



Article

Usability Evaluation of Wearable Smartwatches Using Customized Heuristics and System Usability Scale Score

Majed A. Alshamari ¹ and Maha M. Althobaiti ^{2,*}

¹ Information System Department, College of Computer Sciences and Information Technology, King Faisal University, P.O. Box 400, Hofuf 79820, Saudi Arabia; smajed@kfu.edu.sa

² Department of Computer Science, College of Computing and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

* Correspondence: maha_m@tu.edu.sa

Abstract: The mobile and wearable nature of smartwatches poses challenges in evaluating their usability. This paper presents a study employing customized heuristic evaluation and use of the system usability scale (SUS) on four smartwatches, along with their mobile applications. A total of 11 heuristics were developed and validated by experts by combining Nielsen's heuristic and Motti and Caines' heuristics. In this study, 20 participants used the watches and participated in the SUS survey. A total of 307 usability issues were reported by the evaluators. The results of this study show that the Galaxy Watch 5 scored highest in terms of efficiency, ease of use, features, and battery life compared to the other three smartwatches and has fewer usability issues. The results indicate that ease of use, features, and flexibility are important usability attributes for future smartwatches. The Galaxy Watch 5 received the highest SUS score of 87.375. Both evaluation methods showed no significant differences in results, and customized heuristics were found to be useful for smartwatch evaluation.

Keywords: heuristic evaluation; system usability scale; smartwatches; wearable devices; usability evaluation



Citation: Alshamari, M.A.; Althobaiti, M.M. Usability Evaluation of Wearable Smartwatches Using Customized Heuristics and System Usability Scale Score. *Future Internet* **2024**, *16*, 204. <https://doi.org/10.3390/fi16060204>

Academic Editor: Stefano Rinaldi

Received: 26 April 2024

Revised: 15 May 2024

Accepted: 31 May 2024

Published: 6 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The relationship between humans and technology is best exemplified through wearable technology, which increases human potential through the provision of smart gadgets that can help to monitor and track body parameters. Wearable devices provide continuous monitoring of an individual's health, activity, and fitness, which has significant potential for enhancing human activities and standard of living. There are sensors embedded in these devices that gather personal information and exchange this information through Wi-Fi, Bluetooth, cellular technology, etc. [1]. Wearable technologies offer a wide range of functionalities for unobtrusive health monitoring, such as heart rate tracking, blood sugar tracking, blood pressure analysis, gait analysis, sleep score, as well as a variety of other health factors. These devices are the subject of extensive research [2] and can measure physical activity, such as steps walked, calories burnt, or workout intensity, using a bracelet-like gadget worn on the wrist. Then, the data are transmitted to a mobile application, either wirelessly through Bluetooth synchronizing or through the connection of the device to a smartphone, where objectives, progress, and activity may be recorded [3]. The most widely used wearable devices are wrist-worn smartwatches, which can receive and send notifications, messages, or calls. For example, the Apple watch series can receive incoming calls and has a comparatively large screen, but is more expensive than other smartwatches. The mainstream wrist-worn devices include the Apple watch, Samsung Gear S, Mi Band, Huawei Honor, Fitbit Surge, and Jawbone Up3, with many others available for purchase.

Currently, the major aspects of wearable devices are extensively researched, such as the dependability and accuracy of evaluation. Recently, researchers have raised concerns

regarding the long-term usage of wearable devices, highlighting the importance of combining behavioral change approaches such as goal setting, feedback, and rewards with other evidence-based techniques [4]. In addition, whilst wearable technology has a wider range of potential applications, users do face usability issues with these devices [5]. Furthermore, studies have found that the acceptability of wearable items for regular consumers, as well as their comfort level, is critical to a product's success [6]. For these reasons, it is necessary to identify and address the usability issues of wearable devices because the success of these devices relies on users' product experience. As user experience is one of the core aspects contributing to a product's popularity, it is crucial that the device performs its functionality in an easy, comfortable, and intuitive way. Therefore, continuous usability testing is essential to observe users as they interact with a product or prototype and ensure a satisfactory user experience, contributing to the widespread and quick adoption of wearable devices.

Usability is a multifaceted concept that can be explored in various ways. The term "usability" has been used extensively for the past few decades, and various individuals interpret it differently. Some people associate usability with ease of use or convenience and examine it from the standpoint of an interactive interface; meanwhile, others refer to usability as a conceptual scale for the real-time evaluation of a product's functionality obtained from feedback from potential users [7]. Therefore, despite the testing paradigm, usability testing assesses a product's ability to fulfill its intended functions. Food, consumer items, websites or web applications, computer interfaces, papers, and electronics are all examples of products that benefit from usability testing.

There are various factors in developing usability criteria. According to Nielsen, usability has five key characteristics: learnability, satisfaction, efficiency, low mistake rate/quick error recovery, and memorability [8]. The International Standards Organization (ISO) describes the effectiveness and efficiency of, and satisfaction with, a product as usability parameters [9]. Shackel defined usability attributes as effectiveness, learnability, flexibility, and user attitude [10]. According to Hix and Hariston, the usability parameters are initial performance, long-term performance, learnability, retainability, advanced feature usage, first impressions, and long-term user satisfaction [11]. Furtado defined the usability parameters as ease of use and learning [12]. The parameters defined by ISO and Nielsen are commonly used for usability evaluation [13]. There are several usability evaluation techniques. Any approach or technique used to perform usability evaluation or testing to enhance the usability of an interactive system at any point of its development is known as a usability evaluation method (UEM). Several usability testing methods have been introduced, such as laboratory-based formative assessment with users, heuristics, questionnaires, and other expert-based usability evaluation techniques, model-based analytic approaches, all types of expert assessment, and the remote evaluation of interactive software after field deployment.

The factors that influence the adoption of smart wearable devices among potential customers are wearability, ease of use, compelling design, functionality, and price. Wearability refers to the existence of pain, degree of comfort, ease in wearing, requiring support to wear, or willingness to wear again. Meanwhile, ease of use means no interference in daily activities, comprehensibility, and learnability. Compelling design means visually appealing. Functionality refers to the core features. Although conventional usability evaluation methods such as thematic analysis, heuristic evaluation, interviews, and think-aloud and cognitive walkthroughs can be used, the main shortcoming of these methods is that the findings in a laboratory can sometimes be difficult to illustrate [14]. Alternatively, the Agency for Healthcare Research and Quality has suggested that questionnaires can be effectively used for usability evaluation [15]. The most widely used questionnaires are the system usability scale (SUS) and UTUAT, the post-study system usability questionnaire covering use, satisfaction, ease, etc.

The dependability and accuracy evaluation of wearable devices are currently the focus of extensive research. Concerns have been raised relating to the use of wearable devices, and researchers have indicated the importance of combining behavioral change approaches

such as goal setting, feedback, and rewards alongside other evidence-based techniques [15]. A mixed response to adopting wearable devices has also been observed, with the response of users being less positive than expected [5]. In addition, whilst wearable technology represents a significant range of potential applications, usability issues still remain to be solved by research [5]. Two key features have also been identified that are critical for a product's success: the acceptability of the items to regular consumers and their comfort level [6].

Therefore, the usability issues of wearable devices remain to be solved, as their success depends on users' experience; therefore, it is critical that a wearable device is easy and intuitive to use and comfortable to wear. As a result, continuous usability testing is essential for manufacturers to ensure the widespread and rapid adoption of their particular product. This testing observes users interacting with the product or prototype and is a key factor in ensuring a satisfactory user experience.

The key aim of this paper is to investigate the usability and other issues relating to the existing commercially available wearable smartwatches. The usability will be analyzed based on the customized heuristic evaluations and the system usability scale (SUS). These usability evaluations of smartwatches will provide a comprehensive, user-specific, and customized measure of usability approach, which will provide quantitative and statistical data for benchmarking and further design improvements.

2. Literature Review

Smartwatches have received significant attention for their versatility, which satisfies a wide range of consumer interests, including fitness and health monitoring [4]. Other than basic health and fitness features, smartwatches offer a wide range of features varying from person to person, such as real-time vital signs and overall health monitoring in senior people with Parkinson's, heart disease, or other chronic conditions. In addition, they may capture both crucial and trivial data regarding patient location and behavior more quickly and precisely [16]. According to recent IDC surveys on wristwatch usage, the industry will continue to expand exponentially, with 373 million units expected to be shipped in 2020, up from 100 million in 2016. In 2016, smartwatches made up 25% of all smart wearables. Although smartwatches have a wide range of functionalities, they still face technical and usability challenges, such as aged people being less familiar with new technologies or ease of use or comfort issues for patients [17]. The usability testing needs to examine how consumers really use their smartwatches rather than observing how they were intended to be used. The usability test aims to understand consumers' requirements of these devices through the examination of usability concerns connected to the real tasks that users undertake using their smartwatches [12].

Despite many studies exploring the usability, perceived value, and role of smartwatches in patient monitoring, only limited research has evaluated the relationship of usability and brand factors influencing usability using different usability evaluation methods. This section provides a summary of the fragmented research on wearable technology. Y. Wu et al. [18] proposed a novel method for evaluating a smartwatch's usability based on eye movement tracking. The eye tracker recorded the testers' eye movements, and the eye movement data were added to the system for calculating the usability rating index. In the study, 10 participants were asked to perform specific tasks with Motorola 360 smartwatches and were interviewed afterwards. The results of the task test showed that eye movement data can accurately assess the icons on smartwatch interfaces and illustrate how users search for certain features. J. Chun et al. [19] completed a study to identify challenges in smartwatch usability using in-depth interviews. Many users appeared to find mobile devices convenient for checking pushed alerts. However, smartwatches obtained a poor response when individuals tried to interact with visually rich material. According to the study, individuals like to use smartwatches to listen to music and check the weather. The study of N. Anggraini et al. [20] observed how usability, brand, and price influenced customers' impressions of smartwatches in Indonesia. In order to conduct the

study, 116 Indonesian respondents were surveyed. The participants were less concerned with usability and instead focused on brands and pricing when making their purchases. Most of them did not consult evaluations and recommendations from others when they already had a brand in mind. However, one of the shortcomings was the limited number of participants involved, as with a greater number of participants, the results could be generalized to Indonesia's smartwatch users.

M. Bang et al. [21] designed a nurses' watch app with the Motorola 360 smartwatch in order to automate patient monitoring and checklist systems. The usability of the IT systems, including complexity, training, and need for support for the user interface, was evaluated using the system usability scale (SUS). The mean score in the SUS resulted in the average score as the question on comfort and security from the SUS questionnaire received very poor marks. Neuropsychiatric diseases are the primary cause of disability globally, but current mental health monitoring methods rely on subjective DSM-5 specifications; however, developments in EEG and video monitoring technology have not been extensively embraced because of inconvenience. Kamdar and Wu [22] presented a novel platform—the Passive, Real-time Information for Sensing Mental Health (PRISM)—through the integration of heart rate, light, and motion data from a smartwatch application and text input from a web application. The SUS questionnaire was used to evaluate its usability, and a total of 13 healthy participants were asked to wear the Samsung Gear S smartwatch for usability evaluation. The SUS questionnaire demonstrated that participants had a positive attitude toward PRISM. The participants said that the system was simple to use, requiring little expertise or training; however, they lacked the motivation to use it regularly.

In the study of C.R. Laborde et al. [23], the authors evaluated user satisfaction, usability, and compliance with the help of a real-time, online assessment and mobility monitoring (ROAMM) mobile app designed for smartwatches. In the study, 28 participants were asked to wear smartwatches and fill out a standardized questionnaire. The ROAMM wristwatch app received high marks from older people with knee osteoarthritis, indicating their satisfaction with the app. The condition of atrial fibrillation (AF) is difficult to diagnose since it often presents with mild symptoms. Smartwatches can be used for long-term, non-invasive monitoring, which could improve AF care. The main objective of the study of E.Y. Ding et al. [24] was to evaluate the efficacy of arrhythmia discrimination using a wristwatch. A total of 40 participants were observed, and a questionnaire was completed evaluating several aspects of the device's usability. A real-time algorithm was used to analyze the pulse recordings. The results showed that the majority of participants thought the smartwatch was very usable. The general level of comfort and the level of data privacy when wearing a wristwatch for rhythm monitoring were both positively correlated with younger age and past cardioversion, respectively. The participants deemed the smartwatch to be extremely acceptable despite their age, lack of knowledge of smartwatches, and a significant load of comorbidities. Although smartwatches may be potential tools for atrial fibrillation identification, elderly stroke patients have not been given enough attention.

E. Dickson et al. [25] introduced the pulse watch, a smartwatch-based AF detection system, and evaluated its precision, usability, and adherence in stroke patients. The participants filled out questionnaires to evaluate numerous psychosocial factors and health-related behaviors. The results demonstrated that older stroke patients found the scheme useful and would stick to the monitoring schedule. Smartwatches have also been demonstrated to accurately assess blood pressure in glaucoma patients; however, their usability evaluation has been neglected in previous research. For this purpose, S.B. Bhanvadia et al. [26] conducted an experiment where adult participants received a wristwatch blood pressure monitor for indoor monitoring using the Omron BP monitor and an associated mobile app. Usability testing methodologies included the post-study system usability questionnaire (PSSUQ), which was used for assessing aspects of user satisfaction such as the overall system usefulness, information quality, and interface quality, and the system usability scale (SUS), which was used to assess the overall usability of smartwatches. Furthermore, usability on the basis of age, gender, and race was also evaluated. The usability evaluations

demonstrated that the smartwatches had satisfactory usability ratings, although older age was linked to lower levels of perceived usability and user experience. Table 1 shows the current state-of-the-art methods to evaluate the usability of smartwatches.

Table 1. A summary of the literature review on usability evaluation of smartwatches.

Reference # ¹	Objective	#. Of Participants ²	Evaluation Method	Evaluation Heuristics	Smart Watch
[21]	Design and evaluation of a smartwatch platform for ICU nurses	7	SUS, interviews	Comfort, security, user interface	Unknown
[25]	Usability of smartwatch-based AF detection system in stroke patients	-	Questionnaires	Accuracy, usability, adherence	Unknown
[26]	Usability of BP on glaucoma patients	53	SUS, PSSUQ,	Overall, system usefulness, information quality, and interface quality	Omron heart guide
[23]	Usability of ROAMM mobile app for individuals with knee osteoarthritis	28	Standardized questionnaire	Satisfaction, usability, compliance	Unknown
[24]	Usability of pulse irregularities among older participants	40	SUS	Rhythm analysis, ease of use, user importance of daily activities, comfort, privacy, and associated stress	Unknown
[20]	Users' perception of Smartwatch for Indonesian customers	116	Online survey	Usability, brands, and prices	N/A
[19]	To identify users' needs/requirements regarding the usability of smartwatches for the representative tasks	30	Interim surveys	Info display, control, learnability, interoperability, preference	LG, Sony, Pebble, Samsung Gear, Moto 360
[18]	Usability evaluation method of smartwatch-based eye movement tracking app	10	Customized questionnaires	Smartwatch interface	Moto360
[22]	Usability test of mental health monitoring platform	13	SUS	PRISM platform	Samsung Gear
Proposed work	Usability evaluation of smartwatches	10, 20	Customize Heuristics, SUS	Self-descriptiveness, consistency and standards, aesthetic and minimalist design, reduce short-term memory, match between system and real world, error tolerance, privacy, satisfaction, ease of use, features, etc.	Fitbit Versa 2 and Charge 5, Samsung Galaxy 4 and 5

¹ Reference Number. ² Number of participants who participate in the study.

The interface of smartwatches is one of the main factors contributing to a satisfactory user experience. Everything visualized on the screen should be clear and self-explanatory so that users do not become confused. However, this is limited by its size. The real dilemma is to identify usability issues while improving the user experience for smartwatches. The buyers' judgment of the level of satisfaction for a certain product might vary. Smartwatches are particularly adapted to record the variances in mood, exhaustion, sleep quality, etc., remotely and conveniently as experienced by users. These are the factors that influence the usability of smartwatches, and a single heuristic method alone may be insufficient when evaluating the usability of smart devices, which also require more context-specific heuristics. However, as Table 1 shows, researchers have infrequently used heuristics for the usability evaluation. In addition, there is only one evaluation method in each of the studies; however, if we conduct usability evaluations using more than one method, we may reveal more smartwatch usability issues. Therefore, in this study, we sought to evaluate the

usability of four different smartwatches which includes Samsung Galaxy watch 5, Samsung Galaxy watch 4, Fibit Charge 5, and Fibit Versa 2. Our main contributions are the following:

- We used two evaluation methods for the usability evaluation of these watches to find the technical, design issues, etc., in the real world. One is heuristic evaluation with customized heuristics, and the other is using the SUS score. The details are provided in Sections 3 and 4.
- We developed a total of 11 customized heuristics through a combination of Neilson's heuristics for the user interface evaluation of watches with Motti and Caine's [27] heuristics, which are specifically utilized for the usability evaluation of wearable devices, i.e., smartwatches.
- A total of 10 usability evaluators used the smartwatches for 10 days and completed the customized survey. Alongside the customized heuristics questionnaire, a further 20 users completed the SUS (system usability scale) questionnaire.

3. Usability Evaluation Methods

3.1. Heuristic Evaluation

Heuristic evaluation, proposed by Nielsen and Molich, is a method of usability engineering that employs standard heuristics to evaluate smart devices' usability and to identify design problems [28]. Its classical form involves multiple evaluators exploring and investigating the product separately and listing the problems encountered while interacting with the system. Finally, they assign a severity score to each problem and present a report with suggested solutions to the identified problems [28]. Nielsen and Molich's heuristics are a common, effective tool for usability evaluation, are widely recognized and have been successfully applied to various digital interfaces, desktop applications, and website interfaces [28].

To qualify the usability problems identified in the questionnaires, various research has employed a severity scale that graded the magnitude of the identified usability concern, making it possible to detect problems that prevent smart devices' proper functioning [29]. Figure 1 shows Jakob Nielsen's widely recognized set of 10 usability principles that were used to improve user interface through identifying usability problems [30]. These heuristics principles have been developed by Nielsen and Molich through the introduction of heuristic evaluation for usability inspection [31].

Heuristic (1)	Heuristic (2)	Heuristic (3)	Heuristic (4)	Heuristic (5)
Visibility of system status	Match between the system and the real word	User control and freedom	Consistency and standards	Error prevention
Heuristic (6)	Heuristic (7)	Heuristic (8)	Heuristic (9)	Heuristic (10)
Recognition rather than recall	Flexibility and efficiency of use	Aesthetic and minimalist design	Help users recognize, diagnose, and recover from errors	Help and documentation

Figure 1. Jakob Nielsen's 10 heuristic evaluation criteria [30].

Some researchers have suggested, however, that Nielsen and Molich's heuristics may be insufficient or unsuitable for evaluating smartwatches' usability [32], as the screen size, input capability, battery life, and interaction time of wearable interfaces is limited, yet their heuristics can provide a starting point in the usability evaluation to supplement heuristics or guidelines specifically tailored to smartwatch interfaces. In 2014, Motti and Caine designed a specific heuristic for wearable devices with a human-centered focus [27]. Their heuristics also consider the aesthetics of smartwatch interfaces, which are crucial for user acceptance and satisfaction. Motti and Caine's 20 heuristics for human-centered wearable devices are mostly applicable to the tested device [27]. One or more heuristics

can be attributed to each problem identified. Table 2 shows the 20 principles enumerated by Motti and Caine.

Table 2. Motti and Caine’s 20 heuristics for wearable devices.

ID	Heuristic Name
H1	Aesthetics
H2	Affordance
H3	Comfort
H4	Contextual Awareness
H5	Customization
H6	Ease of Use
H7	Ergonomy
H8	Fashion
H9	Intuitiveness
H10	Obtrusiveness
H11	Overload
H12	Privacy
H13	Reliability
H14	Resistance
H15	Responsiveness
H16	Satisfaction
H17	Simplicity
H18	Subtlety
H19	User-friendliness
H20	Wearability

Aesthetics embrace the attractiveness of any wearable object [33]; an attractive design improves its desirability [34]. Affordance relates to a device’s intuitiveness in terms of physical interactions between the user and the device [35]. Comfort requires an acceptable temperature, texture, shape, weight, and tightness and implies freedom from pain and discomfort [36]. After wearing a device for some time, users should be adequately comfortable and no longer feel it [37]. Contextual awareness, embracing the scenarios in which wearable devices are used, must be clearly understood and considered during the design process.

Comfort is strongly affected by a device’s purpose [38] and varies significantly by social context. Humans vary in shape, size, and dimensions, as well as in preferences, interests, and wishes. Thus, wearable devices’ look and feel should allow customization, accounting for factors such as users’ sensibilities, wishes, and interests [39]. Ease of use recognizes the need for a simple, straightforward, intuitive interface [38], which enhances device usability and increases user engagement. Input and output interfaces should be easy to use [40]. Ergonomy refers to a device’s physical shape, constraints, ergonomic aspects related to bodily anatomy, and how users perceive it [41].

Fashion includes the perception of a wearable device’s desirability [38]. In other words, it indicates how stylish the technology is and contributes to its becoming more (or less) ubiquitous. Intuitiveness describes how interactions occur, such as those involving the existing buttons, keys, commands, and features [38]. This heuristic applies the concept of affordance to the cognitive aspects of interactions. Physiological sensors have various degrees of intrusiveness, which may involve using body tissue to diagnose a given physiological state or condition. Devices that are non-intrusive are often obtrusive and, to some extent, cumbersome. Devices should be anatomically transparent [27], allowing natural body movements, and the design should duly consider the human body’s anatomical characteristics and constraints.

Unlike technology, which has continually grown in capacity, humans have a finite processing capacity and can perform only a limited number of concurrent activities before experiencing cognitive overload, posing a distinct challenge to designers of wearable devices. A mobile interface that does not account for human cognitive capabilities in its

design may hinder a user's primary task [38]. Privacy refers to how confidential interactions can be performed when using the device [42]. Reliability describes users' extent of trust and confidence in a device [36]. Resistance involves understanding the context in which a wearable device is used so as to improve its resistance to specific types of wear; this heuristic helps practitioners identify acceptable levels of resistance, with special consideration given to impact, temperature, humidity, flexure, and laundering. To ensure durability, devices must stand up to wearing and cleaning [37].

Users tend to be less patient on the move than at a desk, making it imperative to give them feedback in near real-time, thus offering outstanding system responsiveness [38]; users' efficiency and productivity are enhanced when they can be responsive to their tasks [39]. Satisfaction concerns how a device meets users' expectations, wishes, and requirements. A device's simplicity embraces ease of use, intuitiveness, and affordance, enabling users to interact more efficiently through straightforward interaction options and necessary feedback [38]. The principles of minimalistic design are respected by including only those features and interaction options that are fundamental to accomplishing available tasks. Subtlety refers to the transparency of the device's communication; for example, notifications for the device's owner should not disturb bystanders. In other words, notifications should not cause social problems [43]. User-friendliness accommodates the mental model of the end user, proposing options that easily and intuitively facilitate interaction. Recovery from errors should be possible [36]. Wearability considers objects' physical shapes and devices' active relationships with the human form [38].

The combination of Nielsen and Molich's and Motti and Caine's heuristics has helped to categorize and formulate better solutions to smartwatches' usability problems.

3.2. SUS Score

The system usability scale (SUS) is a widely used, standardized questionnaire that assesses perceived usability. In 2009, researchers reported that 43% of post-study questionnaires used in industrial usability studies incorporated the SUS [44]. Survey respondents rate statements about the system's complexity and whether they believe training or support is needed to use it effectively. According to Brooke [45], this simple, 10-item post-test questionnaire quickly assesses a product's usability without requiring complicated analysis. Table 3 presents the questionnaire.

Table 3. The system usability survey questionnaire.

#	Questions	Strongly Disagree 1	2	3	4	Strongly Agree 5
1	I think that I would like to use this system frequently.					
2	I found the system unnecessarily complex.					
3	I thought the system was easy to use.					
4	I think that I would need the support of a technical person to be able to use this system.					
5	I found the various functions in this system were well integrated.					
6	I thought there was too much inconsistency in this system.					
7	I would imagine that most people would learn to use this system very quickly.					
8	I found the system very cumbersome to use.					
9	I felt very confident using the system.					
10	I needed to learn a lot of things before I could get going with this system.					

In general, the SUS can quickly, easily, and comprehensively evaluate usability. The SUS comprises 10 closely related questions on a 5-point Likert scale, of which questions 1, 3, 5, 7, and 9 are positive, and 2, 4, 6, 8, and 10 are negative. A higher SUS score implies better product usability.

Evaluating the usability criteria requires expertise in both the customization of usability principles and their specific domain for wearable smartwatches. So, evaluators need a thorough understanding of all the challenges and considerations that should be kept in mind to develop effective heuristics. In turn, this involves extensive research, analysis, and iteration to ensure the heuristics accurately reflect the requirements and expectations of smartwatch users. Moreover, participants involved in the heuristics should be representative of the target user population, i.e., those who are professionals or fitness enthusiasts, which is a challenging task. Thus, the data acquisition, its analysis, and interpretation is a complex problem that is resolved in this study through careful planning and comprehensive analysis of the evaluation methods.

4. Proposed Methodology

In this section, we outline our research design and describe the study conducted for the evaluation of smartwatch usability. The human–computer interface (HCI) has been using usability evaluation methods since the early 1980s, spurred on by the need to improve the usability of smart devices and applications. Usability testing methods include field studies, laboratory experiments, expert-based inspection methods, etc. [46]. There is only scarce research that compares the evaluation methods for smartwatches; however, there have been studies investigating wearable usability using different methods. Therefore, in this research, we will be testing the usability of four smartwatches using two usability evaluation methods. We conducted a heuristic evaluation and used a system usability scale to evaluate the usability of the four smartwatches, comprising the Samsung Galaxy Watch 4, Samsung Galaxy Watch 5, Fitbit Charge 5, and Fitbit Versa 2, along with their mobile applications, Gear Samsung for Samsung Watches, and the Fitbit mobile application for Fitbit watches. For the heuristic evaluation, we used customized heuristics combining Nielsen’s 10 heuristics [47] and Motti and Caine’s 20 heuristics [27,28] designed and validated by HCI experts.

4.1. Smartwatches and Applications to Be Evaluated

In this study, we use the following four watches to investigate their usability with their mobile applications:

1. Samsung Galaxy Watch 4, Galaxy Wearable (Samsung Gear) app;
2. Samsung Galaxy Watch 5, Galaxy Wearable (Samsung Gear) app;
3. Fitbit Charge 5, Fitbit application;
4. Fitbit Versa 2, Fitbit Application.

The Samsung Galaxy Watch 4 (Figure 2a) was released in August 2021 alongside other Samsung watches but became the most popular due to some significant changes, including health and fitness features [48]. This device is equipped with an optical heart rate sensor, an electrical heart rate sensor, and a bioelectrical impedance analysis (BIA) sensor, which provides data to improve health. The advanced BIA sensor measures body skeleton mass, body water, body fat mass, and body mass index (BMI) and provides insights to the user to help them manage their health more effectively [49]. Samsung introduced this after the significant concerns of increasing obesity rates, not just among seniors but also among young adults and children. This has been a particular feature of the post-COVID-19 lockdowns around the world [50].

The Samsung Galaxy Watch 5 (Figure 2b) was released in August 2022 and included a new temperature feature through the addition of an infrared temperature sensor. This can determine the user’s basal body temperature and uses it as a baseline to determine different changes in body temperature. Moreover, the battery life has been improved to last 50 h due to the powerful Exynos-W920 chipset along with 1.5 GB RAM and 16 GB

storage. In addition to existing body monitoring sensors, the Samsung Galaxy Watch 5 watch provides improved sleep tracking. The Galaxy Watch 5 was the most popular watch for Samsung in 2022.



Figure 2. (a) Samsung Galaxy Watch 4; (b) Samsung Galaxy Watch 5; (c) Fitbit Charge 5; (d) Fitbit Versa 2.

Fitbit Charge 5 (Figure 2c) was also released in August 2021 and is considered the best fitness tracker among Fitbit watches due to its improved features, including heart rate monitoring, sleep tracking, stress-management tools, GPS tracking, and activity tracking. The Fitbit Versa marked a move into smartwatch territory with a much larger, squared-off screen and a few extra features beyond health and fitness tracking. Fitbit Versa 2 is the latest device released by Fitbit Inc. The Fitbit Versa 2 improved on its predecessor with a raft of updated features, including Alexa support, better sleep tracking, and Fitbit Pay on all models [51]. Fitbit is a popular brand of wearable trackers [52]. Fitbit Versa 2 and its value-price works for anyone who is intrigued with the idea of a smartwatch and, thus, has served 27.6 million users worldwide in 2018, selling over 13.9 million units [53]. More recent statistics state that Fitbit had 31 million users by the end of 2020 [54].

Fitbit Versa 2 (Figure 2d) is the most affordable Fitbit smartwatch. As these watches were only released in recent years but are used widely, only a handful of articles cover their functionality; therefore, we selected these widely used smartwatches for our usability evaluation.

4.2. Study Design for Heuristic Evaluation

A customized heuristic evaluation was performed using Nielsen's 10 heuristics and Motti and Caine's [27] 20 principles, with 10 evaluators analyzing the usability of the four smartwatches and their four mobile applications. The Nielsen and Molich study originally found that three to five evaluators were sufficient for detecting the majority of usability issues; however, this number remains under debate [55]. For this reason, we chose 10 evaluation experts in the field of usability testing to perform the heuristic evaluations. Our group's HCI experts chose this variation in the heuristic evaluation based on the fact that it proved to be simpler and faster. For the usability evaluation of the interfaces of smartwatches and their mobile applications, we used Nielsen's heuristics. Motti and Caines's heuristics are design decisions toward the human-centered aspects in the wearable domain. Combining these sets of principles, the HCI experts designed a questionnaire comprising a total of 11 heuristics, as shown in Table 2, with details of each heuristic; these are relevant to the context in which our evaluation is conducted. In terms of user interfaces, qualities such as self-descriptiveness, consistency and standards, aesthetic design, reducing short-term memory load, and matching the system with the real world are important. Additionally, interface evaluation is crucial for usability evaluation. The interface is the

primary means through which users interact with a smartwatch. Its design and usability greatly impact the user experience. Ease of use, flexibility, and efficiency of use and features evaluate the intuitiveness, simplicity, and availability or functionality of specific features the smartwatch provides, respectively [18]. These heuristics provide valuable guidelines for assessing different aspects of the user experience. HCI experts validated their evaluation findings through multiple discussion sessions.

The evaluators assigned a severity level to each problem they found in either the interface, the design of a mobile application, or the smartwatch. An estimation of how much more usability work may be needed was also based on the severity ratings, with the results informing the decision on the allocation of resources to the most serious problems. The severity ratings are a combination of how many times a problem is occurring, its impact, and its persistence [56]. Therefore, for the overall assessment of the usability problems, a rating scale was used that combined all three factors to facilitate decision-making, as presented in Table 4. The severity rating value scale can be seen in Table 4. This severity rating scale followed the tradition established in previous research studies using the system usability scale (SUS) in order to maintain consistency with previous research. The severity rating scale used in our study is a widely studied and common approach documented in academic papers [57]. This method ensured that our evaluation results were consistent with current usability research standards. While the severity rating scale employed in our study's design was adapted from existing studies, we also ensured that it was appropriate and fit our study aims. The evaluators also provided comments on why they allocated this severity level to a particular problem and suggested possible solutions.

Table 4. Severity scale for heuristic evaluation.

Scale	Description
1	Not a usability problem, no flaws, and issues in the application and watch
2	Cosmetic usability problem, minor issues that can remain unnoticed
3	Minor usability problem but low priority
4	Major usability problem, usability improvements are required
5	Usability catastrophe, must solve the issue

4.2.1. Heuristic Customization

Due to the radical changes in technology over the years, evaluating usability with the 10 heuristics (HEs) of Nielsen and Molich alone could not provide sufficient insights. In addition, because smart devices demand more context-specific heuristics rather than a generic set, the usability requirements of different user interfaces, audiences, and tasks require customized heuristics. In the process of tailoring the heuristics to the specific interface and task, the evaluation can be more efficient due to orienting the focus to the most relevant usability issues [58]. Based on the existing information in different studies, scant research on the customization of usability heuristics using these methods is available. Therefore, we combined two heuristic principles: one for the interface usability analysis and the other specifically designed for wearable devices. The HCI experts of our group designed 11 heuristics for the evaluation of smartwatches and applications. The researchers state that a more human-centered approach to wearable design must take into account human aspects, facilitating consideration of human factors during design phases [27]. These human factors would include users' needs, preferences, and expectations; moreover, human-centered wearable design contributes to a better user experience. As an innovative approach to design, human-centered design starts with understanding users' perspectives and designing accordingly [59]. Therefore, our selection of six heuristics (visibility of system status, match between system and real-world, consistency and standards, aesthetic and minimalist design, flexibility and efficiency of use, and error tolerance) were sourced from Nielsen's heuristics, while Motti and Caine's heuristics (privacy, wearability, comfort, ease of use, and satisfaction) are also relevant and helpful in the design process of wearables and mobile applications. These are the heuristics that were unanimously agreed on in this

study and used in the subsequent evaluation process. They were further reviewed by three human–computer interaction (HCI) experts to ensure the consistency and validation of these heuristics. In order to provide a comprehensive assessment and evaluation of HEs, each item was broken down into multiple items. These 11 heuristics are defined below:

- **Self-descriptiveness:** This heuristic is used to evaluate whether the design is user-friendly or not. Each screen header should define the purpose of that screen; additionally, the user should be able to know where they are when looking at the screen header and the sentences and words that have been used, i.e., when looking at the time, the header of that screen should be self-explanatory, or when the user touches on a button, there should be visual feedback that describes that action has been noted. For instance, if users delete something, successful deletion, or that object disappearing, provides the visual feedback. In total, we designed five items to obtain the mean severity value for this heuristic.
- **Consistency and standards:** This heuristic identifies the need for user interface elements across different tasks or across different versions to be consistent in their language, icons, and symbolism. We designed four items specifically related to this heuristic.
- **Aesthetic and minimalist design:** The Neilson and Norman group [4] presented aesthetics and minimalism as their eighth usability heuristic. This principle was summarized by Donald Norman as follows: interfaces should not contain unnecessary information or information that is rarely required. The evaluators were given six items under this heuristic.
- **Reduce short-term memory load:** Short-term memory load can be reduced by showing pull-downs and menus on screens. As Nielsen says, the more you recognize something, the easier it is to remember it. Thus, it is important that objects, actions, and options can be accessed by the user in a way that minimizes the user's memory load. We presented three items for this category.
- **Match between the system and the real world:** This heuristic is the second usability principle presented by Nielsen and indicates that the language of the system should be the same as that of the user. The words and phrases should be simple to understand. Experts can score three items with severity ratings under this category in the questionnaire.
- **Error tolerance:** If the user makes a mistake, there should always be a way to go back instead of penalizing the user. Systems should also provide options for user customization.
- **Privacy:** The brand should ask for permission before collecting their user's personal data.
- **Ease of use:** This heuristic represents the capacity of the smartwatch to let its users perform tasks effectively and quickly while relishing the experience. This is the heuristic which fundamentally influences the adoption of smartwatches.
- **Flexibility and efficiency of Use:** This represents the speed with which the system responds to user actions or requests. We designed a total of six items under this heuristic.
- **Features:** This requires investigation as to whether these smartwatches provide basic functionality with good accuracy or are only focused on the design. Evaluator's seven items available for scoring that relate to the time a task took, the response time of the system, etc.
- **Overall satisfaction:** This heuristic evaluates if the user is satisfied with the application and the smartwatch's performance, features, etc.

4.2.2. Procedure for Performing Heuristic Evaluation

A total of 10 evaluators were recruited to perform the usability evaluation. They were HCI experts and had evaluated other interfaces and applications. This study was conducted within a controlled environment in a soundproof laboratory. Throughout this study, the laboratory remained vacant, devoid of any extraneous disturbances. The researchers ensured that every participant's mobile phone was connected to the internet

and their watches were charged and working properly before conducting usability sessions. The smartwatches and their corresponding mobile applications were evaluated by the 10 evaluators for 10 days, and evaluations lasted up to 1 h every day. The evaluators used the applications and smartwatches and wrote daily notes about usability problems they identified. After 10 days of use, they completed the designed questionnaire, giving each problem a severity rating and also providing comments on how that problem could be solved or the reason underlying the allocation of this severity number. The expert evaluators investigated the user interface of mobile applications and smartwatches, as well as the hardware design, the features of each watch, and their assessment of the data accuracy.

Finally, we performed heuristic evaluation calculations using Equation (1) [60], as follows:

$$\sum Hx = (x1) + (x2) + (x3) + (x4) + \dots + (xn) \quad (1)$$

where $\sum Hx$ = total rating score on all sub-aspects of each usability heuristic; $x1, x2, x3, \dots, xn$ are the usability ratings on each question of that heuristic. The severity rating of each usability aspect 2 was calculated using Equation (2):

$$\text{Severity} = \sum (Hx/n) \quad (2)$$

where n is the number of usability sub-aspects using the severity rating value to indicate the magnitude of the problems identified.

4.3. Study Design for Usability Evaluation Using SUS Score

In addition to the heuristic evaluation, we chose SUS for usability evaluation as it has been widely adopted in the usability evaluation of products. Its versatility makes it useful for the usability evaluation of mobile devices and wearables using a short 10-item questionnaire. A total of 20 participants who had no prior experience with Samsung and Fitbit watches were recruited; their consent was obtained, and they were given the smartwatches. The participants were all aged between 20 and 30 years old. They were asked to use the smartwatches for 30 days. At the end of the 30-day period, they were invited to complete an evaluation SUS questionnaire to rate their experiences with these products. For the descriptive statistical analysis, their basic information, including age and gender, was collected and are presented in the Results section. We calculated the SUS score as follows [61]:

1. The respondent's response to statements of odd numbers was minus by 5 (see Equation (3));
2. Then, 25 was subtracted from the respondent's response to statements of even numbers (see Equation (4));
3. The results from Equations (3) and (4) were summed up and multiplied by 2.5 (see Equation (5))

$$X = (Q1, Q3, Q5, Q7, Q9) - 5 \quad (3)$$

$$Y = 25 - (Q2, Q4, Q6, Q8, Q10) \quad (4)$$

$$\text{SUS_Score} = (X + Y) \times 2.5 \quad (5)$$

5. Results and Discussion

In this section, the results from the questionnaire and findings are discussed. The usability evaluations were performed on both wearable devices and apps, using both the heuristic evaluation and SUS. This section also provides a comparison between the results of this research with the current research. We conducted the heuristic evaluation using 10 evaluators who used the watches for 10 days and completed the questionnaires, adding, if necessary, brief comments or explanations about problems and their solutions. In the group of evaluators, six were females and the rest were males. Three of the users were

Samsung watch users, and seven were Fitbit users. In our study, the Fitbit and Samsung Watch users were randomly selected as evaluators. Therefore, the distribution of Samsung Watch users and Fitbit users was coincidental. For the SUS survey, 20 users used the watches for 30 days and completed the SUS questionnaire. Of the users, 15 were male and 5 were female. None of them had any prior experience with the Samsung Galaxy Watch 4, Galaxy Watch 5, Fitbit Charge 5, or Fitbit Versa 2.

5.1. Evaluation Method 1: Heuristic Evaluation Results

The HCI experts were given the task of installing the application on mobile phones, connecting it to the watches, and analyzing them for usability issues. A total of 20 usability principles were evaluated by each of the 10 experts independently. We compiled the usability issues encountered by each usability expert to create this heuristic evaluation report and present it in Tables 5 and 6. These tables show the number of criteria violated per severity rating for each heuristic listed. Three hundred and seven (307) usability issues were identified in all four watches, where one-hundred and nine (35.5%) were discovered in the Fitbit Charge 5 with twenty-two minor, nineteen major, and fifteen disastrous problems. However, only 46 (14.9%) usability issues were reported for the Samsung Galaxy Watch 5. Out of 307, 66 issues (21.4%) were reported in the Galaxy Watch 4, and the remaining 86 (28%) were reported in the Fitbit Versa 2. Most of the problems were located in H8 “Ease of Use”, H9 “Flexibility and Efficiency of use”, and H10 “Features”.

Table 5. Usability issues encountered by evaluators with severity for the Galaxy watches.

ID	Heuristics	Severity				Frequency	Severity				Frequency
		Galaxy Watch 4					Galaxy Watch 5				
		Cosmetic	Minor	Major	Disastrous		Cosmetic	Minor	Major	Disastrous	
H1	Self-Descriptiveness	3	0	0	0	3	1	0	0	0	1
H2	Consistency and Standards	2	0	0	0	2	4	0	0	0	4
H3	Aesthetic and Minimalist Design	1	2	1	0	4	2	1	0	0	3
H4	Reduce Short Term Memory Load	4	1	3	0	8	1	1	1	0	3
H5	Match Between System and Real-world	5	1	0	1	7	2	0	0	0	2
H6	Error Tolerance	7	0	2	0	9	6	1	0	0	7
H7	Privacy	2	0	0	0	2	2	0	0	0	2
H8	Ease of Use	8	1	2	0	11	7	1	0	0	8
H9	Flexibility and Efficiency of Use	1	2	1	0	4	3	2	0	0	5
H10	Features	2	5	0	0	7	4	2	0	0	6
H11	Satisfaction	4	4	1	0	9	4	1	0	0	5
	Total	39	16	10	1	66	36	9	1	0	46

Table 6. Usability issues encountered by evaluators with the severity for the Fitbit watches.

ID	Heuristics	Severity				Frequency	Severity				Frequency
		Fitbit Charge 5					Fitbit Versa 2				
		Cosmetic	Minor	Major	Disastrous		Cosmetic	Minor	Major	Disastrous	
H1	Self-Descriptiveness	4	1	0	0	5	3	0	0	0	3
H2	Consistency and Standards	2	2	0	0	4	4	4	0	0	8
H3	Aesthetic and Minimalist Design	4	2	1	1	8	2	1	0	3	6
H4	Reduce Short Term Memory Load	6	2	5	1	14	3	3	1	0	7
H5	Match Between System and Real-world	9	3	2	1	15	5	2	1	0	8
H6	Error Tolerance	3	1	2	3	9	6	1	2	2	11
H7	Privacy	2	1	0	0	3	2	0	0	0	2
H8	Ease of Use	11	3	4	2	20	7	1	3	1	12
H9	Flexibility and Efficiency of Use	1	2	1	4	8	3	2	2	3	10
H10	Features	6	2	3	1	12	4	2	3	2	11
H11	Satisfaction	5	3	1	2	11	4	1	2	2	8
	Total	53	22	19	15	109	43	17	14	13	86

The Galaxy Watch 4 demonstrates huge advances in its hardware and software due to its Exynos W920 architecture, Samsung’s first 5 nm wearable processor [61]. There are two Cortex-A55 cores on the Exynos W920, along with a Mali-G68 MP2 GPU, designed to deliver 20 percent faster CPU performance and 10 times faster GPU performance than the Exynos W9110. In addition to higher power efficiency, the 5 nm processor also results in longer smartwatch battery life. A dedicated low-power processor on the Exynos W920

handles always-on displays and other tasks while consuming very little power [62]. This architecture also manages heart rate and notifications in the background more easily. The heuristic evaluations observed that it has a great UI design as compared to other watches. The Galaxy Watch 4's UI design combines the look and feel of Samsung's Tizen platform and Wear OS [63]. The watch has a virtual rotating bezel, two buttons on the right-hand side, and a colorful and intuitive interface that allows for easy navigation and customization. Watch 4 has a redesigned notification system that allows more actions and interactions, including a more consistent and seamless experience across Samsung devices, such as syncing settings, installing apps, and transferring data. The customizable watch face editor allows you to choose from various customization options, colors, and styles. The simplified settings menu is easier to navigate and adjust [64], again allowing user customization, including gesture features that are so sensitive that even moving the wrist can swipe up or down the layout on the watch.

A few evaluators commented that its haptic bezel is not very smooth. The one other design feature is the access strap tucked underneath the wrist strap; therefore, there is no extra flap that can bother the user, dangle, or become lost. This is conclusively a good design approach and increases its wearability. Samsung Galaxy Watch 5 (GW5) has the same chipset, RAM, and storage as in the GW4. The main difference is its battery size. The watch has 284 milliampere per hour (MAH), which offers 40 h of battery life. However, it was noted by evaluators that it has 100% battery and stays live for around 32 h, with a touch-sensitive bezel for easy navigation, as in the GW4. Another difference is the Bluetooth version, which is upgraded to 5.2. The GW4 uses a C-type cable to charge and provides a 45% charge in 30 min. According to the evaluators, it also has sapphire glass upgraded from gorilla glass, which provides a more solid and expensive feel.

Fitbit Versa 2 only weighs 40 gm, which makes it comfortable to wear; however, the Samsung watches weigh less. Versa 2 provides a voice input feature in the watch to set timers and reminders using Alexa. The evaluators observed that its voice recognition was more than 95% effective. The evaluators found completing tasks with it was quite easy and fast, and the system information was understandable, as shown in Table 5. For example, the Fitbit obtained the highest score for the heuristic relating to flexibility and efficiency of use. On the other hand, the Fitbit Charge 5 is a discrete fitness tracker band, that is simple to use and with a bright display, but at the expense of the battery. The Fitbit is also easily wearable in daily life as it is very light, has auto activity detection, works accurately, and runs an electrodermal activity (EDA) sensor instead of a button, making it more aesthetically pleasing. Table 5 clearly shows that minimum usability issues were found in the Galaxy Watch 5. A report on the heuristic evaluation was produced based on the usability issues encountered by each usability expert in Tables 5 and 6. Out of 307 usability issues, 109 (35.5%) were found in the Fitbit Charge 5, with 22 minor, 19 major, and 15 disastrous problems. However, only 46 (14.9%) usability issues were reported for the Samsung Galaxy Watch 5, 66 (21.4%) were reported for the Galaxy Watch 4, and the remaining 86 (28%) were reported for the Fitbit Versa 2.

The evaluators mostly mentioned self-descriptiveness problems as cosmetic issues, indicating that this was not something requiring immediate attention for these watches. Most problems identified as being located in H8 as "Ease of Use", H9 was "Flexibility and Efficiency of Use", and H10 was "features". The usability problems are explained above. Smartwatches from Samsung and Fitbit require user consent before storing any personal data. Samsung takes privacy seriously and seeks user consent before collecting personal data through their devices, including smartwatches. As part of the set-up process for Samsung or Fitbit smartwatches, the user is required to agree to the terms and conditions. These terms include information about data collection and usage. The brands collect data such as your name, email, location, health metrics, payment information, etc. These brands also state that they ask for your consent before collecting or sharing your data [65]. In addition, the evaluators reported that they could manage app permissions to control what data they wanted to provide. The usability problems are explained above. The evaluators

have also provided comments regarding usability and design issues and the better features of smartwatches, and we categorized those comments into positive and negative comments, which are presented in Tables 7 and 8. Table 7 presents the comments on Samsung watches and Table 8 presents those for Fitbit watches.

Table 7. Participants' comments on Galaxy smartwatches sorted by the usability principle.

Heuristics	Positive Comments	Negative Comments
Aesthetic and Minimalist Design	- GW5 has a versatile design and scratch-resistance screen - New watch faces in GW5 are great	- GW5 should have a physical rotating bezel as well
Match Between System and Real World	- GW5 is responsive - GW5 application loads quickly	- Need initial learning in order to be familiar with smartwatch
Error Tolerance	- No comment	- GW4 updates make touch display non-functional - GW4 error messages closed faster than the time to read them
Ease of Use	- The GW4 rotating bezel is a handy navigation tool - Typing keyboard is better in GW5	- No comment
Flexibility and Efficiency of Use	- Phone calls sound clean on GW5	- Samsung Watch 4 disconnects from the watch and does not automatically reconnect - Automatic email notifications are annoying on GW4
Features	- GW5 has skin temperature sensor, which I found cool. - Bigger battery in GW5 - Six evaluators were satisfied with the accuracy of sleep and heart monitoring features in GW5	- GW4 has inconsistent battery life - GW5 has most of the features that are the same as Watch 4

Table 8. Participants' comments on Fitbit smartwatches sorted by usability principle.

Heuristics	Positive Comments	Negative Comments
Aesthetic and Minimalist Design	- Fitbit application is not too complicated or distracting	- Six evaluators commented that the Fitbit Charge 5 wristband seems loose on their wrists, hence, they find it difficult to wear it every day
Match Between System and Real World	- The language is understandable and the information is accurate	- Charge 5 does not have a voice assistant
Error Tolerance	- Charge 5 allows going back to the previous menu from anywhere	
Ease of Use	- In Versa 2, alarm setting and reminders are easy	- Versa 2 does not pair after resetting - Charge 5 does not have a microphone or speaker
Features	- The Fitbit watch feature to monitor the intake of calories sounds great - Heart rate and all other features are present in FB Charge 5 despite the small size - Versa 2 allows use of Alexa	- Versa 2 sleep tracking is sometimes worrying
Satisfaction	- Fitbit Charge 5 is comfortable to wear - Versa 2 has a better battery life	- Charge 5 battery drains rapidly - Heart rate monitoring requires immediate improvements

Negative comments relate to the usability problems, and it is observed that Samsung watches need improvement for quick and efficiency synchronization with the mobile applications. Table 7 shows that the GW5 has more positive comments than the GW4 from the evaluators. Although some improvements are still required, i.e., relating to the accuracy

of activity tracking features in the GW5 because four of the evaluators were dissatisfied with the results. We also observed that Charge 5 has design issues, missing voice input, and with the speakers and voice assistant; however, Versa 2 has a better battery life and a voice assistant that makes setting reminders and alarms easier. From the results in Table 6 and the comments from Table 8, we concluded that Fitbit Versa 2 has fewer usability problems than Fitbit Charge 5 according to the evaluators. However, the Galaxy Watch 5 still has the lowest number of usability problems compared to the other three watches. We present these usability problems in Table 9 with their severity rating value.

Table 9. Usability problems with their severity value based on comments.

Rating Usability Problems	Samsung Galaxy Watch 4, Galaxy Wearable (Samsung Gear) App	Samsung Galaxy Watch 5, Galaxy Wearable (Samsung Gear) App	Fitbit Charge 5, Fitbit Application	Fitbit Versa 2, Fitbit Application
1	- Gear Mobile Application is fine	- Fits best on the wrist - Bigger and better battery life - Fast charging - Has versatile design and scratch-resistance screen - mobile application is fine	- none	- Alexa feature makes it quite interactive
2	- Character limit of notifications on the watch should increase - Sleep tracking requires improvement	- Sleep tracking can be finer and more accurate	- none	- Battery life needs improvements
3	- No option to change the font size from the watch - GW4 error messages closed faster than the time to read them	- Display does not respond to touches on 3rd party watch-face	- No voice assistant - Setting alarms and reminders by hand	- Sleep tracking is sometimes worrying
4	- Updates make touch display non-functional on watch.	- Have to update and activate Google assistant every time after booting up watch	- Not for everyday use as it keeps falling off - synching watch is problematic: no automatic synching	- Does not pair automatically after resetting - synching errors
5	- Battery drains fast. - After update, watch does not work properly.	- A few times, watch does not record steps correctly	- Rapid battery discharge - Problems with Bluetooth connection - Heart rate monitoring requires immediate improvements	- Touch screen is not functional after updates

5.2. Evaluation Method 2: SUS Evaluation Results

An SUS score above 68 is considered above average, and a score below 68 is considered below average; thus, 68 is the average score, making it the 50th percentile [55]. An SUS score greater than 80 is considered excellent, while a score below 51 is considered an awful design in terms of efficiency, effectiveness, and overall ease of use. Once the 20 participants filled out the SUS survey for each watch, we calculated the mean for each watch regarding that item number (#), as shown in Table 10. GW5 obtained the highest SUS score of 87.375 among all four watches, and was considered as possessing excellent usability.

The battery life of these devices is one of the most significant obstacles to their acceptance and usability in consumer markets [66]. With a battery life of up to 50 h, the GW5 is the most useful in daily life, even though it is almost identical to GW4. In addition to improving battery life, the GW5 has some features that make it more appealing, such as an aluminum metal frame with a sapphire crystal display. However, GW5 is much more expensive than GW4. Moreover, the SUS score in Table 7 shows Fitbit Versa 2 has a greater usability score than Fitbit Charge 5. In the Samsung Galaxy range, the Galaxy Watch 5

is better. From these results, we also analyzed that the latest launch of the GW5 had a higher usability value than other Samsung watches (GW4). However, in the Fitbit series, Fitbit Versa 2 was more efficient than the Fitbit Charge 5. Galaxy has higher values in odd questions, which indicates that the users mostly agree that this system is easy to use and efficient, and lower values in even questions indicating that they did not find the system unnecessarily complex and inefficient, as compared to Fitbit.

Table 10. System usability score for each watch.

#	Item	GW4	GW5	FB Charge 5	FB Versa 2
1	I think that I would like to use this system frequently.	4.45	4	4	3.95
2	I found the system unnecessarily complex.	1.55	1.6	1.95	1.9
3	I thought the system was easy to use.	4.05	4.15	4.15	4.15
4	I think that I would need the support of a technical person to be able to use this system.	1.25	1.25	1.7	2
5	I found the various functions in this system were well integrated.	4.75	4.8	4.35	4.2
6	I thought there was too much inconsistency in this system.	1.35	1.3	1.35	1.45
7	I would imagine that most people would learn to use this system very quickly.	4.45	4.65	4.4	4.3
8	I found the system very cumbersome to use.	1.4	1.4	1.6	1.6
9	I felt very confident using the system.	4.15	4.2	3.8	3.75
10	I needed to learn a lot of things before I could get going with this system.	1.3	1.3	1.35	1.3
	SUS Score	86.875	87.375	81.875	82.75

From these evaluation results, we concluded that the customized heuristics and SUS both favored Galaxy Watch 5 because the evaluators identified fewer usability issues, and there were more positive comments among all four watches in terms of usability. Hence, this also demonstrated that the customized heuristics validated the heuristics for the usability evaluation of watches. In this work, we overcame the limitations of previous research with a combination of different evaluation methods and the customization of 11 heuristics for usability evaluation.

We performed a comparative analysis of our findings with the previous research to find if we have generalized values with other studies. The Galaxy Watch 4's haptic bezel was identified as not being very smooth, which sometimes made navigation difficult. As a result, the user lost control and freedom, and the system became intolerable. The findings were similar to those found in other studies [67], where users found it difficult to navigate between pages when applications lacked backward and forward navigation buttons. As a result, this aspect of the design made it difficult to navigate a smartwatch. Fitbit Versa 2 does not pair automatically after resetting, thus violating the heuristic of ease of use. In addition, Fitbit Charge 5 does not support voice assistants because it does not have a microphone and speaker, making alarm settings difficult and violating the heuristic of ease of use. In a previous study [68], it was shown that ease of use in the system improves the quality experience of the user with that system.

The results of the SUS survey for smartwatches showed that the users found them easy to learn and use, resulting in improved performance. This aligns with previous studies [69], which also found the systems easy to learn. The Watch 5 was found to be particularly effective for everyday use due to its long battery life, customization options, and wearability, as also reported by the authors of [70]. Overall, the usability results were positive for the effectiveness of the watch applications. However, issues were reported after updates, which confirmed the findings of earlier studies by the authors of [69,71] that the applications were efficient for users. Additionally, the participants reported that they would easily remember how to use the watches in the future, a finding supported by the authors of [72,73]. The users also encountered fewer errors while using the watch applications, a result consistent with earlier studies by [71,72,74]. The users were also satisfied with the features, functionalities, design, information, and display quality of the Watch 5 among all four watches, as reported in studies by [69,74,75]. The Galaxy watches were also found to

be easy to use, leading to user satisfaction, as reported in studies by [75,76]. The new haptic bezel feature of the Galaxy watch was highly rated for its aesthetic appeal, as also observed in [76]. The Fitbit Charge 5 provided advanced fitness tracking features, as reported in [75], which found the application useful in achieving fitness-related goals.

6. Conclusions

The goal of our study was to evaluate the usability of smartwatches using different usability evaluation methods. To achieve this goal, we inspected four smartwatches, including the Samsung Galaxy Watch 4, the Galaxy Watch 5, the Fitbit Charge 5, and the Fitbit Versa 2 with their respective mobile applications, Samsung Gear and the Fitbit mobile. We employed customized heuristics and the system usability scale (SUS). The customized heuristics employed a collaborative evaluation using Nielsen and Molich's ("Usability 101: Introduction to Usability," n.d.) heuristics and Motti and Caine's [27] principles for the usability of wearable devices. A total of 10 experts used the watches for 10 days and investigated the usability issues that arose from the watches and their applications. Some identified problems were that there was no automatic reconnection between the smartwatch and the application, the automatic settings after updates were reported as annoying, and inconsistent battery life impeded everyday use. Evaluators also reported problems regarding the wearability of the Fitbit because the strap of the watch appeared to be loose. However, the Galaxy Watch 5 obtained the highest usability mean value among all the watches. Most of its usability problems were found in ease of use, flexibility, and efficiency of use.

The system usability survey was completed by 20 participants who had no prior experience with Samsung and Fitbit watches. The results from the SUS score indicated that the Galaxy Watch 5 had a slightly higher SUS score than the Galaxy Watch 4 and thus was rated as having the highest usability. In addition, the Fitbit Versa 2 obtained a higher score value than the Charge 5. We also observed that people rated the GW5 the highest because of the battery life and its classic design. Therefore, this paper conclusively contributes to the usability evaluation of smartwatches, overcoming problems encountered with customized heuristic evaluation and the system usability scale, and identifies the watch with the best usability among all watches. The results achieved in this study will be influential in providing better design features and shaping the user experience of smartwatches as they are continuously gaining popularity among users due to their innovative use of technology. Future work will include extending the parameters of the user evaluation study and investigating how usability evaluation methods can be better adapted to capture relevant usability issues in other wearable devices, i.e., headsets, smart rings, and smart shoes.

Author Contributions: Conceptualization: M.A.A.; methodology: M.A.A.; formal analysis and investigation: M.M.A.; writing—original draft preparation: M.A.A.; review and editing: M.M.A.; supervision: M.A.A.; project administration: M.M.A.; funding acquisition: M.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deanship of Graduate Studies and Scientific Research, Taif University.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to acknowledge the Deanship of Graduate Studies and Scientific Research, Taif University for funding this work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lee, J.; Kim, D.; Ryoo, H.Y.; Shin, B.S. Sustainable Wearables: Wearable Technology for Enhancing the Quality of Human Life. *Sustainability* **2016**, *8*, 466. [CrossRef]
2. Mejia, C.; Ciarlante, K.; Chheda, K. A wearable technology solution and research agenda for housekeeper safety and health. *Int. J. Contemp. Hosp. Manag.* **2021**, *33*, 3223–3255.
3. Lunney, A.; Cunningham, N.R.; Eastin, M.S. Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes. *Comput. Human. Behav.* **2016**, *65*, 114–120. [CrossRef]
4. Sullivan, A.N.; Lachman, M.E. Behavior Change with Fitness Technology in Sedentary Adults: A Review of the Evidence for Increasing Physical Activity. *Front. Public Health* **2016**, *4*, 289. [CrossRef]
5. Jia, Y.; Wang, W.; Wen, D.; Liang, L.; Gao, L.; Lei, J. Perceived user preferences and usability evaluation of mainstream wearable devices for health monitoring. *PeerJ* **2018**, *6*, e5350. [CrossRef] [PubMed]
6. Mercer, K.; Giangregorio, L.; Schneider, E.; Chilana, P.; Li, M.; Grindrod, K. Acceptance of Commercially Available Wearable Activity Trackers among Adults Aged Over 50 and with Chronic Illness: A Mixed-Methods Evaluation. *JMIR Mhealth Uhealth* **2016**, *4*, e7. [CrossRef] [PubMed]
7. What Is Usability Testing? *IxDF*. Available online: <https://www.interaction-design.org/literature/topics/usability-testing> (accessed on 1 February 2023).
8. Usability Engineering—Jakob Nielsen. Google Books. Available online: [https://books.google.com.pk/books?hl=en&lr=&id=95As2OF67f0C&oi=fnd&pg=PR9&dq=Nielsen,+J.+\(1994\).+Usability+engineering.+Morgan+Kaufmann&ots=3cFCzocvXt&sig=HvT1YG47dXgA3BbUFzuwzUfiteA#v=onepage&q=Nielsen,%20J.%20\(1994\).%20Usability%20engineering.%20Morgan%20Kaufmann&f=false](https://books.google.com.pk/books?hl=en&lr=&id=95As2OF67f0C&oi=fnd&pg=PR9&dq=Nielsen,+J.+(1994).+Usability+engineering.+Morgan+Kaufmann&ots=3cFCzocvXt&sig=HvT1YG47dXgA3BbUFzuwzUfiteA#v=onepage&q=Nielsen,%20J.%20(1994).%20Usability%20engineering.%20Morgan%20Kaufmann&f=false) (accessed on 1 February 2023).
9. ISO 9241-11:1998; Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs)—Part 11: Guidance on Usability. ISO—International Organization for Standardization: Geneva, Switzerland, 1998. Available online: <https://www.iso.org/standard/16883.html> (accessed on 1 February 2023).
10. Human Factors for Informatics Usability. Google Books. Available online: https://books.google.com.mt/books?id=KSHrPgLiMJIC&printsec=copyright&source=gbs_pub_info_#v=onepage&q&f=false (accessed on 1 February 2023).
11. Hix, D.; Hartson, R.H. *Developing User Interfaces: Ensuring Usability through Product & Process* (Wiley Professional Computing); John Wiley & Sons: Hoboken, NJ, USA, 1993.
12. Furtado, E.; Vasco Furtado, J.J.; Mattos, F.L.; Vanderdonckt, J. Improving Usability of an Online Learning System by Means of Multimedia, Collaboration and Adaptation Resources. In *Usability Evaluation of Online Learning Programs*; IGI Global: Hershey, PA, USA, 2011. [CrossRef]
13. Seadle, M.; Greifeneder, E. Defining a digital library. *Libr. Hi Tech* **2007**, *25*, 169–173. [CrossRef]
14. Hudson, D.; Kushniruk, A.; Borycki, E.; Zuege, D.J. Physician satisfaction with a critical care clinical information system using a multimethod evaluation of usability. *Int. J. Med. Inform.* **2018**, *112*, 131–136. [CrossRef] [PubMed]
15. Johnson, C.M.; Johnston, D.; Crowley, P.K.; Culbertson, H.; Rippen, H.; Damico, D.; Plaisant, C. *EHR Usability Toolkit: A Background Report on Usability and Electronic Health Records*; Rockville: Agency for Healthcare Research and Quality: Rockville, MD, USA, 2011; pp. 1–68.
16. Lu, T.C.; Fu, C.M.; Ma, M.H.M.; Fang, C.C.; Turner, A.M. Healthcare Applications of Smart Watches. A Systematic Review. *Appl. Clin. Inform.* **2016**, *7*, 850–869. [CrossRef]
17. Kruse, C.S.; Mileski, M.; Moreno, J. Mobile health solutions for the aging population: A systematic narrative analysis. *J. Telemed. Telecare* **2017**, *23*, 439–451. [CrossRef]
18. Wu, Y.; Cheng, J.; Kang, X. Study of smart watch interface usability evaluation based on eye-tracking. In *Design, User Experience, and Usability: Technological Contexts*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2016; Volume 9748, pp. 98–109. [CrossRef]
19. Chun, J.; Dey, A.; Lee, K.; Kim, S.J. A qualitative study of smartwatch usage and its usability. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2018**, *28*, 186–199. [CrossRef]
20. Anggraini, N.; Kaburuan, E.R.; Wang, G.; Jayadi, R. Usability Study and Users' Perception of Smartwatch: Study on Indonesian Customer. *Procedia Comput. Sci.* **2019**, *161*, 1266–1274. [CrossRef]
21. Bang, M.; Solnevik, K.; Eriksson, H. The Nurse Watch: Design and Evaluation of a Smart Watch Application with Vital Sign Monitoring and Checklist Reminders. *AMIA Annu. Symp. Proc.* **2015**, *2015*, 314–319. [PubMed]
22. Kamdar, M.R.; Wu, M.J. PRISM: A data-driven platform for monitoring mental health. In *Proceedings of the Pacific Symposium on Biocomputing, Big Island, HI, USA, 4–8 January 2016*; pp. 333–344. [CrossRef]
23. Laborde, C.R.; Cenko, E.; Mardini, M.T.; Nerella, S.; Kheirkahan, M.; Ranka, S.; Fillingim, R.B.; Corbett, D.B.; Weber, E.; Rashidi, P.; et al. Satisfaction, Usability, and Compliance with the Use of Smartwatches for Ecological Momentary Assessment of Knee Osteoarthritis Symptoms in Older Adults: Usability Study. *JMIR Aging* **2021**, *4*, e24553. [CrossRef] [PubMed]
24. Ding, E.Y.; Han, D.; Whitcomb, C.; Bashir, S.K.; Adaramola, O.; Soni, A.; Saczynski, J.; Fitzgibbons, T.P.; Moonis, M.; Lubitz, S.A.; et al. Accuracy and Usability of a Novel Algorithm for Detection of Irregular Pulse Using a Smartwatch among Older Adults: Observational Study. *JMIR Cardio* **2019**, *3*, e13850. [CrossRef] [PubMed]

25. Dickson, E.L.; Ding, E.Y.; Saczynski, J.S.; Han, D.; Moonis, M.; Fitzgibbons, T.P.; Barton, B.; Chon, K.; McManus, D.D. Smartwatch monitoring for atrial fibrillation after stroke—The Pulsewatch Study: Protocol for a multiphase randomized controlled trial. *Cardiovasc. Digit. Health J.* **2021**, *2*, 231–241. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Bhanvadia, S.B.; Brar, M.S.; Delavar, A.; Tavakoli, K.; Saseendrakumar, B.R.; Weinreb, R.N.; Zangwill, L.M.; Baxter, S.L. Assessing Usability of Smartwatch Digital Health Devices for Home Blood Pressure Monitoring among Glaucoma Patients. *Informatics* **2022**, *9*, 79. [\[CrossRef\]](#)
27. Motti, V.G.; Caine, K. Human Factors Considerations in the Design of Wearable Devices. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2014**, *58*, 1820–1824. [\[CrossRef\]](#)
28. Capeleti, B.S.; Ferreira, J.P.B.; Domingute, G.L.; Pereira, M.R.; Freire, A.P. Heuristic evaluation and user tests of wearable mobile health monitoring applications: What results do different methods yield? In Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems, IHC '20, Diamantina, Brazil, 26–30 October 2020; pp. 1–10. [\[CrossRef\]](#)
29. Quiñones, D.; Rusu, C. How to develop usability heuristics: A systematic literature review. *Comput. Stand. Interfaces* **2017**, *53*, 89–122. [\[CrossRef\]](#)
30. Nielsen, J. “10 Usability Heuristics for User Interface Design”. Available online: <https://www.nngroup.com/articles/ten-usability-heuristics,1994>. (accessed on 22 May 2024).
31. Nielsen, J.; Molich, R. Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'90, Seattle, WA, USA, 1–5 April 1990; pp. 249–256. [\[CrossRef\]](#)
32. Benaida, M. Developing and extending usability heuristics evaluation for user interface design via AHP. *Soft Comput.* **2023**, *27*, 9693–9707.
33. Gemperle, F.; Kasabach, C.; Stivoric, J.; Bauer, M.; Martin, R. Design for wearability. In Proceedings of the Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215), Pittsburgh, PA, USA, 19–20 October 1998; pp. 116–122. [\[CrossRef\]](#)
34. Angelini, L.; Caon, M.; Carrino, S.; Bergeron, L.; Nyffeler, N.; Jean-Mairet, M.; Mugellini, E. Designing a desirable smart bracelet for older adults. In Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, UbiComp 2013 Adjunct, Zurich, Switzerland, 8–12 September 2013; pp. 425–433. [\[CrossRef\]](#)
35. Svanæs, D. Interaction design for and with the lived body. *ACM Trans. Comput. Hum. Interact.* **2013**, *20*, 8. [\[CrossRef\]](#)
36. Jeong, K.S.; Yoo, S.K. Electro-Textile Interfaces: Textile-Based Sensors and Actuators. In *Smart Clothing Technology and Applications*; CRC Press: Boca Raton, FL, USA, 2010; Chapter 4.
37. Tharion, W.J.; Buller, M.J.; Karis, A.J.; Mullen, S.P. Acceptability of a wearable vital sign detection system. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2007**, *51*, 1006–1010. [\[CrossRef\]](#)
38. Siewiorek, D.; Smailagic, A.; Starner, T. *Application Design for Wearable Computing*; Synthesis Lectures on Mobile & Pervasive Computing; Springer: Cham, Switzerland, 2008. [\[CrossRef\]](#)
39. Nazneen Boujarwah, F.A.; Sadler, S.; Mogus, A.; Abowd, G.D.; Arriaga, R.I. Understanding the challenges and opportunities for richer descriptions of stereotypical behaviors of children with asd: A concept exploration and validation. In Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS'10, Orlando, FL, USA, 25–27 October 2010; pp. 67–74. [\[CrossRef\]](#)
40. Cho, H.; Yen, P.Y.; Dowding, D.; Merrill, J.A.; Schnall, R. A multi-level usability evaluation of mobile health applications: A case study. *J. Biomed. Inform.* **2018**, *86*, 79–89. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Lin, R.; Kreifeldt, J.G. Ergonomics in wearable computer design. *Int. J. Ind. Ergon.* **2001**, *27*, 259–269. [\[CrossRef\]](#)
42. Lee, W.; Lim, Y.K. Explorative research on the heat as an expression medium: Focused on interpersonal communication. *Pers. Ubiquitous Comput.* **2012**, *16*, 1039–1049. [\[CrossRef\]](#)
43. Hansson, R.; Ljungstrand, P. The reminder bracelet: Subtle notification cues for mobile devices. In Proceedings of the Conference on Human Factors in Computing Systems, The Hague, The Netherlands, 1–6 April 2000; pp. 323–324. [\[CrossRef\]](#)
44. Lewis, J.R. The System Usability Scale: Past, Present, and Future. *Int. J. Hum.-Comput. Interact.* **2018**, *34*, 577–590. [\[CrossRef\]](#)
45. Brooke, J. SUS: A retrospective. *J. Usability Stud.* **2013**, *8*, 29–40.
46. Karlsson, F. *Assessing Usability Evaluation Methods for Smartwatch Applications*; Degree Project Computer Science and Engineering; KTH Royal Institute of Technology: Stockholm, Sweden, 2016.
47. Usability 101: Introduction to Usability. Available online: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/> (accessed on 2 February 2023).
48. Best Smartwatch 2023: The Best Wearables for iPhone and Android. Expert Reviews. Available online: <https://www.expertreviews.co.uk/smartwatches/1405039/best-uk-smartwatches-wearables> (accessed on 2 February 2023).
49. Samsung Galaxy Watch's Bio Active Sensor Data Can Help Reduce Obesity: AJCN. Deccan Herald. Available online: <https://www.deccanherald.com/business/technology/samsung-galaxy-watches-bio-active-sensor-data-can-help-reduce-obesity-ajcn-1162600.html> (accessed on 13 July 2023).
50. Bennett, J.P.; Liu, Y.E.; Kelly, N.N.; Quon, B.K.; Wong, M.C.; McCarthy, C.; Heymsfield, S.B.; Shepherd, J.A. Next-generation smart watches to estimate whole-body composition using bioimpedance analysis: Accuracy and precision in a diverse, multiethnic sample. *Am. J. Clin. Nutr.* **2022**, *116*, 1418–1429. [\[CrossRef\]](#)
51. Fitbit Versa 2, vs.; Fitbit Versa. Digital Trends. Available online: <https://www.digitaltrends.com/wearables/fitbit-versa-2-vs-fitbit-versa/> (accessed on 13 July 2023).

52. Yoon, Y.H.; Karabiyyik, U. Forensic Analysis of Fitbit Versa 2 Data on Android. *Electronics* **2020**, *9*, 1431. [CrossRef]
53. Fitbit. Google Blog. Available online: <https://blog.google/products/fitbit/> (accessed on 13 July 2023).
54. Fitbit Revenue and Usage Statistics (2023)—Business of Apps. Available online: <https://www.businessofapps.com/data/fitbit-statistics/> (accessed on 18 July 2023).
55. Hayat, S.N.; Ramdani, F. A comparative analysis of usability evaluation methods of academic mobile application: Are four methods better? In Proceedings of the 5th International Conference on Sustainable Information Engineering and Technology, Malang, Indonesia, 16–17 November 2020; pp. 136–141. [CrossRef]
56. Measuring Usability with the System Usability Scale (SUS). MeasuringU. Available online: <https://measuringu.com/sus/> (accessed on 2 February 2023).
57. Balafif, S. Website analysis using heuristic evaluation based on severity ratings and usability scale system. *J. Inform. Teknol. Dan Sains (JINTEKS)* **2022**, *4*, 123–130. [CrossRef]
58. Helander, M.G.; Landauer, T.K.; Prabhu, P.V. Behavioral research methods in human-computer interaction. In *Handbook of Human-Computer Interaction*; Elsevier: Amsterdam, The Netherlands, 1997; pp. 203–227.
59. Azmi, L.F.; Ahmad, N. Exploring the Influence of Human-Centered Design on User Experience in Health Informatics Sector: A Systematic Review. In *Innovative Systems for Intelligent Health Informatics*; Lecture Notes on Data Engineering and Communications Technologies; Springer: Cham, Switzerland, 2021; Volume 72, pp. 242–251. [CrossRef]
60. Irsyad, M.; Khairat, S.B.; Priyadi, Y.; Adrian, M. Usability Measurement in User Interface Design Using Heuristic Evaluation & Severity Rating (Case Study: Mobile TA Application based on MVVM). In Proceedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 26–29 January 2022. [CrossRef]
61. Will, T. Measuring and Interpreting System Usability Scale (SUS). Available online: <https://uiuxtrend.com/measuring-system-usability-scale-sus/> (accessed on 10 May 2024).
62. Exynos W920. Wearable Processor. *Samsung Semiconductor Global*. Available online: <https://semiconductor.samsung.com/processor/wearable-processor/exynos-w920/> (accessed on 18 July 2023).
63. Samsung Galaxy Watch 4 Review: The Return of Wear OS. TechRadar. Available online: <https://www.techradar.com/reviews/samsung-galaxy-watch-4-review> (accessed on 18 July 2023).
64. Samsung Galaxy Watch 4 vs. Galaxy Watch 3: What Are the Differences? Available online: <https://www.androidauthority.com/samsung-galaxy-watch-4-vs-galaxy-watch-3-2731869/> (accessed on 18 July 2023).
65. Fitbit Legal: Privacy Policy. Available online: <https://www.fitbit.com/global/us/legal/privacy-policy> (accessed on 18 July 2023).
66. Homayounfar, M.; Malekijoo, A.; Visuri, A.; Dobbins, C.; Peltonen, E.; Pinsky, E.; Teymourian, K.; Rawassizadeh, R. Understanding Smartwatch Battery Utilization in the Wild. *Sensors* **2020**, *20*, 3784. [CrossRef] [PubMed]
67. Höhn, S.; Bongard-Blanchy, K. Heuristic Evaluation of COVID-19 Chatbots. In *Chatbot Research and Design, Proceedings of the 4th International Workshop, CONVERSATIONS 2020, Virtual Event, 23–24 November 2020*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2021; Volume 12604, pp. 131–144. [CrossRef]
68. Usability and Accessibility Evaluation of Pakistan’s E-Commerce Sites. Available online: https://www.researchgate.net/publication/325086084_Usability_and_Accessibility_Evaluation_of_Pakistan%E2%80%99s_E-commerce_Sites (accessed on 2 February 2023).
69. A’bas, N.N.; Rahim, S.S.; Dolhalit, M.L.; Saifudin, W.S.N.; Abdullasim, N.; Parumo, S.; Omar, R.N.R.; Khair, S.Z.M.; Kalaichelvam, K.; Izhar, S.I.N. Development and Usability Testing of a Consultation System for Diabetic Retinopathy Screening. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 178–188. [CrossRef]
70. Ali, G.; Dida, M.A.; Sam, A.E. Heuristic Evaluation and Usability Testing of G-MoMo Applications. *J. Inf. Syst. Eng. Manag.* **2022**, *7*, 15751. [CrossRef]
71. Zakaria, N.; Wahabi, H.; Qahtani, M.A. Development and usability testing of Riyadh Mother and Baby Multi-center cohort study registry. *J. Infect. Public Health* **2020**, *13*, 1473–1480. [CrossRef] [PubMed]
72. Alturki, R.; AlGhamdi, M.J.; Gay, V.; Awan, N.; ur Rehman, A.; Alshehri, M. Privacy, Security and Usability for IoT-enabled Weight Loss Apps. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 258–263. [CrossRef]
73. Sukmasetya, P.; Setiawan, A.; Arumi, E.R. Usability evaluation of university website: A case study. *J. Phys. Conf. Ser.* **2020**, *1517*, 012071. [CrossRef]
74. View of Improving Online Course Based on the Result of Usability Testing Methods. Available online: <http://psychologyandeducation.net/pae/index.php/pae/article/view/3795/3373> (accessed on 2 February 2023).
75. Santesteban-Echarri, O.; Tang, J.; Fernandes, J.; Addington, J. Development and Usability Testing of SOMO, a Mobile-Based Application to Monitor Social Functioning for Youth at Clinical High-Risk for Psychosis. *Digit. Psychol.* **2020**, *1*, 4–19. [CrossRef]
76. Kumar, B.A.; Goundar, M.S.; Chand, S.S. A framework for heuristic evaluation of mobile learning applications. *Educ. Inf. Technol.* **2020**, *25*, 3189–3204. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.