

Assessment Methods of Usability and Cognitive Workload of Rehabilitative Exoskeletons: A Systematic Review

Laura Maria Alessandra La Bara, Luca Meloni, Davide Giusino *  and Luca Pietrantoni 

Department of Psychology, Alma Mater Studiorum, University of Bologna, 40126 Bologna, Italy; lauramaria.labara@studio.unibo.it (L.M.A.L.B.); luca.meloni@studio.unibo.it (L.M.); luca.pietrantoni@unibo.it (L.P.)

* Correspondence: davide.giusino2@unibo.it

Featured Application: The present work has potential applications in the field of exoskeleton-based rehabilitation, within which it may contribute to the development of guidelines and analytical tools for exoskeletons' usability and cognitive workload.

Abstract: Robotic exoskeleton technologies are applied in the medical field to help patients with impaired mobility to recover their motor functions. Relevant literature shows that usability and cognitive workload may influence the patients' likelihood to benefit from the use of rehabilitative exoskeletons. Following the PRISMA method, the present study aimed to systematically review the assessment methods of usability and cognitive workload in the use of exoskeletal devices for motor rehabilitation. The literature search was conducted in the Scopus and Web of Science bibliographical databases, using 16 keywords that were combined into one search query. A final sample of 23 articles was included in the review, from which 18 distinct assessment methods were identified. Of them, 15 aimed to assess usability, whereas 3 aimed to assess cognitive workload in the use of rehabilitative exoskeletons. Some of the identified methods (e.g., SUS, QUEST, SWAT, and NASA-TLX) showed good psychometric properties and were therefore proven to be appropriate to assess usability and cognitive workload while performing exoskeleton-based rehabilitation. The current study may contribute to the development of guidelines and analytical tools for exoskeletons' usability and exoskeleton-related patients' cognitive workload in the domain of medical rehabilitation.

Keywords: exoskeletons; rehabilitation; usability; cognitive workload; assessment



Citation: La Bara, L.M.A.; Meloni, L.; Giusino, D.; Pietrantoni, L. Assessment Methods of Usability and Cognitive Workload of Rehabilitative Exoskeletons: A Systematic Review. *Appl. Sci.* **2021**, *11*, 7146. <https://doi.org/10.3390/app11157146>

Academic Editor: Oscar Reinoso García

Received: 1 July 2021
Accepted: 30 July 2021
Published: 2 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Robotic exoskeletons are wearable devices intended to augment or enhance the physical capabilities of human subjects—both able-bodied persons and persons with impaired mobility [1]. Recently, medical exoskeletons emerged as a useful option to, on the one hand, facilitate the recovery of a patient's functioning to the level before the injury and, on the other hand, alleviate some of the physical demands associated with traditional motor rehabilitation [2]. From the users' perspective, it is possible for an exoskeleton to measure quantitative data to evaluate the patient's condition and residual mobility [3], which then allows for the design of tailored training programs fostering motor recovery [4]. From the operators' perspective, according to Lo and Xie [3], as exoskeletons can mimic the dynamics of human limbs, they may allow for the treatment of patients without the presence of the therapist, enabling more frequent treatment, providing higher quality care, and potentially reducing costs.

Despite the numerous advantages described above, exoskeleton-based rehabilitation is not successful per se. Several factors concur to the patients' likelihood to benefit from the use of rehabilitative exoskeletons. Among others, the relevant literature suggests that usability and cognitive workload be of primary importance [5,6]. Therefore, to remove potential obstacles and maximize the motor training's success, it is crucial to assess the

usability of the device and the patient's cognitive workload while performing this type of robotic rehabilitation. For this task to be accomplished properly, reliable methods are needed, and knowledge about such methods would be beneficial. On this basis, the present study aimed to focus on usability and cognitive workload in the use of rehabilitative exoskeletons, and particularly on the methods that can be deployed to assess these influential dimensions.

1.1. Usability in the Use of Rehabilitative Exoskeletons

Usability can be defined as the extent to which products can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [7]. A broader definition of usability [8] may include the user's experience of the quality of the product or system. Usability is a critical goal for systems' design, as it pertains to how successfully users can achieve their goals in an acceptable amount of time and if they are satisfied with the experience. Specifically, improving the usability of rehabilitative exoskeletons may result in a better user experience [5], that is, the way a person feels about using a product in a specific context of use, and how these feelings shape the image of oneself when utilizing a device [9]. Rehabilitation technology abandonment by end-users is a well-documented phenomenon [10]; however, enhancing the usability of an exoskeleton may promote the patient's intention to continue motor training.

Usability assessments are performed to identify, and possibly resolve, potential difficulties that are encountered during the use of a given product. According to Dumas and Salzman [11], usability assessment methods can be distinguished into four categories, all displaying pros and cons, namely: (1) usability inspections; (2) usability testing; (3) surveys, interviews, and focus groups; and (4) field methods. First, usability inspections entail the use of guidelines and checklists by expert analysts to assess the usability of a system, as well as the tasks that are performed while utilizing it; although this method benefits from experts' knowledge and experience, and requires relatively little time and resources, and the experts' evaluations may not always reflect the actual end-users' experience [11]. Second, usability testing aims to empirically detect the strengths and weaknesses of a system's usability, as well as to measure and compare the usability of different systems [12]; this method entails the participation of a sample of end-users, who are instructed to perform typical or critical tasks for the correct use of the system; although usability testing implies the direct involvement of end-users, it requires large resources, and its scope could be limited by the number of evaluated tasks. Third, surveys, interviews, and focus groups allow for directly assessing the end-users' experience; specifically, surveys are used to gather information from a large sample of users, whereas interviews and focus groups are directed towards deeper analyses among a smaller number of people; although surveys, interviews, and focus groups allow for informative and often inexpensive assessments of usability, these methods require larger sample sizes to achieve satisfactory levels of reliability [11]. Lastly, field methods constitute a set of techniques that aim to evaluate users' interactions with a system within real-world settings; field studies can be tailored to a wide variety of assessment objectives and can adopt both single-evaluation or longitudinal designs, enabling professionals to assess usability based on the users' actual, rather than reported, behaviors, albeit requiring larger resources than other methods.

Usability assessment methods' adoption cannot be fully comprehended if devoid of context [13]. As there are currently no clear and specific guidelines and analytical tools for assessing the usability of rehabilitative exoskeletons [14], a systematic review of the available literature pointing out the psychometric properties of usability assessment methods in the use context of exoskeleton-based rehabilitation might be needed. This is what the present study aimed to provide; thus, it may be considered as a first step towards filling this gap in both current knowledge and practice.

1.2. Cognitive Workload in the Use of Rehabilitative Exoskeletons

Cognitive workload can be defined as the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and experience [15]. In this definition, attentional resources are thought to have a finite capacity and may be allocated to one or more tasks. Regarding the use of rehabilitative exoskeletons, the patient's cognitive workload should be managed properly for the correct use of the device, which should not be too cognitively taxing. This has implications in terms of both the design and implementation of such medical robotic technologies, whose beneficial effect could be significantly undermined should the task of using them pose too many cognitive demands to the patient.

A wide variety of methods are deployed to assess cognitive workload, which is crucial during the entire design and life cycle of complex systems, such as exoskeletons [16]. The most common assessment methods of cognitive workload could be distinguished into three categories, according to whether they measure primary and secondary task performance, physiological parameters, or subjective reports [17]. Assuming that an individual's performance varies based on the task's cognitive demands, primary task performance measures assess an individual's ability to perform a specific task at an acceptable level, that is, with a reasonably low number of errors [17]; conversely, secondary task performance measures assess an individual's capacity to perform an additional secondary task, assuming that, in any dual-task situation whereby one task is prioritized over the other, the secondary task's performance closely relates to the portion of the individual's mental resources that are not required by the primary task, so that an increase in cognitive workload hinders the performance of the secondary task [17]. Physiological measures of cognitive workload entail the assessment of the physiological variables that may be influenced by the increase or decrease of an individual's cognitive load. For instance, heart rate, blood pressure, facial muscle activation, and brain activity are indicators of the individual's cognitive workload variations [17]. Physiological measures of cognitive workload do not require additional mental demands and are applicable in field settings rather than solely in simulations of task executions [16]. Lastly, subjective reports of cognitive workload consist of either qualitative or quantitative ratings provided by individuals about their perceived cognitive workload while executing a task; this method is appealing because it can be deployed easily, quickly, and at a relatively low cost; nevertheless, subjective reports do not always correlate with objective measures of cognitive workload [18].

The quality of information provided by a method depends on its psychometric properties in a specific context of use [19]. Thus, a systematic literature review might also be needed regarding the assessment methods of cognitive workload in the use of rehabilitative exoskeletons.

1.3. Aim of the Study

The present study aimed to systematically review the assessment methods of usability and cognitive workload in the use of rehabilitative exoskeletons. Specifically, the following two research questions and four sub-questions were inspired by a preliminary, unstructured, and exploratory review of the literature aimed to gather a broad understanding of usability and cognitive workload in the use of rehabilitative exoskeletons.

RQ1: "Which methods are deployed to assess the usability of rehabilitative exoskeletons?"; RQ1a: "How can usability assessment methods be categorized, in terms of the deployed type of measure?"; RQ1b: "What are the psychometric properties (i.e., validity and reliability) of the identified usability assessment methods?";

RQ2: "Which methods are deployed to assess cognitive workload in the use of rehabilitative exoskeletons?"; RQ2a: "How can cognitive workload assessment methods be categorized, in terms of the deployed type of measure?"; RQ2b: "What are the psychometric properties (i.e., validity and reliability) of the identified cognitive workload assessment methods?";

2. Materials and Methods

For the development and reporting of the present study, we followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) protocol [20].

2.1. Keywords and Search Query

Based on the frequency of their occurrence as keywords in the preliminary literature review, we identified five main keywords to operationalize the research questions of the systematic literature review, namely “exoskeleton” (i.e., 100% of the time), “rehabilitation” (i.e., 77% of the time), “usability” (i.e., 36% of the time), “evaluation” (32% of the time), and “cognitive workload” (i.e., 27% of the time). After considering potential synonyms and different spellings, we obtained a total of 16 keywords, such as “exoskelet*”, “rehabilitat*”, “medical”, “assist*”, “usab*”, “cognitiv* *load*”, “mental *load*”, “evaluat*”, “assess*”, “method*”, “tool*”, “instrument*”, “checklist*”, “questionnaire*”, “index*”, and “indices”.

The final keywords were combined with the use of Boolean operators (i.e., AND, OR) to produce one single search query, as follows: (“exoskelet*”) AND (“rehabilitat*” OR “medical” OR “assist*”) AND (“usab*” OR “cognitiv* *load*” OR “mental *load*”) AND (“evaluat*” OR “assess*” OR “method*” OR “tool*” OR “instrument*” OR “checklist*” OR “questionnaire*” OR “index*” OR “indices”).

2.2. Bibliographical Databases

The defined search query was inserted both in Scopus and Web of Science (WoS), which we selected as the bibliographical databases as they have been reported to ensure good multidisciplinary coverage in high-quality peer-reviewed articles [21]. As for the search fields, we selected “TITLE-ABS-KEY” in Scopus and “TOPIC” (i.e., title, abstract, and keywords) in WoS. To maximize the sensitivity of the search process and prevent the exclusion of potentially relevant articles from diverse disciplines (e.g., human factors and ergonomics, medicine, and engineering), the timespan and subject were not restricted. The document type was restricted to articles and reviews in both databases.

2.3. Inclusion and Exclusion Criteria

Inclusion criteria were established to determine the eligibility of the gathered articles against our research questions. Articles had to satisfy the following criteria: (a) study participants are either patients following a robotic exoskeleton-based motor rehabilitation training program or informative healthy subjects such as subject-matter experts (SMEs) in the field of exoskeleton-type systems for rehabilitative purposes; (b) the study entails the measurement of usability and/or cognitive workload in the use of a rehabilitative exoskeleton, although the explicit assessment of them does not need to be the primary study purpose; (c) the study may adopt a meta-analysis, systematic review, experimental, quasi-experimental, correlational, and case study design; and (d) the study is written in English.

Exclusion criteria were established to determine the omission of the gathered articles due to their irrelevance in terms of answering our research questions. We excluded articles showing at least one of the following characteristics: (a) study participants are under the age of 18; (b) the study focuses on exoskeletons that have purposes different from rehabilitation (e.g., industrial or military); (c) the study consists of either a narrative discussion or an unstructured literature review; (d) the full-text article is not retrievable; and (e) the study is not even partially consistent with the objectives of the present study.

2.4. Framework for Critical Appraisal of Retrieved Assessment Methods

According to our research questions, we aimed to review the psychometric properties as well as some other practical aspects of the assessment methods of usability and cognitive workload in the use of rehabilitative exoskeletons. To our knowledge, no specific tool is currently available for the analysis of the psychometric properties of the assessment methods of usability and cognitive workload in the use of rehabilitative exoskeletons.

Therefore, we adapted the checklist originally developed by Francis and colleagues [22] to our purposes. Six general appraisal criteria are included in the checklist, namely: (a) conceptual model, that is, the description of and the rationale for the constructs that an instrument intends to measure, as well as the population to which the instrument is targeted; (b) content validity, that is, the degree to which an instrument's items and subscales are relevant to measure the intended construct; (c) reliability, that is, the degree to which an instrument avoids random measurement errors; (d) construct validity, that is, the degree to which an instrument measures the construct it is intended to; (e) scoring and interpretation, that is, the process of assigning a score to the corresponding item and interpreting its meaning; and (f) respondent burden and presentation, that is, the time, effort, and other demands required from responders for the completion of a test. These criteria were translated into 18 dichotomous items that indicate whether the criteria were or were not met by the instrument, each corresponding to a score of either 0 or 1, whereby the former reflects the absence and the latter the presence of a criterion.

In the adaptation process from Francis and colleagues [22], we removed one item that we deemed not to be appropriate for our purposes. The modified version of the checklist is presented in Table 1.

Table 1. Checklist for the critical appraisal of retrieved assessment methods ¹.

A. Conceptual Model
1. Has the construct of usability/cognitive workload been specifically defined?
2. Has the intended respondent population been described?
3. Does the instrument's conceptual model address whether a single construct/scale or multiple subscales are expected?
B. Content Validity
4. Is there evidence that members of the targeted respondent population were involved in the instrument's development?
5. Is there evidence that content experts were involved in the instrument's development?
6. Is there a description of the methodology by which items/questions were determined (e.g., focus groups and interviews)?
C. Reliability
7. Is there evidence that the instrument's reliability was tested (e.g., test-retest and internal consistency)?
8. Are the reported indices of reliability adequate (e.g., ideal: $r \geq 0.80$; adequate: $r \geq 0.70$) or otherwise justified?
D. Construct Validity
9. Is there reported quantitative justification that single or multiple subscales exist in the instrument (e.g., factor analysis or item response theory)?
10. Is the instrument intended to measure change over time? If YES, is there evidence of both test-retest reliability AND responsiveness to change? Otherwise, award 1 point if there is an explicit statement that the instrument is NOT intended to measure change over time.
11. Are there findings supporting expected associations with existing instruments or with other relevant data?
E. Scoring and Interpretation
12. Is there documentation regarding how to score the instrument (e.g., a scoring method such as summing or an algorithm)?
13. Has a plan for managing and/or interpreting missing responses been described (i.e., how to score incomplete surveys)?
14. Is information provided about how to interpret the instrument's scores (e.g., scaling/anchors and what high and low scores represent) and/or normative data?
F. Respondent Burden and Presentation
15. Is the time to complete reported and reasonable? OR, if it is NOT reported, is the number of questions appropriate for the intended application?
16. Is there a description of the literacy level required by the instrument?
17. Is the entire instrument available for public viewing (e.g., published with the citation or information provided about how to access a copy)?

¹ The checklist is adapted from Francis and colleagues [22] for the purposes of the present study.

2.5. Data Extraction

All articles that were deemed eligible for full-text review against our search criteria underwent careful examination by two reviewers (L.M.A.L.B. and L.M.). The relevant characteristics of the selected studies were extracted and represented on a spreadsheet, such as the definition of the targeted construct (i.e., usability and/or cognitive workload) if present, the deployed assessment method, whether it was a qualitative versus quantitative and objective versus subjective method, the number of subscales in the case of a questionnaire, the number of participants involved in the study, and the measures of validity and reliability if any. Extracted data are shown in Appendix A.

3. Results

Figure 1 shows the process and results of our search strategy. The search was performed on Tuesday, 5 May 2020, and yielded 115 initial articles, of which 63 were from Scopus and 52 from WoS. After removing 40 duplicates, we obtained 75 unique articles that were screened against the inclusion and exclusion criteria. Two articles were excluded as the study participants were under the age of 18; 12 articles were excluded as the study focused on exoskeletons that have purposes different from rehabilitation (e.g., industrial or military); 3 articles were excluded as they did not entail the evaluation of an exoskeleton; 1 article was excluded as the full text was not retrievable; and 34 articles were excluded as the study was not even partially consistent with our objectives. This process left a final sample of 23 articles that were included in the review. Of them, 21 articles reported information about assessment methods of usability in the use of rehabilitative exoskeletons, whereas the remaining two reported information about the assessment methods of cognitive workload within the same context of the use of the same robotic wearable device.

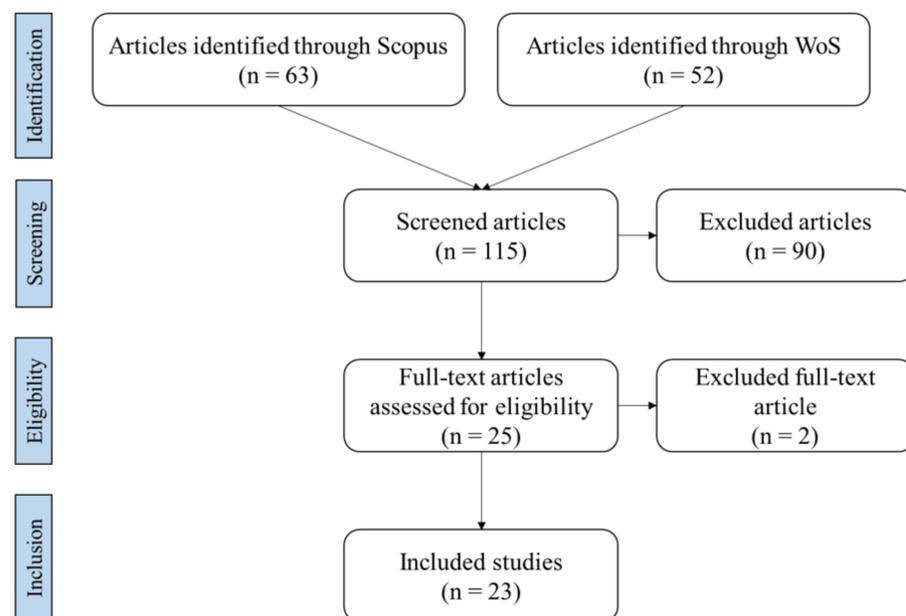


Figure 1. Process and results of the search strategy deployed in the current study.

3.1. Assessment Methods of Usability in the Use of Rehabilitative Exoskeletons

We retrieved 15 assessment methods of usability in the use of rehabilitative exoskeletons. Of these, eight deployed quantitative and subjective measurements, four deployed qualitative and subjective measurements, two deployed quantitative and objective types of measurement, and one deployed a mixed type of measurement. These methods are described in detail within the following subsections.

3.1.1. Quantitative and Subjective Assessment Methods of Usability in the Use of Rehabilitative Exoskeletons

Quantitative and subjective assessment methods are those providing a measurement in the form of an amount or count and relying on individuals' judgments [23].

System Usability Scale (SUS). Reported by seven of the selected studies [24–30], this popular and well-accredited instrument was deployed to conduct quick evaluations of usability with users of a broad range of technological devices and interactive systems. It consists of a 10-item unidimensional psychometric questionnaire that is answered on a 5-point Likert-type scale. The SUS is frequently reported by quantitative evidence (e.g., factor analysis) to show good values of validity and reliability indices, such as construct validity and Cronbach's α . On this basis, this method was assigned a score of 13 out of 17 in our critical appraisal checklist.

Ad hoc questionnaires. Reported by six of the selected studies [30–35], this type of instrument was most often reported without any evaluation of the dimensionality nor validity nor reliability, as well as with any description of scoring procedures. Therefore, this method was assigned a mean score of 5 out of 17 in our critical appraisal checklist.

Quebec User Evaluation of Satisfaction with assistive Technology 2.0 (QUEST 2.0). Reported by five of the selected studies [27,29,36–38], this instrument was deployed to evaluate how users of assistive technologies are satisfied with such devices. It consists of a 12-item bidimensional psychometric questionnaire that is answered on a five-point Likert-type scale. Its dimensions are device (i.e., eight items assessing dimensions/size, weight, adjustments, safety, durability, simplicity of use, comfort, and effectiveness) and service (i.e., four items assessing service delivery, repairs and service of the device, professionalism of the device, and follow-up service), and have been established through statistical techniques such as factor analysis and nomological relatedness. QUEST 2.0 is reported to show good values of validity and reliability indices, such as interclass correlation coefficient (ICC) for the instrument's test-retest and Cronbach's α . On this basis, this method was assigned a score of 11 out of 17 in our critical appraisal checklist.

Visual Analog Scale (VAS). Reported by two of the selected studies [33,35], this instrument was originally developed to evaluate individuals' subjective responses to pain, and can be used to perform usability assessment while using rehabilitative robotic exoskeletons. It consists of one item that is answered on a 10-point Likert-type scale that allows for rating the degree of perceived pain as "no pain" (i.e., 0), "mild" (i.e., from 1 to 3), "moderate" and "severe" (i.e., from 4 to 6), "very severe" (i.e., from 7 to 9), and "worst pain possible" (i.e., 10). The scale is usually graphically complemented by the presence of colored emotion icons, of which the green and smiling ones indicate less pain, whereas the red and suffering ones indicate more pain. Generally, the VAS is reported with no values of psychometric properties. So, this method was assigned a score of 4 out of 17 in our critical appraisal checklist.

AttrakDiff. Reported by one of the selected studies [28], this instrument was used to assess the perceived qualities of a given interactive system. It consists of a 28-item three-dimensional psychometric questionnaire that is answered on a 7-point Likert-type scale. Its dimensions are pragmatic quality, hedonic quality, and attractiveness. Whereas AttrakDiff is reported to show good values of Cronbach's α as an indicator of reliability, it lacks evidence of validity. On this basis, this method was assigned a score of 10 out of 17 in our critical appraisal checklist.

Self-Assessment Manikin (SAM). Reported by one of the selected studies [38], this instrument was deployed to assess the individuals' emotional response to a stimulus, such as using an upper-limb robotic exoskeleton for motor rehabilitation purposes. It consists of three graphical items allowing users to rate pleasure (i.e., from "happy" to "unhappy"), arousal (i.e., from "excited" to "calm"), and dominance (i.e., from "controlled" to "in control") on a nine-point scale. Whereas SAM is reported to show good content validity and acceptable respondent burden, it lacks evidence of reliability, and descriptions of

scoring procedures are not provided. Thus, this method was assigned a score of 11 out of 17 in our critical appraisal checklist.

Heuristic evaluation. Reported by one of the selected studies [31], this instrument was deployed with a hand robotic exoskeleton for rehabilitation. It consists of a set of up to 29 open-ended questions that are answered by SMEs. While providing a clear definition of its target population, the instrument lacks evidence of both validity and reliability. Therefore, this method was assigned a score of 2 out of 17 in our critical appraisal checklist.

Perceived Rate of Exertion (PRE). Reported by one of the selected studies [35], this instrument was deployed to perform assessments of individuals' physical effort, such as in the case of using a wearable robotic exoskeleton for motor rehabilitation of the upper limbs. It consists of one item used to rate the perceived physical effort on a 10-point Likert-type scale, where one corresponds to "very light activity", two and three correspond to "light activity", four to six correspond to "moderate activity", seven and eight correspond to "vigorous activity", nine corresponds to "very hard activity", and ten corresponds to "max effort activity". While PRE is reported to show a very good reliability, it lacks evidence of content and construct validity, as well as a clear definition of its underlying conceptual model and a description of the scoring procedures. On this basis, this method was assigned a score of 7 out of 17 in our critical appraisal checklist.

3.1.2. Qualitative and Subjective Assessment Methods of Usability in the Use of Rehabilitative Exoskeletons

Qualitative and subjective assessment methods are those used to collect data in the form of words, sentences, or descriptions of a phenomenon, and relying on individuals' judgments [23].

Observation. Reported by two of the selected studies [39,40], unstructured observations were conducted to assess the usability of two different rehabilitative exoskeletons. The articles examined included a description of both the target population and the scoring procedure. In addition, SMEs were involved in the design of observations. However, no evidence for validity nor reliability was reported. On this basis, this method was assigned a score of 3 out of 17 in our critical appraisal checklist.

Focus group. Reported by one of the selected studies [41], this method was deployed to detect potential usability issues in a lower-limb rehabilitative exoskeleton. No evidence for construct validity nor reliability was reported. However, a semi-structured facilitator guide was used to standardize the facilitator's conduct across the focus groups. On this basis, this method was assigned a score of 5 out of 17 in our critical appraisal checklist.

Semi-Structured interview. Reported by one of the selected studies [42], this method was deployed to assess the usability of a lower-limb rehabilitative exoskeleton. It consisted of a semi-structured protocol for individual interviews, divided into three themes of investigation, such as capabilities, which can be compensated for or improved with a lower-limb exoskeleton (i.e., behavior capabilities, motor activity capabilities, and protection and resistance capabilities); life habits, which can be improved with a lower-limb exoskeleton (i.e., nutrition, personal care, housing, mobility, responsibility, interpersonal relationships, community life, education, employment, and recreation); and expected technical characteristics of a lower-limb exoskeleton (i.e., appearance, adjustments, comfort, cost, dimensions, durability, effectiveness, installation, weight, repairs and servicing, safety, stigmatization/reaction of others, and usefulness/simplicity of use). The QUEST instrument was used as a reference for the development of the interview protocol. In addition, the transcribed text data were analyzed through thematic analysis with parallel coding to mitigate the subjectivity of interpretation. On this basis, this method was assigned a score of 7 out of 17 in our critical appraisal checklist.

Think-Aloud protocol. Reported by one of the selected studies [25], this method was deployed to assess the usability of a hand/wrist rehabilitative exoskeleton during the prototype development of the system. It consisted of the user thinking aloud while using the robotic device and reporting potentially relevant issues in terms of usability. The think-aloud protocol lacks evidence of both validity and reliability. However, its

findings were positively correlated to other methodologies' results. On this basis, this method was assigned a score of 6 out of 17 in our critical appraisal checklist.

3.1.3. Quantitative and Objective Assessment Methods of Usability in the Use of Rehabilitative Exoskeletons

Quantitative and objective assessment methods are those providing a measurement in the form of an amount or count and intended to measure concrete and observable phenomena or qualities [23].

Toronto Rehabilitation Institute Hand Function Test (TRI-HFT). Reported by one of the selected studies [36], this instrument was originally developed to assess the gross motor function of the hands of patients with motor impairments, and can be deployed to assess the usability of hand robotic exoskeletons for rehabilitation purposes. The test focuses on two main dimensions, namely ability to manipulate and grasp strength. The TRI-HFT is reported to show good values of validity and reliability indices, such as construct validity, content validity, and inter-rater reliability. On this basis, this method was assigned a score of 11 out of 17 in our critical appraisal checklist.

Experimental Characterization. Reported by one of the selected studies [43], this method consists of a procedure that entails the probing and measurement of a system's properties and characteristics, and can be deployed to gauge the usability of a rehabilitative exoskeleton in terms of the torque and speed of the patients' movements. No evidence for validity nor reliability was provided. This method was assigned a score of 6 out of 17 in our critical appraisal checklist.

3.1.4. Mixed Types of Assessment Methods of Usability in the Use of Rehabilitative Exoskeletons

Mixed types of assessment methods are those deploying both quantitative and qualitative, as well as both objective and subjective, measurements.

Framework of Usability for Robotic Exoskeletal Orthoses (FUREO). Reported by one of the selected studies [44], FUREO encompasses six modules (i.e., functional applications, personal factors, device factors, external factors, activities, and health outcomes), each of them corresponding to a set of metrics and measures to assess the usability of current and/or future robotic exoskeletal orthoses for medical motor rehabilitation purposes. Functional applications include community-level mobility, household-level mobility, exercise, and ambulation training. Personal factors are, for instance, fit within device, muscle excitability, trunk stability, neurologic level of injury and severity, and spasticity. Device factors entail, among other elements, noise level, step/stair climbing capability, type of controllers, speed, user interface, battery, size, weight, durability, ease of maintenance, and need for trained assistance. External factors correspond to regulatory approval, cost, availability of device, training, and repair service. Activities include ease and time for donning/doffing, transfers, ascending and descending, reaching, and carrying objects, and running. Finally, health outcomes are bone density, cardiovascular fitness, cholesterol, body composition, glucose intolerance, pressure ulcers, bowel function, depression, mood, pain, fatigue, and sleep. Despite its complexity and comprehensiveness, despite a clear definition of both the construct of usability and its target population (i.e., persons with neurologic conditions such as spinal cord injury, stroke, and multiple sclerosis), and despite a thorough description of the expected subscales of the instrument, to our knowledge, FUREO has never been applied nor deployed. As such, it remains a purely theoretical framework that lacks, at present, any evidence for validity and reliability. On this basis, this method was assigned a score of 3 out of 17 in our critical appraisal checklist.

3.2. Assessment Methods of Cognitive Workload in the Use of Rehabilitative Exoskeletons

We retrieved three assessment methods of cognitive workload in the use of rehabilitative exoskeletons. Of these, two deploy quantitative and subjective measurements, and one deploys quantitative and objective types of measurement. These methods are described in detail within the following subsections.

3.2.1. Quantitative and Subjective Assessment Methods of Cognitive Workload in the Use of Rehabilitative Exoskeletons

Subjective Workload Assessment Technique (SWAT). Reported by one of the selected studies [45], this instrument enables assessing cognitive workload during the use of a rehabilitative exoskeleton. It consists of three multiple-choice items for rating time load, mental effort load, and psychological stress load. SWAT has been reported with a clear definition of its underlying conceptual model, as well as with good values of validity and reliability indices, such as content validity, construct validity, and Cronbach's α . On this basis, this method was assigned a score of 14 out of 17 in our critical appraisal checklist.

NASA-Task Load Index (NASA-TLX). Reported by one of the selected studies [45], this popular and well-accredited instrument was deployed to evaluate individuals' perceptions of task-related workload and can be used to gauge the cognitive workload experienced by users of rehabilitative exoskeletons while performing the task of using such robotic devices. It consists of six items for rating mental demand, physical demand, temporal demand, performance, effort, and frustration on a 10-point Likert-type scale. NASA-TLX has been reported with a clear definition of its underlying conceptual model, as well as with good values of validity and reliability indices, such as content validity, construct validity, and Cronbach's α . On this basis, this method was assigned a score of 14 out of 17 in our critical appraisal checklist.

3.2.2. Quantitative and Objective Assessment Methods of Cognitive Workload in the Use of Rehabilitative Exoskeletons

Event-Related Potential (ERPs). Reported by one of the selected studies [46], amplitude analysis of ERPs corresponds to an electroencephalography-based measurement of peaks in the brain activity and, as such, can be deployed to assess the cognitive workload of exoskeleton users while performing rehabilitative exercises. This technique has been reported with good construct validity and content validity, but no evidence for reliability was provided. On this basis, this method was assigned a score of 9 out of 17 in our critical appraisal checklist.

3.3. Critical Appraisal of Retrieved Assessment Methods

Figure 2 shows the results of applying our framework for the critical appraisal of the retrieved assessment methods of usability and cognitive workload in the use of rehabilitative exoskeletons. Among the methods aiming to assess usability, SUS resulted in the one with the highest score (i.e., 13), whereas heuristic evaluation resulted in the one with the lowest score assigned (i.e., 2). Among the methods aiming to assess cognitive workload, SWAT and NASA-TLX equally resulted as the ones with the highest score (i.e., 14), whereas ERP amplitude analysis resulted in the one with the lowest score assigned (i.e., 9). Nevertheless, a couple of other noteworthy methods obtained relatively high scores, such as QUEST and AttrakDiff, scoring 11 and 10, respectively. None of the retrieved methods were assigned the maximum total score of 17 possible in the adopted checklist, while only SUS and the TRI-HFT received at least 1 point for each evaluative dimension.

Investigated variable	Methodology	Conceptual Model			Content Validity			Reliability		Construct Validity			Scoring and Interpretation			Respondent Burden and Presentation		
		Construct defined	Target population defined	Expected sub-scales described	Participant involvement	Content experts involvement	Description of item development	Reliability tested	Coefficient adequacy	Justification of sub-scales	Longitudinal validity	Convergent validity	Scoring procedure is described	Plan for missing responses	Interpretation of data is described	Completion time is reasonable	Literacy level is described	Publicly available
Usability	TRI-HTF																	
	Experimental characterization																	
	SUS																	
	QUEST 2.0																	
	Ad hoc questionnaires																	
	AttrakDiff																	
	SAM																	
	Heuristic evaluation																	
	PRE																	
	VAS																	
	Focus group																	
	Observation																	
	Semi-structured interview																	
	Think-aloud protocol																	
	FUREO																	
Cognitive workload	ERPs analysis																	
	SWAT																	
	NASA-TLX																	

Figure 2. Critical appraisal of the retrieved assessment methods of usability and cognitive workload in the use of rehabilitative exoskeletons based on the adaptation of the checklist from Francis and colleagues [22].

4. Discussion

In the present study, we systematically reviewed the assessment methods of usability and cognitive workload in the use of rehabilitative exoskeletons.

To answer RQ1, we retrieved the methods that were deployed to assess the usability of rehabilitative exoskeletons. Specifically, we categorized usability assessment methods in terms of the deployed type of measure (RQ1a) and investigated their psychometric properties (RQ1b). We were able to retrieve mostly quantitative and subjective assessment methods, as well as qualitative and subjective ones. These methods most often consist of questionnaires and interviews, for both individuals and groups. This finding may be thought of as reflecting the need for researchers to take advantage of usually inexpensive

methodologies in order to conduct usability assessments of rehabilitative exoskeletons whereby end-users are directly involved in providing informed feedback [11]. On the one hand, quantitative methods allow for standardized evaluations of a rehabilitative exoskeleton's overall usability, thus being more straightforward and less ambiguous when it comes to the interpretation, comparability, and generalizability of results across different experiments. On the other hand, qualitative methods may offer deeper insights into potential issues that might foster or hinder the usability of a rehabilitative exoskeleton, thus providing researchers with valuable information regarding specific usability issues encountered by participants while using the wearable robotic device during some rehabilitation exercise; in turn, this may enable an easier resolution of problems and may reveal especially important information during the early phases of a rehabilitative exoskeleton's development. It is reasonable to expect the validity and reliability of qualitative methods (i.e., interviews and focus groups) to be affected by the facilitators' ability to conduct the discussion with the interviewees or to manage a smooth interaction between participants. In addition, the lack of a clearly defined conceptual model and an obscure description of the methodological procedure may render qualitative methods less effective. Nevertheless, these issues can apply to quantitative methods too. Considering all of the above, it can be concluded that quantitative and qualitative methods should not be considered as competing nor as being mutually exclusive. Rather, these two types of measurements can be integrated and can complement each other to compensate for each other's limitations and maximize each other's pros.

Among the retrieved methods that were deployed to assess the usability of rehabilitative exoskeletons, SUS, QUEST, and AttrakDiff showed the best psychometric properties and therefore proved appropriate to assess usability while performing robotic exoskeleton-based motor rehabilitation. The only notable caveat about SUS concerns the fourth item of the questionnaire, namely "I think that I would need the support of a technical person to be able to use this system". As exoskeleton-based rehabilitation requires a therapist's assistance throughout the exercises' performance, the final score for this item may not necessarily reflect the actual usability of the exoskeletal device. This conclusion is supported by findings by Tsai and colleagues [30], who reported high internal consistency values as an indicator of the reliability of SUS, except for item 4, thus suggesting the presence of a low correlation between this item's score and the total score of SUS. Regarding QUEST, it aims to also evaluate other dimensions of satisfaction with the use of a given rehabilitative exoskeleton that are not directly related to usability, such as safety, size, comfort, weight, and durability. Therefore, the adoption of this method may lead to a criterion contamination problem, whereby the actual criterion (i.e., the parameter that is used to measure a construct) includes variables it should not, ultimately leading to measurement errors. Moreover, QUEST overlooks some sub-dimensions of usability (e.g., learnability and efficiency) that may be informative for both research and practice purposes. Therefore, it can be concluded that QUEST may provide researchers with valid and reliable assessments of rehabilitative exoskeletons' usability, although they ought to be aware of its limitations and cautious against potential measurement errors. In this regard, the deployment of only its "device" subscale may be advisable, because of its resemblance with the construct of usability [27,36–38]. Similarly, the conceptual model underlying AttrakDiff partially encompasses the construct of usability within its "pragmatic quality" subscale, and the overall method aims to assess the interactive qualities of a rehabilitative exoskeleton beyond its mere usability, such as an individual's emotional response to the use of the wearable robotic medical device. Again, this may be expected to lead to a criterion contamination problem, as the instrument includes two other sub-dimensions (i.e., "hedonic quality" and "attractiveness") that are not strictly representative of the construct of usability. So, deploying only the "pragmatic quality" subscale may be a potential solution to this issue, if support for its validity and reliability is provided. Furthermore, an additional notation is worthwhile regarding two other quantitative and subjective assessment methods of the usability of rehabilitative exoskeletons, such as VAS and PRE. These instruments aim to

measure the subjective response to pain and the perceived physical effort in users during robotic exoskeleton-based rehabilitation exercises, respectively. However, as rehabilitation itself tends to imply difficult and painful exercises according to the severity of the condition suffered by patients, the assessments performed via VAS and PRE may not necessarily reflect the usability of the wearable medical device.

As for objective methods, they constitute a minority among the retrieved assessment methods of usability in the use of rehabilitative exoskeletons. One major issue with this type of method is that it does not provide a clear indication of the validity for the usability construct. Rather, usability was conceptualized as enjoyment, difficulty, and comfort [33]; ease of use [39]; or torque and speed [43]. As the term “usability” is often intended as a synonym of “feasibility” within the engineering field, the authors of these studies might have deployed instruments that were consistent with this mental model.

To answer RQ2, we retrieved the methods that were deployed to assess the cognitive workload associated with the use of rehabilitative exoskeletons. Specifically, we categorized cognitive workload assessment methods in terms of the deployed type of measure (RQ2a) and investigated their psychometric properties (RQ2b). Compared with the usability assessment methods, we were able to retrieve a smaller number of methods. Of them, we retrieved mostly quantitative and subjective assessment methods. None of the retrieved assessment methods deployed qualitative types of measurements. Although the limited number of studies included prevented us from drawing definite conclusions, this result may be thought to reflect a lack of interest in these methods’ adoption on the part of researchers. Among the retrieved methods that were deployed to assess the cognitive workload associated with the use of rehabilitative exoskeletons, SWAT and NASA-TLX showed the best psychometric properties and were therefore proven to be appropriate to assess cognitive workload while performing robotic exoskeleton-based motor rehabilitation. As objective and subjective measures of cognitive workload in the use of rehabilitative exoskeletons do not strongly correlate with each other [18,46], it is advisable to adopt both types of methodologies in order to achieve a comprehensive evaluation of the participants’ mental effort.

The results and conclusions of the present study should be considered in light of some limitations. First, our search strategy might have produced different outcomes if using additional bibliographical databases, which in turn may have yielded alternative knowledge and, therefore, led to dissimilar results and conclusions compared with those hereby achieved. Nevertheless, as the nature of the targeted study topic touches on a broad range of diverse applied disciplines (e.g., human factors and ergonomics, medicine, and engineering), we purposely selected Scopus and WoS because of their multidisciplinary coverage of high-quality peer-reviewed articles [21], while using more specialized sources of information (e.g., PsychInfo and PubMed) might have yielded too narrowed results. So, even if this review cannot address a big amount of data, it could be of interest to various dedicated readers. Second, the results from our critical appraisal of the retrieved assessment methods of usability and cognitive workload in the use of rehabilitative exoskeletons should be cautiously considered in light of a potential bias, whereby the components and characteristics of the adopted evaluative framework might appear to be more favorable towards quantitative methods compared with the qualitative ones. Nonetheless, the checklist was revealed to be useful for appraising all of the retrieved assessment methods. Finally, rehabilitative exoskeletons vary widely in their properties and modes of operation between upper limbs and lower limbs, as well as those operated completely autonomously compared with those operated based on the users’ intention. This variety might influence the selection of usability or cognitive workload assessment methods, as well as the assessment itself, as certain methods may be more suitable to certain properties and modes of operation, and less suitable to others. Although relevant, this issue was not addressed in our study. However, it was beyond the scope of the present review, and it may constitute an interesting direction for future research.

Despite the above limitations, the present study holds relevant practical implications about contributing to the development of the currently lacking [14] guidelines and analytical tools for exoskeletons' usability and exoskeleton-related patients' cognitive workload in the domain of medical rehabilitation. The results may inform the design of benchmarking protocols [47] for rehabilitative exoskeletons, thus enabling the assessment of exoskeletons' performance and characteristics at different developmental stages, by adopting valid and reliable assessment methods. In addition, the assessment methods that have been reported to show good psychometric properties may be regarded as valuable resources to guide both the development and implementation of rehabilitative exoskeletons in medical settings. Particularly, these methods may prove useful in the context of user-centered design processes, whereby measures of usability and cognitive workload might be used to identify users' requirements and to implement appropriate solutions within the system's design [10].

5. Conclusions

Robotic exoskeletons are deployed in the medical field to help patients with impaired mobility to recover their motor functions. Usability and cognitive workload may influence the patients' likelihood to benefit from the use of rehabilitative exoskeletons [5,6]. To remove potential obstacles and maximize the motor training's success, both researchers and practitioners would benefit from the use of valid and reliable methods to assess usability and cognitive workload associated with performing rehabilitation exercises while wearing an exoskeletal device. For this assessment task to be accomplished properly, SUS, QUEST, and AttrakDiff (i.e., usability), as well as SWAT and NASA-TLX (i.e., cognitive workload), may prove suitable.

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

Funding: This research received funding from the European Union's Horizon 2020 research and innovation program, via an open call issued and executed under project EUROBENCH (grant agreement no. 779963). <http://eurobench2020.eu> (accessed on 1 July 2021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available within the article.

Acknowledgments: The authors would like to thank Guido Tolomelli from the Department of Psychology of the Alma Mater Studiorum—University of Bologna, for his helpful comments to the first draft of the manuscript.

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

Appendix A

Table A1. Data extracted from the literature search.

Reference	Construct	Methods	Measure Type	Dimensions	Items	Participants	Psychometrics	Comments
Almenara et al. (2017) [31]	Usability	A. Heuristic evaluation and questionnaire for SMEs; B. Questionnaire for end-users. Both administered at T0 and one-year follow-up	Qualitative, subjective for both	1 for both	A. 21 (open-ended) at T0, 29 at follow-up; B. 13 (6-point Likert) at T0, 14 at follow-up	37 (28 SMEs and 9 end-users)	N/A	Usability testing for developmental purposes
Ambrosini et al. (2019) [24]	Usability	System Usability Scale (SUS)	Quantitative, subjective	1	10 (5-point Likert)	7	N/A	N/A
Amirabdollahian et al. (2014) [25]	Usability	A. Think aloud protocol (development); B. SUS (testing)	A. Qualitative, subjective; B. Quantitative, subjective	B. 1	B. 10 (5-point Likert)	23 (12 completed)	N/A	Methods may differ based on the phase of exoskeleton development
Bortole et al. (2015) [39]	Usability	A. Unstructured observation; B. Ratings of ease of use	A. Qualitative, objective; B. Quantitative, subjective	1 for both	B. 1 (10-point Likert)	3	N/A	Usability was conceptualized as ease of use
Bryce et al. (2015) [44]	Usability	Framework of Usability for Robotic Exoskeletal Orthoses (FUREO)	Quantitative and qualitative, subjective and objective	6 (applications, personal, device, external factors, activities, and health outcomes)	N/A	N/A	N/A	N/A
Casey et al. (2019) [26]	Usability	SUS	Quantitative, subjective	1	10 (5-point Likert)	4	Reliability. High internal consistency; Validity. Reference to validation study	Learnability was also considered as a dimension of usability
Deeny et al. (2014) [46]	Cognitive workload	Event-related potential (ERP) amplitude analysis	Quantitative, objective	N/A	N/A	20	Validity. The study is a validation study in itself	N/A

Table A1. Cont.

Reference	Construct	Methods	Measure Type	Dimensions	Items	Participants	Psychometrics	Comments
Dudley et al. (2019) [27]	Usability	A. SUS; B. Quebec User Evaluation of Satisfaction with assistive Technology (QUEST)	Quantitative, subjective for both	A. 1; B. 2 (device and services)	A. 10 (5-point Likert); B. 12 (5-point Likert)	1	N/A	Only 8 items of the device sub-scale of QUEST were used
Eicher et al. (2019) [28]	Usability	A. SUS; B. AttrakDiff	Quantitative, subjective for both	A. 1; B. 3 (pragmatic quality, hedonic quality, and attractiveness)	A. 10 (5-point Likert); B. 28 (7-point Likert)	13	N/A	N/A
Heinemann et al. (2020) [41]	Usability	Focus group	Qualitative, subjective	N/A	N/A	35	Reliability. A semi-structured facilitator guide was used to standardize facilitator's conduct across focus groups	Focus group data were analyzed through thematic analysis
Koumpuros (2016) [29]	Usability	A. SUS; B. QUEST	Quantitative, subjective	N/A	N/A	31 articles	Validity. Reference to validation studies	N/A
Lajeunesse et al. (2018) [42]	Usability	Semi-structured individual interview	Qualitative, subjective	3 (capabilities, life habits, and expected technical characteristics)	N/A	13	Reliability. Semi-structured protocol and parallel coding of text data	QUEST was used as a reference
Liu et al. (2017) [45]	Cognitive workload	A. Subjective Workload Assessment Technique (SWAT); B. NASA Task Load Index (NASA-TLX)	Quantitative, subjective for both	A. 3 (time load, mental effort load, and psychological stress load); B. 6 (mental demand, physical demand, temporal demand, performance, effort, and frustration)	A.3; B. 6 (10-point Likert)	6	Validity. Good content and construct validity. Reliability. Cronbach's $\alpha > 0.8$ for both	N/A

Table A1. Cont.

Reference	Construct	Methods	Measure Type	Dimensions	Items	Participants	Psychometrics	Comments
López-Larraz et al. (2016) [37]	Usability	QUEST (modified)	Quantitative, subjective	2 (device and services)	9 (5-point Likert)	4	N/A	Only the device sub-scale of QUEST was used
Nam et al. (2019) [32]	Usability	Ad-hoc questionnaire	Quantitative, subjective	1	10 (7-point Likert)	20 (only 2 patients)	N/A	Usability assessment for developmental purposes
Ozkul and Barkana (2013) [35]	Usability	A. Perceived Rate of Exertion Scale (PRES); B. Visual Analog Scale (VAS); C. Ad-hoc questionnaire	Quantitative, subjective for all	1 each	A. 1 (10-point Likert); B. 1 (10-point Likert); C. 10 (5-point Likert)	9	N/A	Comfort, performance, and safety were also part of the assessment
Resquín et al. (2017) [38]	Usability	A. QUEST; B. Self-Assessment Manikin (SAM)	Quantitative, subjective for both	A. 2 (device and services); B. 3 (pleasure, arousal, and dominance)	A. 7 (5-point Likert); B. 3 (9-point Likert)	3	N/A	Only the device sub-scale of QUEST was used
Schneider et al. (2016) [33]	Usability	A. Ad-hoc questionnaire; B. Visual analysis of facial expressions	A. Quantitative, subjective; B. Quantitative, objective	1 each	A. 3 (5-point Likert)	8	N/A	Usability was conceptualized as enjoyment, difficulty, and comfort
Schrade et al. (2018) [40]	Usability	Field behavioral observation	Qualitative, subjective	N/A	N/A	2	N/A	Usability was conceptualized as rehabilitative progress and effectiveness
Stellin et al. (2018) [34]	Usability	Ad-hoc questionnaire	Quantitative, subjective	1	10 (5-point Likert)	16 (of which 6 patients)	N/A	Usability was conceptualized as satisfaction
Tsai et al. (2019) [30]	Usability	A. SUS; B. Ad-hoc questionnaire	Quantitative, subjective for both	1 each	A. 10 (5-point Likert); B. 5 (5-point Likert)	22 (of which 18 patients)	Reliability. A. Cronbach's $\alpha = 0.64$; B. $\alpha = 0.85$	N/A

Table A1. Cont.

Reference	Construct	Methods	Measure Type	Dimensions	Items	Participants	Psychometrics	Comments
Vitiello et al. (2016) [43]	Usability	Experimental characterization A. Toronto Rehabilitation Institute Hand Function Test (TRI-HFT); B. Korean-Quebec User Evaluation of Satisfaction with assistive Technology 2.0 (K-QUEST 2.0)	Quantitative, objective	N/A	N/A	3	N/A	Torque and speed were used
Yoo et al. (2019) [36]	Usability	Korean-Quebec User Evaluation of Satisfaction with assistive Technology 2.0 (K-QUEST 2.0)	A. Quantitative, objective; B. Quantitative, subjective	A. 2 (ability to manipulate and grasp strength); B. 2 (device and services)	A. 19; B. 12 (5-point Likert)	10	A. Inter-rater reliability = 1.0 ($p < 0.01$)	Only the device sub-scale of QUEST was used

N/A—Not applicable.

References

1. Gams, A.; Petrič, T.; Debevec, T.; Babič, J. Effects of robotic knee exoskeleton on human energy expenditure. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 1636–1644. [CrossRef] [PubMed]
2. Wall, A.; Borg, J.; Palmcrantz, S. Clinical application of the Hybrid Assistive Limb (HAL) for gait training—A systematic review. *Front. Syst. Neurosci.* **2015**, *9*, 1–10. [CrossRef]
3. Lo, H.S.; Xie, S.Q. Exoskeleton robots for upper-limb rehabilitation: State of the art and future prospects. *Med. Eng. Phys.* **2012**, *34*, 261–268. [CrossRef]
4. Chen, G.; Chan, C.K.; Guo, Z.; Yu, H. A review of lower extremity assistive robotic exoskeletons in rehabilitation therapy. *Crit. Rev. Biomed. Eng.* **2013**, *41*, 343–363. [CrossRef] [PubMed]
5. Kim, T. Factors influencing usability of rehabilitation robotic devices for lower limbs. *Sustainability* **2020**, *12*, 598. [CrossRef]
6. Stirling, L.; Siu, H.C.; Jones, E.; Duda, K. Human Factors considerations for enabling functional use of exosystems in operational environments. *IEEE Syst. J.* **2018**, *13*, 1072–1083. [CrossRef]
7. International Organization for Standardization. *Ergonomic Requirements for Office Work with Visual Display Terminals. Part 11: Guidance on Usability*; ISO Standard No. 9241-11:1998; International Organization for Standardization: Geneva, Switzerland, 1998. Available online: <https://www.iso.org/standard/16883.html> (accessed on 1 August 2021).
8. Earthy, J.; Jones, B.; Bevan, N. ISO standards for user-centered design and the specification of usability. In *Usability in Government Systems*; Buie, E., Murray, D., Eds.; Morgan Kaufman: Burlington, MA, USA, 2012; Volume 1, pp. 267–283.
9. Kaasinen, E.; Roto, V.; Hakulinen, J.; Heimonen, T.; Jokinen, J.P.; Karvonen, H.; Tokkonen, H. Defining user experience goals to guide the design of industrial systems. *Behav. Inf. Technol.* **2015**, *34*, 976–991. [CrossRef]
10. Hill, D.; Holloway, C.S.; Ramirez, D.Z.M.; Smitham, P.; Pappas, Y. What are user perspectives of exoskeleton technology? A literature review. *Int. J. Technol. Assess. Health Care* **2017**, *33*, 160–167. [CrossRef]
11. Dumas, J.S.; Salzman, M.C. Usability assessment methods. *Rev. Hum. Factors Ergon.* **2006**, *2*, 109–140. [CrossRef]
12. Helander, M. Design of human-computer interaction. In *A Guide to Human Factors and Ergonomics*, 2nd ed.; Helander, M., Ed.; Taylor & Francis: Boca Raton, FA, USA, 2006; Volume 1, pp. 119–143.
13. Furniss, D.; Blandford, A.; Curzon, P. Usability Evaluation Methods in Practice: Understanding the Context in Which They Are Embedded. In Proceedings of the 14th European Conference on Cognitive Ergonomics: Invent! Explore! New York, NY, USA, 28–31 August 2007; Brinkman, W.P., Ham, D.H., Wong, B.L.W., Eds.; Association for Computing Machinery: New York, NY, USA, 2007; pp. 253–256. [CrossRef]
14. Pietrantonio, L.; Rainieri, G.; Fraboni, F.; Tušl, M.; Giusino, D.; De Angelis, M.; Tria, A. Usability of Exosystems: A Review. In *CHIItaly19: Adjunct Proceedings of the 13th Biannual Conference of the Italian SIGCHI Chapter: Designing the Next Interaction, Padua, Italy, 23–25 September 2019*; Gamberini, L., Pittarello, F., Spagnolli, A., Eds.; Association for Computing Machinery: New York, NY, USA, 2019; Available online: https://www.researchgate.net/publication/336044737_Usability_of_exosystems_A_review (accessed on 23 June 2021).
15. Young, N.; Stanton, N. Mental workload: Theory, measurement, and application. In *International Encyclopedia of Ergonomics and Human Factors*, 2nd ed.; Karwowski, W., Ed.; Taylor & Francis: Boca Raton, FA, USA, 2006; Volume 3, pp. 866–869.
16. Stanton, N.A.; Salmon, P.M.; Rafferty, L.A.; Walker, G.H.; Baber, C.; Jenkins, D.P. Mental workload assessment methods. In *Human Factors Methods: A Practical Guide for Engineering and Design*, 2nd ed.; Stanton, N.A., Salmon, P.M., Rafferty, L.A., Walker, G.H., Baber, C., Jenkins, D.P., Eds.; Taylor & Francis: Boca Raton, FA, USA, 2017; pp. 301–364.
17. Young, M.S.; Brookhuis, K.A.; Wickens, C.D.; Hancock, P.A. State of science: Mental workload in ergonomics. *Ergonomics* **2015**, *58*, 1–17. [CrossRef]
18. Yakobi, O. Determinants of association and dissociation between subjective and objective measures of workload. *Proc. Hum. Factors Ergon.* **2018**, *62*, 222–226. [CrossRef]
19. Souza, A.C.D.; Alexandre, N.M.C.; Guirardello, E.D.B. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol. Serv. Saúde* **2017**, *26*, 649–659. [CrossRef] [PubMed]
20. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann. Intern. Med.* **2009**, *151*, 264–269. [CrossRef] [PubMed]
21. Mingers, J.; Leydesdorff, L. A review of theory and practice in scientometrics. *Eur. J. Oper. Res.* **2015**, *246*, 1–19. [CrossRef]
22. Francis, D.O.; McPheeters, M.L.; Noud, M.; Penson, D.F.; Feurer, I.D. Checklist to operationalize measurement characteristics of patient-reported outcome measures. *Syst. Rev.* **2016**, *5*, 1–11. [CrossRef] [PubMed]
23. UserFocus. Usability test data. Available online: <https://www.userfocus.co.uk/articles/datathink.html> (accessed on 25 June 2021).
24. Ambrosini, E.; Zajc, J.; Ferrante, S.; Ferrigno, G.; Gasperina, S.D.; Bulgheroni, M.; Baccinelli, W.; Schauer, T.; Wiesener, C.; Russold, M.; et al. A hybrid robotic system for arm training of stroke survivors: Concept and first evaluation. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 3290–3300. [CrossRef]
25. Amirabdollahian, F.; Ates, S.; Basteris, A.; Cesario, A.; Buurke, J.; Hermens, H.; Hofs, D.; Johansson, E.; Mountain, G.; Nasr, N.; et al. Design, development and deployment of a hand/wrist exoskeleton for home-based rehabilitation after stroke—SCRIPT project. *Robotica* **2014**, *32*, 1331–1346. [CrossRef]
26. Casey, A.; Azhar, H.; Grzes, M.; Sakel, M. BCI controlled robotic arm as assistance to the rehabilitation of neurologically disabled patients. *Disabil. Rehabil. Assist. Technol.* **2021**, *16*, 525–537. [CrossRef]

27. Dudley, D.R.; Knarr, B.A.; Siu, K.C.; Peck, J.; Ricks, B.; Zuniga, J.M. Testing of a 3D printed hand exoskeleton for an individual with stroke: A case study. *Disabil. Rehabil. Assist. Technol.* **2021**, *16*, 209–213. [[CrossRef](#)] [[PubMed](#)]
28. Eicher, C.; Haesner, M.; Spranger, M.; Kuzmicheva, O.; Gräser, A.; Steinhagen-Thiessen, E. Usability and acceptability by a younger and older user group regarding a mobile robot-supported gait rehabilitation system. *Assist. Technol.* **2019**, *31*, 25–33. [[CrossRef](#)] [[PubMed](#)]
29. Koumpouros, Y. A systematic review on existing measures for the subjective assessment of rehabilitation and assistive robot devices. *J. Healthc. Eng.* **2016**, *2016*, 1–10. [[CrossRef](#)]
30. Tsai, Y.; Huang, J.; Pu, S.; Chen, H.; Hsu, S.; Chang, J.; Pei, Y. Usability assessment of a cable-driven exoskeletal robot for hand rehabilitation. *Front. Neurobot.* **2019**, *13*, 1–5. [[CrossRef](#)] [[PubMed](#)]
31. Almenara, M.; Cempini, M.; Gómez, C.; Cortese, M.; Martín, C.; Medina, J.; Vitiello, N.; Opisso, E. Usability test of a hand exoskeleton for activities of daily living: An example of user-centered design. *Disabil. Rehabil. Assist. Technol.* **2017**, *12*, 84–96. [[CrossRef](#)]
32. Nam, H.S.; Hong, N.; Cho, M.; Lee, C.; Seo, H.G.; Kim, S. Vision-assisted interactive human-in-the-loop distal upper limb rehabilitation robot and its clinical usability test. *Appl. Sci.* **2019**, *9*, 3106. [[CrossRef](#)]
33. Schneider, J.C.; Ozsecen, M.Y.; Muraoka, N.K.; Mancinelli, C.; Della Croce, U.; Ryan, C.M.; Bonato, P. Feasibility of an exoskeleton-based interactive video game system for upper extremity burn contractures. *PM&R* **2016**, *8*, 445–452. [[CrossRef](#)]
34. Stellin, G.; Sale, P.; Masiero, S.; Becchi, F.; Sieklicki, W. Development and test of fex, a fingers extending exoskeleton for rehabilitation and regaining mobility. *Int. J. Mech. Control* **2018**, *19*, 3–15.
35. Ozkul, F.; Barkana, D.E. Upper-extremity rehabilitation robot RehabRoby: Methodology, design, usability and validation. *Int. J. Adv. Robot. Syst.* **2013**, *10*, 1–13. [[CrossRef](#)]
36. Yoo, H.J.; Lee, S.; Kim, J.; Park, C.; Lee, B. Development of 3D-printed myoelectric hand orthosis for patients with spinal cord injury. *J. Neuroeng. Rehabil.* **2019**, *16*, 1–14. [[CrossRef](#)]
37. López-Larraz, E.; Trincado-Alonso, F.; Rajasekaran, V.; Pérez-Nombela, S.; Del-Ama, A.; Aranda, J.; Minguez, J.; Gil-Agudo, A.; Montesano, L. Control of an ambulatory exoskeleton with a brain-machine interface for spinal cord injury gait rehabilitation. *Front. Neurosci.* **2016**, *10*, 1–15. [[CrossRef](#)]
38. Resquín, F.; Gonzalez-Vargas, J.; Ibáñez, J.; Brunetti, F.; Dimbwadyo, I.; Carrasco, L.; Alves, S.; Gonzalez-Altad, C.; Gomez-Blanco, A.; Pons, J.L. Adaptive hybrid robotic system for rehabilitation of reaching movement after a brain injury: A usability study. *J. Neuroeng. Rehabil.* **2017**, *14*, 1–15. [[CrossRef](#)]
39. Bortole, M.; Venkatakrisnan, A.; Zhu, F.; Moreno, J.C.; Francisco, G.E.; Pons, J.L.; Contreras-Vidal, J. The H2 robotic exoskeleton for gait rehabilitation after stroke: Early findings from a clinical study wearable robotics in clinical testing. *J. Neuroeng. Rehabil.* **2015**, *12*, 1–14. [[CrossRef](#)]
40. Schrade, S.O.; Dätwyler, K.; Stücheli, M.; Studer, K.; Türk, D.; Meboldt, M.; Gassert, R.; Lambercy, O. Development of VariLeg, an exoskeleton with variable stiffness actuation: First results and user evaluation from the CYBATHLON 2016. *J. Neuroeng. Rehabil.* **2018**, *15*, 1–18. [[CrossRef](#)] [[PubMed](#)]
41. Heinemann, A.W.; Kinnett-Hopkins, D.; Mummidisetty, C.K.; Bond, R.A.; Ehrlich-Jones, L.; Furbish, C.; Field-Fote, E.; Jayaraman, A. Appraisals of robotic locomotor exoskeletons for gait: Focus group insights from potential users with spinal cord injuries. *Disabil. Rehabil. Assist. Technol.* **2020**, *15*, 1–11. [[CrossRef](#)] [[PubMed](#)]
42. Lajeunesse, V.; Routhier, F.; Vincent, C.; Lettre, J.; Michaud, F. Perspectives of individuals with incomplete spinal cord injury concerning the usability of lower limb exoskeletons: An exploratory study. *Technol. Disabil.* **2018**, *30*, 63–76. [[CrossRef](#)]
43. Vitiello, N.; Cempini, M.; Crea, S.; Giovacchini, F.; Cortese, M.; Moisé, M.; Posteraro, F.; Carrozza, M.C. Functional design of a powered elbow orthosis toward its clinical employment. *IEEE-ASME T. Mech.* **2016**, *21*, 1880–1891. [[CrossRef](#)]
44. Bryce, T.N.; Dijkers, M.P.; Kozlowski, A.J. Framework for assessment of the usability of lower-extremity robotic exoskeletal orthoses. *Am. J. Phys. Med. Rehabil.* **2015**, *94*, 1000–1014. [[CrossRef](#)]
45. Liu, D.; Chen, W.; Pei, Z.; Wang, J. A brain-controlled lower-limb exoskeleton for human gait training. *Rev. Sci. Instrum.* **2017**, *88*, 1–13. [[CrossRef](#)] [[PubMed](#)]
46. Deeny, S.; Chicoine, C.; Hargrove, L.; Parrish, T.; Jayaraman, A. A simple ERP method for quantitative analysis of cognitive workload in myoelectric prosthesis control and human-machine interaction. *PLoS ONE* **2014**, *9*, e112091. [[CrossRef](#)] [[PubMed](#)]
47. Torricelli, D.; Gonzalez-Vargas, J.; Veneman, J.; Mombaur, K.; Tsagarakis, N.; del Ama, A.; Gil-Agudo, A.; Moreno, J.; Pons, J.L. Benchmarking bipedal locomotion: A unified scheme for humanoids, wearable robots, and humans. *IEEE Rob. Autom. Mag.* **2015**, *22*, 103–115. [[CrossRef](#)]