

Clear Skies, Clear Interface – User-Centered Design of a Mobile Duty Roster Application for Pilots

Jan Kuhn^{1,†}, Stephan Schlögl^{1,*,†} and Aleksander Groth^{1,†}

¹MCI – The Entrepreneurial School, Innsbruck, Austria

Abstract

While continuous digitalization efforts have certainly transformed entire industries and led to more efficient work processes, they have also introduced a number of challenges for users. One of these challenges is found in aviation, where the increasing use of portable electronic devices in commercial aircraft cockpits raises concerns about distraction and cognitive overload, potentially threatening flight safety. To this end, our work has examined the usability of a prototype mobile duty roster application for pilots, which aims at minimizing such cognitive load. The prototype was developed in accordance with well-established usability heuristics for mobile applications and subsequently evaluated via formal usability testing, including a retrospective think-aloud analysis. In this, quantitative data was collected via the Single Ease Question and System Usability Scale questionnaires. Our respective findings show that, while generally the prototype was perceived to be intuitive and easy to learn and operate, challenges connected to scrolling issues, interface confusions, and cognitive load peaks remained. The latter were particularly apparent when evaluating the data from the additional tapping task study participants had to perform while interacting with the prototype.

Keywords

Mobile Application Design, User-Centered Design, Usability Heuristics, Usability Testing, Cognitive Load, Aviation Industry

1. Introduction

During the past two decades, the proliferation of digital technologies and Internet connectivity has dramatically transformed industries and work processes [1, 2]. A digital revolution has created a connected society, where constant access to information and communication has become the norm, and the lines between work and personal life have increasingly blurred [3]. And although this technology penetration has led to numerous benefits, such as greater productivity and easier access to information, it has also introduced new challenges, particularly with respect to perceived techno-stress and sensory overstimulation [4, 5]. Respective cognitive overload, characterized by excessive exposure to information and the demands of managing multiple devices and applications simultaneously [6, 7], can lead to distraction, reduced concentration, and decreased productivity [8, 9]. This has also affected the aviation industry. Today's cockpits, once dominated by analog instruments, are now equipped with an array of digital displays and portable electronic devices [10]. While these technologies offer pilots valuable tools for flight management and communication, they also contribute to a complex and potentially overwhelming information environment [11]. That is, pilots are consistently exposed to a multitude of stimuli, including aircraft system alerts, navigation displays, air traffic control communications, and weather radar, all while needing to maintain situational awareness. The complexity resulting from such an interplay of information and technology can lead to an excessive mental workload and an increased risk of human error, potentially jeopardizing flight safety.

Recognizing this impact of poorly designed interfaces on cognitive load and flight safety, the work we report on in this article has been focusing on identifying and addressing the specific usability challenges

SQAMIA 2025: Workshop on Software Quality Analysis, Monitoring, Improvement, and Applications, September 10–12, 2025, Maribor, Slovenia

*Corresponding author.

† These authors contributed equally.

✉ jo.kuhn@mci4me.at (J. Kuhn); stephan.schloegl@mci.edu (S. Schlögl); aleksander.groth@mci.edu (A. Groth)

ORCID 0000-0001-7469-4381 (S. Schlögl); 0000-0003-4698-2321 (A. Groth)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

associated with mobile applications in aviation. The design and subsequent evaluation of a mobile duty roster application for pilots has thereby served as a starting point and testbed for investigations focusing on the following research question:

To what extent can established usability heuristics for mobile applications help design interfaces for everyday tasks in the aviation industry?

Hereinafter we report on the initial phase of this research agenda. We start with a short introduction of the given context in Section 2, which is followed by Section 3, describing the methodology we have used for our analyses. Section 4 then summarizes key findings, before Section 5 concludes our report and Section 6 outlines limitations and future research direction.

2. Context

Along with security, usability counts as a primary concern in mobile applications design, focusing on the intuitiveness of interfaces and their effectiveness in helping users achieve goals [12, 13]. Compared to websites and other types of desktop applications, mobile applications face greater constraints when it comes to screen size [14], multifunctional buttons, and limited processing power [15]. Furthermore, they usually depend on wireless operations and Internet connectivity, which may significantly influence interaction experiences [16] and consequently affect perceived task complexity. In the aviation industry, those challenges are further complemented by often rather complex interfaces that are overcharged with information [17]. As the amount of information increases, cognitive load rises [18], potentially hindering learning and increasing the likelihood of errors [19]. Excessive cognitive load may also compromise flight safety, even during ground operations [11]. Conversely, low cognitive load fosters user satisfaction and performance [20]. Numerous researchers have therefore emphasized the importance of usability and cognitive load considerations in aviation (e.g. [21, 22, 23]), with Clamann & Kaber [17] pointing to formal usability testing and heuristic evaluation as the most effective measures for applications design, although limited budgets and difficulties in the recruitment of suitable participants may hamper respective efforts. We aimed to follow this call and thus employed a user-centered design approach for the development of a mobile duty roster application for pilots.

3. Methodology

Aiming to investigate potential usability and cognitive load challenges associated with a mobile duty roster application for pilots, we employed a combination of heuristic evaluation and cognitively aligned usability testing. We followed a 2-step approach which first used mobile usability heuristics to inform the design of our prototype and subsequently evaluated the impact of this design via formal usability tests.

3.1. Prototype Development

A user interface prototype of the mobile duty roster application for pilots was developed using Figma (<https://www.figma.com/>) and Quant UX (<https://www.quant-ux.com/>). Quant UX enables the creation of interactive screens with functional widgets, allowing for realistic user interactions. Its ‘Design View’ facilitates the visual design and import of elements from Figma, while its ‘Prototype View’ enables the definition of screen interactions and dynamic properties.

Our user interface prototype was designed based on the usability heuristics of Da Costa et al. [24] whose appropriateness were subsequently evaluated in formal usability tests with actual pilots.

3.2. Usability Tests

Five full-time pilots (four male, one female) participated in the formal usability tests of our user interface prototype. Their age ranged from 36 to 49 years ($Mean = M = 39.6$ years, $StandardDeviation =$

$SD = 5.5$ years), and they had between 12 and 25 years of experience as pilots ($M = 15.2$ years, $SD = 5.5$ years). To assess their familiarity with our mobile application test platform (i.e., iOS), participants were asked about their personal smartphone usage. Three of them reported to privately use iOS devices, while two indicated the private use of the Android platform.

3.2.1. Procedure:

The usability tests were conducted remotely, utilizing Quant UX and the participants' employer-issued iPhone 15 models. These iPhones had replaced an older generation of iPhones two months prior to the study. Data collection employed a comprehensive approach, including:

1. *Remote observation:* User interactions with the mobile duty roster application were closely observed and recorded. This involved capturing the smartphone screen, the participant's operating hand, and their facial expressions using webcams.
2. *Additional tapping task:* A simultaneous tapping test was administered to simulate more realistic cognitive pressure and identify critical situations, providing insights into cognitive load variations during application usage.
3. *Retrospective think-aloud analysis:* Participants were asked to verbalize their thoughts and actions retrospectively, shedding light on their cognitive processes while interacting with the application. Such allowed for in-depth exploration of their experiences, providing valuable insights and context.
4. *Questionnaires:* Post-task and post-test questionnaires were administered to gather quantitative and qualitative feedback on user experience and satisfaction.

3.2.2. Tasks:

With the following 10 tasks our usability test aimed to replicate realistic scenarios and interactions a pilot might encounter while using the duty roster application. Instructions were clear and concise, providing contextual information without explicitly stating the steps required to complete the task. The order of tasks was randomized for each participant to minimize learning effects.

1. *Locate a colleague's PK-number:* This task simulated a situation where the user had to report a colleague's sickness but had forgotten their PK-number.
2. *Find the pick-up time:* This task replicated a common scenario where the user had to check the pick-up time for transportation on the next morning.
3. *Identify the date of the next pilots meeting:* This task reflected a user's need to find information about upcoming events or meetings.
4. *Determine the deadline for submitting schedule requests:* This task simulated a user's need to manage their schedule preferences and deadlines.
5. *Retrieve flight information:* This task replicated a scenario where the user had to access specific flight details for check-in.
6. *Access duty roster changes:* This task simulated a user's need to stay updated on changes to their duty roster.
7. *Customize display settings:* This task allowed users to personalize the application's display settings to suit their preferences.

8. *Find information about holiday capacity*: This task replicated a user's need to access information about holiday scheduling and availability.
9. *Confirm crew dispositions*: This task simulated a user's need to acknowledge and confirm crew dispositions received at the hotel.
10. *Verify flight times*: This task replicated a user's need to check and confirm flight times recorded in their personal flight book.

3.2.3. Additional tapping task:

An additional tapping task was incorporated to increase the workload and simulate cognitive conditions that are more similar to a cockpit environment, where the application would typically be used. This secondary task involved participants tapping with their non-dominant hand at a steady rhythm while completing tasks on the mobile application [25]. The tapping task also aimed to identify areas of high cognitive load and potential user frustration within the application [26]. Tapping was recorded by the same webcam that captured a participant's iPhone screen. The post-test analysis then examined those areas of the mobile application where tapping slowed down or stopped, driven by the assumption that slow or erratic tapping indicates an additional cognitive burden or even overload. This approach was complemented by a post-test think-aloud analysis, where participants verbalized their thoughts and actions retrospectively, providing insights into their cognitive processes during application usage.

3.2.4. Retrospective think-aloud analysis:

The so-called think-aloud analysis [27] involves participants verbalizing their thoughts while using an interface, offering insights into their cognitive processes. Yet, as such would have interfered with our tapping task, we decided to use a retrospective think-aloud analysis (RTA) [28]. This allowed users to complete the tasks first and then retrospectively verbalize their thought processes, minimizing potential cognitive overload during the task.

3.2.5. Questionnaires:

To complement the qualitative feedback coming from the RTA, we used both a post-task and a post-test questionnaire. The Single Ease Question (SEQ) [29] served as the post-task questionnaire, capturing initial impressions on the perceived task complexity, and the System Usability Scale (SUS) [30] was employed as the post-test questionnaire to assess the overall usability experience, identify likes and dislikes, and gather improvement suggestions.

4. Findings

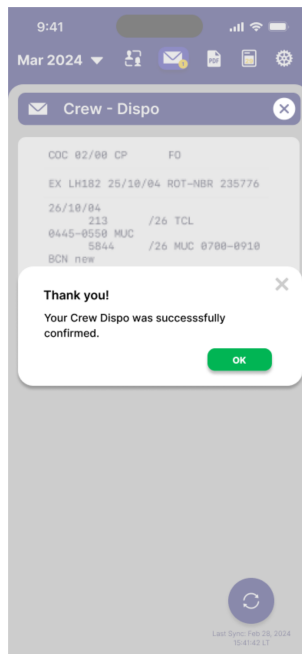
Following we report on our insights from designing the mobile duty roster application for pilots in accordance with the heuristics by Da Costa et al. [24] and the leanings that came from the subsequently conducted usability tests.

4.1. Usability Heuristics as Prototyping Guidelines

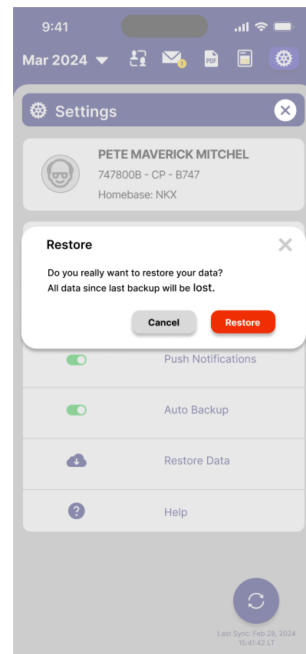
As outlined above, we used the heuristics for mobile application usability by Da Costa et al. [24] as guidelines for the design of our duty roster application prototype. These heuristics emphasize various aspects of usability, which were implemented as follows:

4.1.1. Visibility of system status:

Implemented through confirmation messages and clear visual cues to indicate state changes, such as using green color for successful actions (cf. Fig. 1a).



(a) Confirmation messages.



(b) Pop-up window.

Figure 1: Confirmation messages and pop-up windows inform about the system status and help prevent errors.

4.1.2. Correspondence between the application and the real world:

Achieved through intuitive scrolling behavior, familiar icons, and logical information presentation (cf. Fig. 2b).

4.1.3. User control and freedom:

Provided through easily accessible exit options and toggle buttons for settings (cf. Fig. 2a).

4.1.4. Consistency and standards:

Maintained through a consistent color scheme, standardized icons, and familiar terminology (cf. Fig. 2b).

4.1.5. Error prevention:

Implemented through confirmation pop-up windows for critical actions and careful button placement to avoid accidental clicks (cf. Fig. 1b).

4.1.6. Minimize the user's memory load:

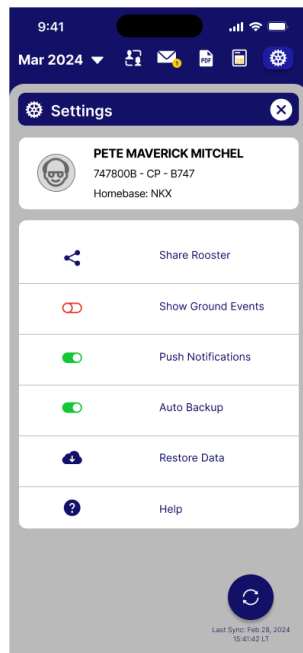
Achieved through a clear information hierarchy and progressive disclosure, revealing details as needed rather than overwhelming the user (cf. Figs. 3a & 3b).

4.1.7. Efficiency of use and performance:

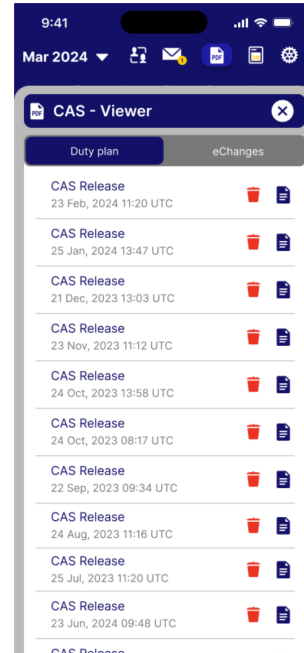
Optimized by minimizing clicks for frequent actions and ensuring smooth transitions.

4.1.8. Aesthetic and minimalist design:

Achieved through compact information presentation, clear visual hierarchy, and limited menu options (cf. Figs. 3a & 3b).



(a) Toggle buttons.



(b) Recycle bin.

Figure 2: Familiar icons such as toggle buttons and the recycle bin foster the intuitiveness of an interface.

4.1.9. Help users recognize, diagnose, and recover from errors:

Implemented through clear messages with simple language and relevant icons (cf. Fig. 1b).

4.1.10. Help and documentation:

Provided through a dedicated help area accessible within the app.

4.1.11. Pleasant and respectful interaction with the User:

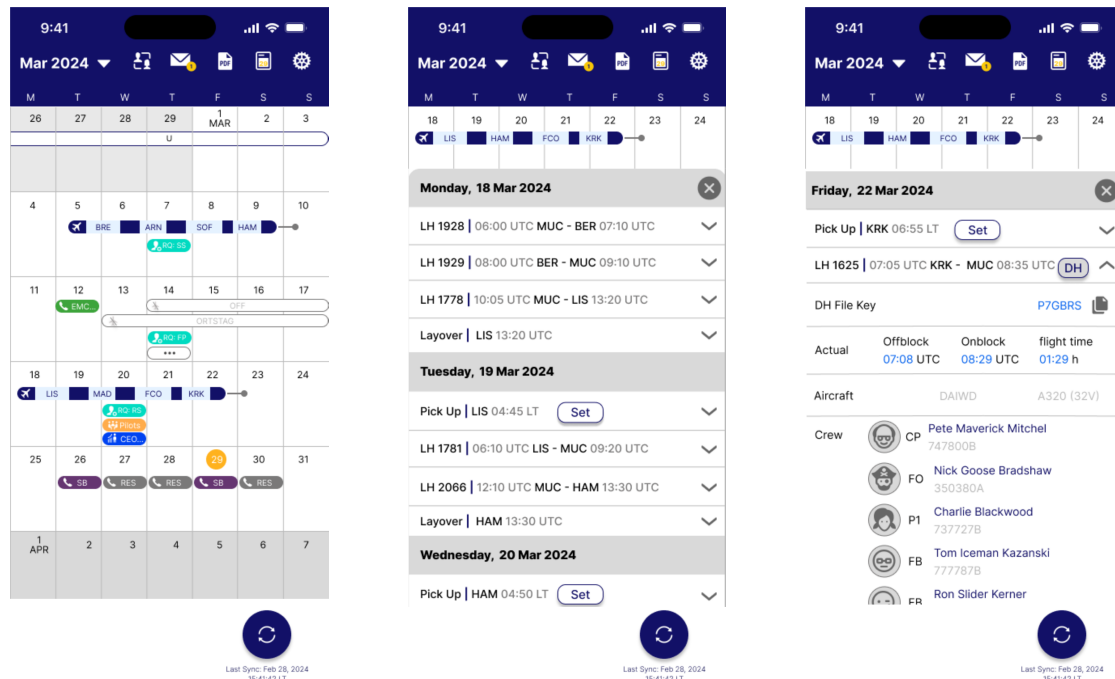
Achieved through familiar design elements, efficient interaction, and accessible button placement.

4.1.12. Privacy:

Addressed by leveraging the device's built-in security features and limiting the display of sensitive information.

4.2. Learnings from the Usability Tests

In order to test the usability of our designed prototype, $n = 5$ pilots were asked to complete a series of tasks designed to evaluate various aspects of the application, ranging from basic navigation and information retrieval to more complex functions like managing schedule changes and vacation requests (cf. Section 3.2.2). As outlined above, we employed a mixed-methods approach, combining observation of user interactions, retrospective analysis where participants verbalized their thought processes, and physiological data in the form of a tapping task to assess cognitive load during task completion. Analysis results show that, overall the application was well-received by our participants. Tasks involving core functions, such as finding crew lists (Task 1) and identifying pick-up times (Task 2), were found to be intuitive and easy to navigate. Similarly, features like toggling the display of ground events (Task 7) and confirming crew rosters (Task 9) were perceived as straightforward and user-friendly. However, certain aspects of the application's design and functionality elicited critical feedback. For example, several participants encountered difficulties with accurately selecting specific days within multi-day flight duty



(a) Overview screen.

(b) Details screen.

(c) More details.

Figure 3: Details are revealed when required so as to not overwhelm users.

events (Task 1), highlighting a potential issue with the application's scrolling functionality and date selection precision. The inclusion of a 'set button' next to the pick-up time (Task 2) caused confusion for one participant who misinterpreted its purpose (cf. Fig. 3c). Furthermore, the icons representing specific events, such as pilot meetings (Task 3) and roster deadlines (Task 4), were not always deemed sufficiently clear and informative. Also, the use of abbreviations within these icons was not universally understood (cf. Fig. 3c). Challenges were also encountered when accessing information related to 'Dead Head Flights' (Task 5) due to the small size and placement of the relevant button. Similarly, navigating schedule changes (Task 6) led to confusions, as the time it took to update the duty roster was perceived as too long. The most significant challenges, however, were encountered in Task 8, which involved locating information about vacation requests. To successfully complete this task, it was necessary to click on the three dots icon in the calendar, providing additional entries for a day, among which was also the vacation request, i.e. "U-RQ" (cf. Fig. 3a, March 20). Unfortunately, this connection between vacation requests and the duty roster application was not apparent, leading to irritation and frustration among participants. Subjective workload assessments, measured via the SEQ, indicated that most tasks were perceived as easy or very easy, with Tasks 1, 2, 3, 7, and 9 receiving particularly high SEQ scores (note: the average task completion time was 30.46 seconds). In contrast, Task 8 received a significantly lower SEQ score as only 3 of the 5 participants were able to complete this task. Overall, however, our interface prototype was found to be intuitive, illustrated by an average SUS score of 92 out of 100 points (95% CI; 87.4|96.6). The physiological data from the tapping task generally corroborated this participant feedback. That is, tasks that elicited comments expressing confusion or difficulty were often accompanied by observable disruptions in participants' tapping rhythm, indicating increased cognitive load. This was particularly evident during Task 8, where the frequent interruptions in tapping rhythm reflected struggles with navigation and comprehension. Overall, we may thus argue that our study has provided valuable insights into the usability of our application prototype, revealing some challenges which require further refinement. That is, while overall the application prototype demonstrated strong functionality in core areas, it was shown that several interface elements, iconography, and task flows would benefit from optimization so as to enhance user experience and minimize cognitive load.

5. Conclusions

The development of our pilot duty roster application prototype was guided by Da Costa et al.'s heuristics for mobile applications [24], which emphasize minimizing cognitive load. We believe that this approach was beneficial, as may be seen by the positive usability test results and participant feedback. The application of these heuristics furthermore triggered the integration of various system status messages, confirmations, and error messages, which may have contributed to a potentially smoother and more user-friendly experience.

However, the usability test also revealed areas for improvement. While most tasks were completed without major difficulties, certain features and design elements caused confusion or challenges for some participants. Specifically, scrolling functionality, icon clarity, abbreviation usage, and the interface for managing vacation requests require further refinement. These findings highlight the importance of empirical testing in identifying usability issues that may not be apparent during the initial design phase, even when guided by established heuristics.

Despite these minor issues, the overall usability of the prototype was rated highly. The positive SUS scores and SEQ ratings suggest that it was perceived as user-friendly and easy to learn. The observed learning effects between similar tasks further support this conclusion. The additional tapping task provided further insights into cognitive load, revealing subtle fluctuations that may not have been captured by subjective measures alone.

In conclusion, one may therefore argue that the use of heuristics proved valuable in guiding the development of this pilot duty roster application prototype. However, usability testing remained crucial for identifying and addressing more specific areas for improvement.

6. Limitations & Future Outlook

While the heuristics proposed by Da Costa et al. [24] served as a valuable starting point for the development of the pilot duty roster application prototype, their broad interpretability and limited scope necessitate further considerations. Involving multiple developers in the heuristic evaluation process could have enhanced the identification of potential usability issues, which might have ensured a more comprehensive assessment. Additionally, heuristics, while helpful, may not fully address the nuances of specific design challenges, such as minimizing cognitive load. Thus, developers should have supplemented their understanding of these heuristics with additional knowledge and design expertise.

As for the evaluation method, the remote nature of our usability tests facilitated detailed analysis of user interactions, yet limited the observation of people's body language. The additional tapping task proved to be a sensitive measure of cognitive load fluctuations but should be interpreted cautiously and in conjunction with other methods, as rhythmic tapping may not always reflect genuine understanding or task success. Finally, while the SEQ and SUS questionnaires provided valuable quantitative data, the small sample size requires careful consideration when interpreting their validity.

A broader analysis with more participants is required to tackle these limitations. Such is planned for when the identified interface issues have been addressed. Furthermore, we plan to explore usability within the broader context of the pilot application ecosystem, examining the interplay between multiple applications and their impact on workflow and cognitive load in and around the cockpit.

Declaration on Generative AI

During the preparation of this work, the authors used Gemini 2.5 Flash in order to summarize text fragments and check them for grammar and spelling. After using this tool, the authors reviewed and edited the content as needed and thus take full responsibility for the publication's content.

References

- [1] K.-H. Huarng, T. H.-K. Yu, W. Lai, Innovation and diffusion of high-tech products, services, and systems, 2015. URL: <https://www.sciencedirect.com/science/article/pii/S0148296315002295>. doi:<https://doi.org/10.1016/j.jbusres.2015.06.001>.
- [2] M. Graham, W. H. Dutton, Society and the internet: How networks of information and communication are changing our lives, Oxford University Press, 2019.
- [3] M. Mazmanian, Avoiding the trap of constant connectivity: When congruent frames allow for heterogeneous practices, *Academy of Management journal* 56 (2013) 1225–1250.
- [4] J. Rennecker, L. Godwin, Delays and interruptions: A self-perpetuating paradox of communication technology use, *Information and Organization* 15 (2005) 247–266.
- [5] J. Barjis, A. Gupta, R. Sharda, Knowledge work and communication challenges in networked enterprises, *Information Systems Frontiers* 13 (2011) 615–619.
- [6] S. A. Grandhi, Q. Jones, S. R. Hiltz, Technology overload: is there a technological panacea?, *AMCIS 2005 Proceedings* (2005) 493.
- [7] P. Yin, C. X. Ou, R. M. Davison, J. Wu, Coping with mobile technology overload in the workplace, *Internet Research* 28 (2018) 1189–1212.
- [8] L. Rosen, A. Samuel, Conquering digital distraction, *Harvard business review* 93 (2015) 110–113.
- [9] P. Karr-Wisniewski, Y. Lu, When more is too much: Operationalizing technology overload and exploring its impact on knowledge worker productivity, *Computers in Human Behavior* 26 (2010) 1061–1072.
- [10] M. A. Orhan, S. Castellano, I. Khelladi, L. Marinelli, F. Monge, Technology distraction at work. Impacts on self-regulation and work engagement, *Journal of Business Research* 126 (2021) 341–349.
- [11] A. F. Stokes, C. D. Wickens, Aviation displays, in: *Human factors in aviation*, Elsevier, 1988, pp. 387–431.
- [12] J. Buranatrived, P. Vickers, An investigation of the impact of mobile phone and PDA interfaces on the usability of mobile-commerce applications, in: *Proceedings 3rd IEEE International Workshop on System-on-Chip for Real-Time Applications*, IEEE, 2002, pp. 90–95.
- [13] G. Joyce, M. Lilley, Towards the development of usability heuristics for native smartphone mobile applications, in: *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 3*, Springer, 2014, pp. 465–474.
- [14] N. Ahmad, A. Rextin, U. E. Kulsoom, Perspectives on usability guidelines for smartphone applications: An empirical investigation and systematic literature review, *Information and Software Technology* 94 (2018) 130–149.
- [15] Y. S. Lee, S. W. Hong, T. L. Smith-Jackson, M. A. Nussbaum, K. Tomioka, Systematic evaluation methodology for cell phone user interfaces, *Interacting with Computers* 18 (2006) 304–325.
- [16] J. Heo, D.-H. Ham, S. Park, C. Song, W. C. Yoon, A framework for evaluating the usability of mobile phones based on multi-level, hierarchical model of usability factors, *Interacting with Computers* 21 (2009) 263–275.
- [17] M. Clamann, D. B. Kaber, Applicability of usability evaluation techniques to aviation systems, *The international journal of aviation psychology* 14 (2004) 395–420.
- [18] S. Feinberg, M. Murphy, Applying cognitive load theory to the design of web-based instruction, in: *18th Annual Conference on Computer Documentation. IPCC SIGDOC 2000. Technology and Teamwork. Proceedings. IEEE Professional Communication Society International Professional Communication Conference*, IEEE, 2000, pp. 353–360.
- [19] D. S. K. Seong, Usability guidelines for designing mobile learning portals, in: *Proceedings of the 3rd international conference on Mobile Technology, Applications & Systems*, 2006, pp. 25–es.
- [20] J. Pitkänen, Mobile application usability research: Case study of a video recording and annotation application, 2016.
- [21] G. Singer, Filling the gaps in the human factors certification net, in: *Coping with Computers in*

the Cockpit, Routledge, 2018, pp. 87–108.

- [22] E. S. Stein, A critical component for air traffic control systems, in: Human factors in Certification, CRC Press, 2000, pp. 73–80.
- [23] D. Wixon, C. Wilson, The usability engineering framework for product design and evaluation, in: Handbook of Human-Computer Interaction, Elsevier, 1997, pp. 653–688.
- [24] R. P. Da Costa, E. D. Canedo, R. T. De Sousa, R. D. O. Albuquerque, L. J. G. Villalba, Set of usability heuristics for quality assessment of mobile applications on smartphones, IEEE Access 7 (2019) 116145–116161.
- [25] Y. Miyake, Y. Onishi, E. Pöppel, Two types of anticipation in synchronization tapping, Acta Neurobiologiae Experimentalis 64 (2004) 415–426.
- [26] J. P. Tracy, M. J. Albers, Measuring cognitive load to test the usability of web sites, in: Annual Conference-society for technical communication, volume 53, 2006, p. 256.
- [27] J. Nielsen, Usability Engineering, Morgan Kaufmann, 1994.
- [28] Z. Guan, S. Lee, E. Cuddihy, J. Ramey, The validity of the stimulated retrospective think-aloud method as measured by eye tracking, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2006, pp. 1253–1262.
- [29] A. Assila, H. Ezzedine, et al., Standardized usability Questionnaires: Features and Quality Focus, Electronic Journal of Computer Science and Information Technology 6 (2016) 15–31.
- [30] J. Brooke, SUS: A Retrospective, Journal of usability studies 8 (2013) 29–40.