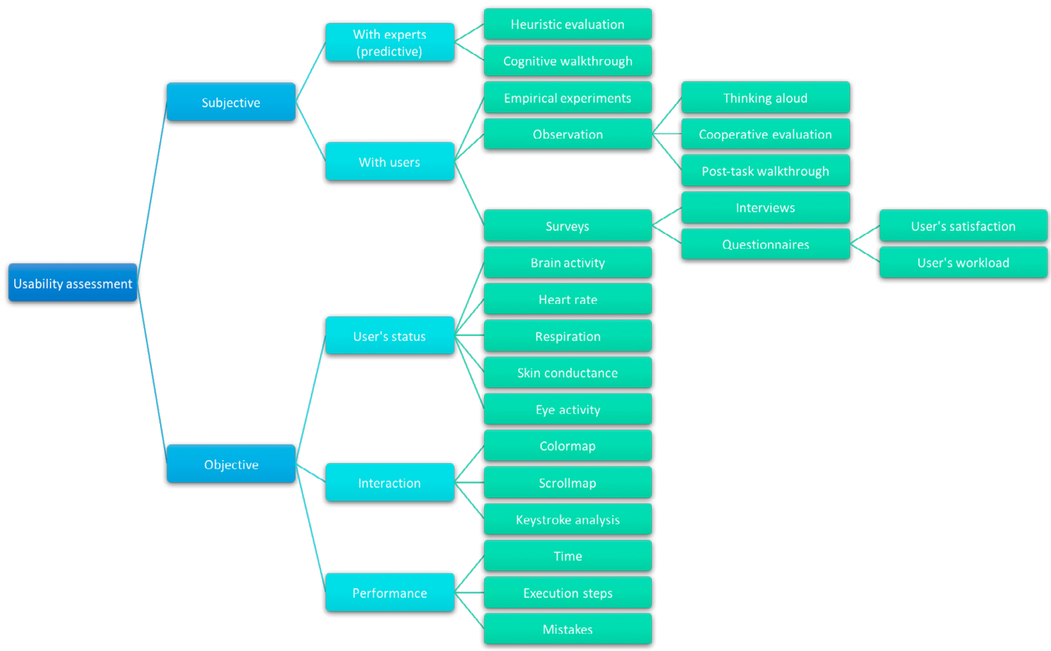

Fig. 1. Overview of the methods for usability assessment of user interfaces.

Heuristic evaluation It is a systematic approach to inspection, not to measurement, which consists in experts (not real users) interacting with the system, trying to find potential problems, and giving prescriptive feedback. Specifically, the system is assessed in terms of its compliance with recognized usability principles or rules of thumb (e.g., the heuristics by Nielsen (Nielsen, 1994)). The evaluators are specialists who are familiar with the evaluation techniques, the domain (not always) and the perspective users. Since they do not use the system to accomplish a real task, the heuristic evaluation can be carried out also on paper prototypes, for interfaces not yet implemented. In the case of evaluators who are not expert on the domain, they should be properly instructed and assisted to learn how to use the interface: to this end, it is needed to describe the typical working scenario, listing the steps that a user would follow to accomplish a real task. The result of the procedure is a list of usability problems with respect to the violated usability principles. The ultimate goal is to find the usability problems in a user interface design so that they can be attended to as part of an iterative design process.

As regards the number of evaluators, a single person will never succeed in finding all the usability problems of an interface. In addition, it has been found that different evaluators find different usability problems (Nielsen, 1994). Thus, it is possible to increase the effectiveness of the approach by enrolling several evaluators and aggregating their evaluations. A good compromise proposed in the literature is to enrol about five evaluators, certainly at least three (Nielsen, 1994). However, the exact number of evaluators to use would depend on a cost-benefit analysis.

The heuristic evaluation of a user interface should follow the following steps. First, each evaluator investigates the system on her/his own, finding the problems of usability and their causes: an analysis report is produced with a short description of the usability problems found, with respect to the usability principles violated by the design. Comments to the interface can be written down by the evaluator or an observer: in this case, the evaluator speaks aloud and the observer takes note. The presence of an observer is not necessary, but might be useful, especially if evaluators are not expert in the domain. Although the observer is a cost, she/he helps having data quickly ordered and can help the evaluator to overcome difficulties, if needed. After the analysis, the evaluators have a debriefing session where all the identified usability problems are considered in detail and compared. Results are discussed and aggregated in a final document where all the problems found are described in a unified language. For each problem, the number of evaluators who found it should be reported, possibly together with a degree of severity ranging from 0 to 4 (less-more severe), useful to order and prioritize problems. In addition, a second debriefing meeting might be held among the evaluators to discuss possible solutions to the problems found and changes to the observed system.

Compared to experiments with users that are discussed in the following, heuristic evaluation is a faster and less expensive technique that can be used also with early prototypes. Moreover, results are pre-interpreted and can be quickly exploited to redesign the interface. The major drawbacks are that it is less accurate since some problems might be ignored or some false positives can be found, and it does not consider real users and their tasks. However, if considered jointly with experiments with end users, the two techniques allow very good results.

Cognitive walkthrough The term walkthrough refers to those techniques that require a detailed review of the sequence of actions performed. Literally, this approach to usability assessment consists in walking through the interaction: an expert performs a detailed review of the sequence of actions that the system requires the user to follow to accomplish a task (Helander, 2014). In particular, it focuses on the learnability of the system, through exploration: for example, the features that increase efficiency but are hidden or difficult to learn will be considered negatively.

The experimental protocol consists in two phases. Preliminarily, the task to accomplish and the objectives to achieve must be defined, together with a complete list of actions-reactions required to complete the task with the system or prototype. Then, the evaluator reviews the prescribed actions step-by-step and for every action, she/he answers 4 questions:

(1) Does the user understand what she/he should do?
(2) Is the user able to identify the interaction tool?
(3) If the user is able to identify the interaction tool, can she/he understand what she/he should do to achieve the goal?
(4) After the action has been performed, can the user understand the answer?

These questions search for the presence of discoverability (question 2), affordances (question 3) and feedback (question 4) in the user interface. Moreover, with respect to the seven stages of an action identified by Norman (Norman, 2013) and depicted in Fig. 2, they inspect nearly all the stages that a user unconsciously passes through when using an interactive system: how the execution of the task is planned is inspected by question 1; how the plan is specified in terms of a sequence of steps is inspected by questions 2 and 3; finally, question 4 inspects how system response is perceived by the user, interpreted and compared to expectations.

Answers to these questions are reported in an evaluation report with a step-by-step description of the actions performed by the evaluator, divided in as many modules as the actions are. Each module reports the answers to the 4 questions for each action and, based on these, the usability problems associated to each single action should be described in a usability report, where every usability problem is accompanied with the action and the question that solicited it. Also in this case, for every problem, it is beneficial to add a severity rating that quantifies how frequent and detrimental it is for the user.

This approach to predictive usability assessment is limited with respect to the other techniques of evaluation, since it investigates only one dimension of usability, and ignores other important features (e.g., global consistency, prevention/management of catastrophic errors). However, one of its strength is that it finds any lack of consistency between the user's conceptual model of the task and the designer's conceptual model. In particular, it highlights any wrong selection of terminology (e.g., menus, buttons) and when there is inadequate feedback. Moreover, it highlights (implicit) assumptions by designers with respect to the knowledge the user has of the task and the conventions in the interface.

2.2 Experiments with users

Usability assessment with users follows two different test settings: lab studies and field studies. In lab studies, users are extracted from their working context and take part in controlled experiments, usually, but not necessarily, in a specialized usability lab. Advantages are that, if well equipped, the lab may have advanced tools for analysis and audio- or videotaping, such as bilateral mirrors, and the test subjects are not interrupted during the task. However, the recorded scenario differs form the real one due to lack of context, lack of interruptions and non natural environment. Lab studies are specifically suited for remote or dangerous systems, or if deliberate manipulation of the context is needed to identify problems or procedure rarely used, or to compare alternative systems. In field studies, the designer or the evaluator enters the working environment of the user to observe the system while in use. As a consequence, it is possible to observe interactions between systems and operators, which cannot be seen in lab, and interaction can be studied during effective use. However, observation is made difficult by high environmental noise, frequent displacements and interruptions (e.g., phone calls). Also in field studies test subjects are influenced by devices for recording and/or analysing their actions: the observed scenario is never the same as the real one.

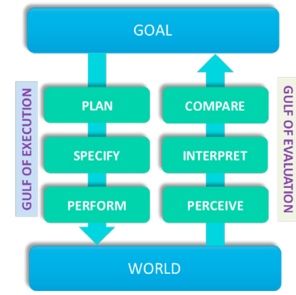

Fig. 2. The interaction with a system goes through seven stages, which are often unconscious (Norman, 2013). Cognitive walkthrough provides a critical analysis of such stages to determine the usability of the system.

Tests can be designed considering specific task scenarios or open ended scenarios. In specific task scenarios, users are given clear guidance on what actions to perform and what features to explore. On the contrary, in open ended scenarios users are not given out a lot of information about how to perform the task, since the idea is rather to watch users uncover the solution on their own and establish a strategy for action.

Given these distinctions on test settings, different testing methodologies exist.

Empirical methods According to this methodology, an experimental protocol is designed to observe a user interacting with the interface to accomplish a predefined task. To this end, a controlled experiment is designed by:

- selecting the hypothesis to test (problem),
- defining the experimental variables,
- setting the control variables (fixed, which represent the situational conditions),
- manipulating independent variables,
- observing (or measuring) dependent variables.

The goal of the experiment is to prove whether there is or not an effect in the dependent variable(s) when the independent one(s) is changed. It is worthwhile noting that, following this approach, it is possible to achieve good observation addressing *specific* aspects of the interaction process with *limited* goals.

As regards the selection of variables, they should be selected considering the assumption that the experiment aims at verifying an hypothesis by changing and measuring variables in controlled conditions. To this end, distinction has to be made between:

- independent variables: they represent those elements in the system that are changed to create different conditions to compare (for example, style of the interface, amount of guidance, number of elements in menus, or design of icons);
- dependent variables: these are the variables that can be measured in the experiment and their value depends on changes to the independent variables. Examples are time taken to accomplish a task, number of mistakes, user's preference and quality of performance.

An example is an experiment that seeks to verify whether the time required to search an item in the user interface changes with the number of elements in menus and their names.

As regards the selection of participants, test subjects should be as much similar as possible to real end users; when experimental testing with real end users is not possible, test subjects should have similar characteristics to real end users (e.g., analogous age, education level, experience with computers and systems similar to the one under test, experience and knowledge of the working scenario and the task). The number of enrolled test subjects is usually limited by practical constraints, such as budget and availability of test subjects. However, the sample size should be big enough for statistical analysis to hold true: in (Dix et al., 2004) it is suggested to enrol at least ten participants.

The experimental design might follow two different schemes: within subjects or between subjects. In the within subject design, each participants tests all the experimental conditions. As a consequence, results can be affected by learning effect, which can be attenuated varying the order of test conditions. Clearly, this approach is useful when learning effect is to be considered. Moreover, it is less expensive because it requires fewer users and reduces the effect of differences among test subjects. In the between subject (also called randomized) design, each participants tests a single experimental conditions. Hence, it requires at least two experimental conditions: the experimental condition, in which the independent condition is manipulated, and the control condition, which is the same as the experimental condition, without manipulation in the experimental condition. Thus, it is important to ascertain that the differences found are due to the manipulation. Following this design, any learning effect is under control. However, a greater number of participants is needed and a great difference between users groups affects the significance of results.

Observational methods These methods for usability assessment consist in observing users during interaction to collect information about the real use of a system. An evaluator observes and records user's actions doing as little as possible in order not to interfere with their work. Different practical approaches can be used to track user's actions, such as writing down notes, audio or video recording, PC recording or analyzing user's notes. However, simply observing is likely to be not sufficient to determine whether the system satisfies user's requirements. To this end, observation should be organized in a structured way to collect useful feedback from the user. In the following thinking aloud, cooperative evaluation and post-task walktrough are discussed.

Thinking aloud According to this approach, the user is requested to describe aloud what she/he is doing while being observed and verbalize her/his thoughts. The main advantage of this method is that it is easy to implement, since it requires little experience. Nonetheless, it can provide good understanding of problems in the HMI. Moreover, it can be used to observe how the system is used in practice, but can be useful also during the whole design, for example, using prototypes in the early phases. As drawbacks, it should be mentioned that reported information is subjective and observation might change the way users perform a task, since having to describe what is being done usually changes the way it is done.

Cooperative evaluation It is a variant of thinking aloud in which the user is encouraged to see her/himself as a collaborator in the evaluation rather than just a subject to be observed. This means that while the user thinks aloud, the evaluator can ask such questions as "Why?" and "What if.....?" and the user can ask the evaluator for clarification if problems arise. As a consequence, the user feels more relaxed than in thinking aloud and is encouraged to actively criticise the system rather than simply suffer it. Additionally, the evaluator can clarify points of confusion so maximising the effectiveness of the approach.

Post-task walkthrough Usually data recorded during unobtrusive observation lack interpretation: the exact sequence of actions performed is known, but little about why they were executed can be inferred and no information is given with respect to possible alternative actions that were not performed. To overcome this issue, post-task walkthrough consists in reconsidering actions soon after the experimental session. Specifically, it is necessary if it is needed to deduce discussion during the experiment, for an unobtrusive observation. Moreover, it is the only way to gather a personal opinion when the user cannot speak during the experiment, for example in the case of tasks too difficult or critical.

Surveys Directly questioning the test subjects is useful to gather their point of view and find problems not considered by designers. Surveys are cheap and easy to manage, but, of course, collected information are subjective. Two survey techniques are interviews and questionnaires.

Interviews Interviews are a direct and structured way to collect information. It is possible to adapt the level of questions to the context and, to this end, the evaluator can investigate more in detail interesting questions as they arise. Typically a top-down approach is used: the evaluator starts with general questions on the task and then continues with more specific questions (e.g., "Why?" or "What happens if...?"). Questions can be also useful to collect information on preferences, impressions and mental behaviours of users.

Questionnaires Questionnaires represent an alternative less flexible than interviews since questions are predefined. On the other side, they allow reaching a greater number of subjects, are less time consuming and can be analysed more systematically. Usability questionnaires can be designed for the application under analysis, including questions that investigate specific aspects that need in depth analysis. However, some standard questionnaires having general validity are widely used, thus providing also a benchmark for tests of different systems with analogous functionalities. In the following we provide an overview of the most used questionnaires. They are mostly related to the assessment of user's satisfaction, which is one of three dimensions of usability (ISO, 2018), or workload.

The *computer user satisfaction* questionnaire (Bailey and Pearson, 1983) measures, among the others, accuracy, reliability, timeliness, relevancy, precision of information output, system flexibility and confidence in the system. The *questionnaire for user interface satisfaction* (Chin et al., 1988) investigates user satisfaction in terms of five different aspects: overall reaction to the interface, organization of the screen, terminology and system information, learnability, and system capabilities. The *computer system usability questionnaire* (Lewis, 1995) contains 19 questions measuring characteristics such as ease of use, learnability, simplicity, effectiveness and organization of the user interface. The *system usability scale* (Brooke et al., 1996) is a validated tool for measuring the usability of a wide variety of products and services and was designed to meet the need of a short, simple tool to be used in industrial settings (Bangor et al., 2008). Examples of questions are "I think that I would like to use this system frequently", "I found the system unnecessarily complex" and "I felt very confident using the system". Other questionnaires for investigating user's satisfaction are the *USE Questionnaire* (Lund, 2001), *after-scenario questionnaire* (Lewis, 1995) and *Purdue Usability Testing Questionnaire* (Lin et al., 1997).

Another widely used tool of the assessment of user interfaces, with specific focus on workload, is the *NASA task load index* (Hart and Staveland, 1988), which measures workload in terms of mental demand, physical demand, temporal demand, own performance, effort and frustration. Other questionnaires for investigating mental workload are the *subjective workload assessment technique* (Reid and Nygren, 1988) and the *multiple resource questionnaire* (Boles and Adair, 2001), which considers workload in terms of different resources drained to the user.

3. OBJECTIVE USABILITY ASSESSMENT

The methods presented in the previous section allow to collect experts' or users' subjective feedback about the user interface. As a consequence, collected information will be unavoidably subjective and, in some cases, affected by the collection procedure itself. To overcome this issues, it is possible to collect an indirect objective assessment of usability of a user interface by measuring the interaction of the user with the HMI. Specifically, information relevant to this end are user's emotional condition, her/his way to interact with the system and performance. From such data it is possible to get an implicit and unsolicited feedback about system interface that can be exploited to reconsider redesign. Moreover, such information could also be used for an on-line adaptation of the HMI to the current measured condition of the user. As an example, if an increase in mental workload is measured, or if it is tracked that the user is struggling with some specific areas of the interface, or her/his performance are dropping, a simplification in the interface or a link to guided procedures could be instantaneously presented to the user.

3.1 Measurement of user's status

Monitoring and interpreting nonverbal communication can provide important insights about a human interacting with a complex system, thus making it possible to achieve implicit feedback about the interaction (Heard et al., 2018). To this end, eye gaze (Rich et al., 2010), facial expression (Gunes et al., 2011), voice, linguistic and paralinguistic (e.g., utterances) features (Gunes et al., 2011), and physiological signals (Kulic and Croft, 2007) have been investigated as indices of subject's affective state, focus, attention and intent. Among physiological signals, most of the studies have employed data related to brain activity, heart rate, respiration, skin conductance and eye (pupil dilatation and blink frequency) activity. Details can be found in (Heard et al., 2018).

Two major issues might limit the use of physiological parameters recording for usability assessment in working environment. First, invasiveness and costs of the recording techniques might discourage their use. However, recently several wearable non-invasive devices that record such parameters with suitable accuracy have been proposed. Examples are commercial smartwatches or armbands that measure heart rate and skin response, wireless EEG headsets and eye trackers. While commercial smartwatches and armbands have reduced costs and are completely transparent to be worn (see, e.g., (Villani et al., 2018)), reliable eye trackers still have very high costs, and EEG headset are cumbersome to be worn, especially in dynamic working environment. Second, recording user's physiological condition to infer feedback about the interaction system poses major challenges in terms of legal and ethics issues, since privacy and freedom of users might be easily violated in case of improper use of such data.

3.2 Measurement of user's interaction with the system

From the way a user navigates through an interface insights about its usability can be inferred. The most relevant examples in this regard are given by colormaps (Bergman et al., 1995), which represent areas of an HMI where users act more frequently, and scrollmaps (Martinez and Rahn, 2000), which provide a representation of areas in an interface where users spend most of the time and how they navigate through the HMI. Such tools can be used to identify the parts of the HMI that mostly absorb the user because they are either crowded of important information or difficult to understand. Moreover, the keystroke analysis is an evaluation technique that consists in predicting how long it will take to users to accomplish low level interaction actions (Card et al., 1980). The execution of a task is described in terms of physical-motoric and cognitive operators, examples of the former being keystroking or pointing to a target on a display with a mouse, and of the latter mentally preparing for executing physical actions. The goal is to provide an estimate of the minimum duration of a task, and this technique can be used to compare different versions of a system, in terms of efficiency. Indeed, it does not replace a usability test, but is a tool to assess the design of an interface, without implementing it, train end users and measure their performance.

3.3 Measurement of user's performance

As in the case of the parameters mentioned in the previous paragraph, performance indicators can be identified and used to identify pitfalls in the user interface. Examples of such indicators are: execution time, reaction time, time for decisions, execution steps, mistakes and redundancies. Reference values for these parameters should be defined depending on the goal of the usability assessment. For example, if the learnability of a new system is to be investigated, then reference values might be measured considering non-novice users; or reference values with an established interface can be measured to assess the goodness of a new one. Measuring these parameters can be done by tracking user's action on the interface, in a way that is completely transparent to the user and does not interfere with her/his task. As mentioned above, measured values can be used also to develop structural knowledge maps and register the training evolution to give support to the user in the future. Also these recording techniques raise ethical and legal issues, since the same data can be used to assess and rank the user rather the interface.

4. CONCLUSION

In this paper an extensive overview of different methods to assess the usability of user interfaces has been presented. In particular, the methods covered by the survey range from subjective assessment to objective quantitative analysis that takes into account user's mental fatigue, clarity of presented information and production efficiency.

ACKNOWLEDGEMENT

The research is carried out within the "Smart and adaptive interfaces for INCLUSIVE work environment" project, funded by the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement N. 723373).

REFERENCES

(2018). ISO 9241-11:2018. ergonomics of human system interaction – part 11: Usability: Definitions and concepts. *ISO/TC 159/SC 4 Ergonomics of human-system interaction. International Standardization Organization (ISO).*

Bailey, J.E. and Pearson, S.W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management science*, 29(5), 530–545.

Bangor, A., Kortum, P.T., and Miller, J.T. (2008). An empirical evaluation of the system usability scale. *Intl. J. Human-Computer Interaction*, 24(6), 574–594.

Bergman, L.D., Rogowitz, B.E., and Treinish, L.A. (1995). A rule-based tool for assisting colormap selection. In *Proc. 6th conference on Visualization*, 118. IEEE.

Boles, D.B. and Adair, L.P. (2001). The multiple resources questionnaire (MRQ). In *Proc. Human Factors and Ergonomics Society Annual Meeting*, volume 45, 1790–1794. SAGE Publications Sage CA: Los Angeles, CA.

Brooke, J. et al. (1996). SUS: A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4–7.

Card, S.K., Moran, T.P., and Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7), 396–410.

Chin, J.P., Diehl, V.A., and Norman, K.L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proc. SIGCHI Conf. Human Factors in Computing Systems*, 213–218. ACM.

Dix, A., Finlay, J.E., Abowd, G.D., and Beale, R. (2004). *Human-Computer Interaction.* Pearson, 3rd edition.

Gunes, H., Nicolaou, M.A., and Pantic, M. (2011). *Continuous Analysis of Affect from Voice and Face*, 255–291. Springer London.

Hart, S.G. and Staveland, L.E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, 139–183. Elsevier.

Heard, J., Harriott, C.E., and Adams, J.A. (2018). A survey of workload assessment algorithms. *IEEE Trans. Human-Machine Systems.*

Helander, M.G. (2014). *Handbook of human-computer interaction.* Elsevier.

Kulic, D. and Croft, E.A. (2007). Affective state estimation for human–robot interaction. *IEEE Trans. Robotics*, 23(5), 991–1000.

Lewis, J.R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57–78.

Lin, H.X., Choong, Y.Y., and Salvendy, G. (1997). A proposed index of usability: a method for comparing the relative usability of different software systems. *Behaviour & information technology*, 16(4-5), 267–277.

Lund, A.M. (2001). Measuring usability with the USE questionnaire. *Usability interface*, 8(2), 3–6.

Martinez, A.E. and Rahn, M.D. (2000). Graphical selection marker and method for lists that are larger than a display window. US Patent 6,147,683.

Nielsen, J. (1994). *Usability engineering.* Elsevier.

Norman, D.A. (2013). *The design of everyday things: Revised and expanded edition.* Basic books.

Reid, G.B. and Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In *Advances in psychology*, volume 52, 185–218. Elsevier.

Rich, C., Ponsler, B., Holroyd, A., and Sidner, C.L. (2010). Recognizing engagement in human-robot interaction. In *5th ACM/IEEE Int. Conf. Human-Robot Interaction (HRI)*, 375–382. IEEE.

Villani, V., Sabattini, L., Secchi, C., and Fantuzzi, C. (2018). A framework for affect-based natural human-robot interaction. In IEEE (ed.), *27th IEEE Int. Symp. Robot and Human Interactive Communication (RO-MAN)*, 1038–1044.