

# An Analysis of Usability Evaluation Practices And Contexts of Use In Wearable Robotics

Jan Thomas Meyer (✉ [jan.meyer@hest.ethz.ch](mailto:jan.meyer@hest.ethz.ch))

Swiss Federal Institute of Technology in Zurich <https://orcid.org/0000-0002-0361-1152>

Roger Gassert

Swiss Federal Institute of Technology: Eidgenössische Technische Hochschule Zurich

Olivier Lambercy

Swiss Federal Institute of Technology Zurich: Eidgenössische Technische Hochschule Zurich

---

## Research

**Keywords:** wearable robotics, usability evaluation, user-centered design, exoskeletons

**Posted Date:** August 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-764275/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Journal of NeuroEngineering and Rehabilitation on December 1st, 2021. See the published version at <https://doi.org/10.1186/s12984-021-00963-8>.

RESEARCH

# An Analysis of Usability Evaluation Practices and Contexts of Use in Wearable Robotics

Jan Thomas Meyer<sup>1\*</sup>, Roger Gassert<sup>1,2</sup> and Olivier Lambercy<sup>1,2</sup>

<sup>1</sup> Rehabilitation Engineering Laboratory, Department of Health Sciences and Technology, ETH Zurich, Switzerland

<sup>2</sup> Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence And Technological Enterprise (CREATE), Singapore

## Correspondence:

elab.publications@hest.ethz.ch

Rehabilitation Engineering

Laboratory, Department of Health

Sciences and Technology, ETH

Zürich, Switzerland

Full list of author information is

available at the end of the article

**Abstract**

**Background:** User-centered design approaches have gained attention over the past decade, aiming to tackle the technology acceptance issues of wearable robotic devices to assist, support or augment human capabilities. While there is a consensus that usability is key to user-centered design, dedicated usability evaluation studies are scarce and clear evaluation guidelines are missing. However, the careful consideration and integration of user needs appears to be essential to successfully develop an effective, efficient, and satisfactory human-robot interaction.

**Methods:** Through an online survey for developers of wearable robotics, we wanted to understand how the design and evaluation in actual daily practice compares to what is reported in literature. With a total of 31 questions, we analyzed the most common wearable robotic device applications and their technology maturity, and how these influence usability evaluation practices.

**Results:** A total of 158 responses from a heterogeneous population were collected and analyzed. The dataset representing contexts of use for augmentation (16.5%), assistance (38.0%), therapy (39.8%), as well as few other specific applications (5.7%), allowed for an insightful analysis of the influence of technology maturity on user involvement and usability evaluation. We identified functionality, ease of use, and performance as the most evaluated usability attributes and could specify which measures are used to assess them. Also, we could underline the frequent use of qualitative measures alongside the expected high prevalence of performance-metrics. In conclusion of the analysis, we derived evaluation recommendations to foster user-centered design and usability evaluation.

**Conclusion:** This analysis might serve as state-of-the-art comparison and recommendation for usability studies in wearable robotics. We believe that by motivating for more balanced, comparable and user-oriented evaluation practices, we may support the wearable robotics field in tackling the technology acceptance limitations.

**Keywords:** wearable robotics; usability evaluation; user-centered design; exoskeletons

### 3 Background

4 Wearable robotic devices (WRD) to assist, support, or augment human functions  
5 and activities are gaining popularity and practicability. A great diversity from fully  
6 wearable and soft, to stationary and rigid devices has been developed over the  
7 past decades, showing promising ranges of functionalities [1, 2, 3]. Nonetheless, the  
8 number of WRD in daily use is still very low due to missing availability on the  
9 market, as well as technology acceptance limitations [1, 4, 5]. More specifically,  
10 the usability - defined as “extent to which a system, product or service can be  
11 used by specified users to achieve specified goals with effectiveness, efficiency, and  
12 satisfaction in a specified context of use” [6] - appears to be a key factor limiting  
13 the market translation and technology acceptance of WRD. Particularly, devices for  
14 medical applications such as robot-aided therapy or advanced assistive technologies  
15 often struggle to comply with the complex contexts of use, as users with health  
16 issues have more specific and personalized needs [7, 8, 9, 10].

17 A solution to tackle and minimize usability limitations is user-centered design  
18 (UCD), which aims to involve the technology stakeholders throughout the device  
19 development to more successfully address and meet user needs [11, 12, 13]. As  
20 part of the iterative UCD process, a structured evaluation of the technical solu-  
21 tion solving a specific human problem is considered essential [14, 15, 16, 17]. The  
22 evaluation of WRD is a challenging endeavor with many facets to consider, rang-  
23 ing from technical characteristics to human factors. From a technical perspective,  
24 device effectiveness and safety are crucial when developing robots that assist, sup-  
25 port, or augment human activities. On the other hand, the users themselves bring  
26 physical and psycho-social factors into play, such as varying skills, knowledge, prior  
27 experiences or expectations, which need to be investigated and valued as design  
28 criteria with equal importance [7, 18]. Unfortunately, there is a lack of standards  
29 and guidelines for the evaluation of the complex human-robot interaction of WRD.

30 More specifically, the application of relevant and appropriate usability evaluation  
31 measures remains a fundamental challenge in wearable robotics development [19].  
32 Recent studies analyzing WRD evaluation practice have shown a vast landscape of  
33 measures being used, with no apparent best-practice identifiable [20, 21, 22]. De-  
34 spite application-specific efforts to address these limitations with novel evaluation  
35 frameworks [23, 24, 25], there still is a need for more general guidelines on the  
36 best-practice for the development and usability evaluation of wearable robots.

37 In this work, we report the insights from an online survey sent to developers  
38 of WRD from academic, industrial, and clinical backgrounds to analyze current  
39 practices in usability evaluation, considering contexts of use, technology maturity  
40 and methods used for evaluation. The objectives of the survey were to investigate: 1)  
41 whether specific contexts of use for WRD are more advanced in technology maturity  
42 than others, 2) whether usability evaluation with active user involvement is current  
43 practice in WRD development, as expected from UCD, 3) whether current usability  
44 evaluation practices are predominantly device-focused, i.e. limited in capturing the  
45 user's perspective, and 4) whether best-practice for WRD usability evaluation can  
46 be inferred from the most frequently reported methods and measures in the field.  
47 Our analysis provides a novel understanding of the predominant WRD contexts  
48 of use, current technological maturity of the field and current best-practice for  
49 usability evaluation. From our data, we propose recommendations that could help  
50 harmonize usability evaluation of WRD, and help tackling the acceptance issue of  
51 wearable robotics.

## 52 **Methods**

### 53 **Study Design**

54 The data for this study was collected with an online survey that was administered  
55 using the QuestionPro survey software (QuestionPro Inc., Austin, TX, USA). A  
56 total of 31 questions, primarily close-ended and multiple-choice, covered three topics

57 of interest related to wearable robotics: (a) context of use, (b) technology maturity  
58 and user involvement, and (c) usability evaluation practice. The total number of  
59 questions for each respondent depended on the answers given, as logic branching  
60 with follow-up questions was applied. The study aims, usage of data and other  
61 informed consent information was provided on the survey landing page. The full  
62 survey including all questions is provided in the Additional file 1.

### 63 *Context of Use*

64 Building upon definitions of the International Organization for Standardization  
65 (ISO) as terminological ground truth, usability is dependent on the context in which  
66 the systems, in our case WRD, are used and investigated. Therefore, we aimed to  
67 first understand which contexts of use the collected data represented, such that we  
68 could investigate our research questions on development and evaluation practices.  
69 The context of use can be specified as “combination of users, goals and tasks, re-  
70 sources, and environment” [6]. The respondents were asked to specify the general  
71 usage purpose of their device, alongside of the usage environments, the forms of  
72 supervision needed, as well as the target populations.

### 73 *Technology Maturity and User Involvement*

74 To further understand the technology maturity of their WRD, respondents were  
75 asked to specify their Technology Readiness Level (TRL). The nine distinct TRL  
76 were introduced as follows, adapted from the Horizon 2020 guideline: TRL 1 = Basic  
77 research and principles observed, TRL 2 = Technology concept formulated, TRL 3  
78 = Experimental proof of concept, TRL 4 = Technology tested in lab environment,  
79 TRL 5 = Technology tested in intended environment, TRL 6 = Technology validated  
80 and demonstrated in intended environment, TRL 7 = Demonstration in operational  
81 environment, TRL 8 = System complete and ready for commercialization, TRL 9 =  
82 Full commercial application [26]. For the purpose of congregated data analysis, we  
83 grouped the TRL into three Technology Readiness Phases (TRP): ”Concept” (TRL

84 1 – 3), "Prototype" (TRL 4 – 7), and "Product" (TRL 8 – 9). Also, respondents  
85 were asked to specify the time since project initiation, as a second measure of project  
86 maturity. Lastly, the number of users who tested the reported WRD was requested  
87 to understand the extent of user involvement, and to validate the TRL estimations.  
88 The respondents were thereby asked to distinguish between target users (real end-  
89 users) and mock users (neurologically intact controls, team-members, themselves).

### 90 *Usability Evaluation Practice*

91 A core interest of this study was the investigation of evaluation practices for WRD  
92 and, more specifically, of measures used for usability evaluation. From the ISO ter-  
93 minology, we can define usability by the three dimensions effectiveness, efficiency,  
94 and satisfaction. Effectiveness reflects "the accuracy and completeness with which  
95 users achieve specified goals", efficiency represents the "resources (time, human ef-  
96 fort, costs & materials) used in relation to the results achieved", and satisfaction  
97 is the "extent to which the user's physical, cognitive and emotional responses that  
98 result from the use of a system, product, or service meet the user's needs and ex-  
99 pectation" [6]. In order to understand their evaluation focus, the respondents were  
100 asked to distribute their current evaluation efforts as a total of 100% to the three  
101 usability dimensions. Although the usability dimensions narrow down the room for  
102 terminological interpretation, specific attributes such as comfort, functionality, ease  
103 of use, and safety are more frequently used to collect and analyze user feedback or  
104 device performance [21]. As one of the aims of this study was to understand how  
105 WRD developers define and approach usability evaluation, we provided a list of  
106 34 popular usability attributes (full list in additional materials). Respondents were  
107 asked to pick up to five attributes on which they are focusing their usability evalu-  
108 ation. If their preference was not among the listed attributes, new entries were pos-  
109 sible. Adapted from the ISO TR 16982-2002, eight specific usability methods were  
110 then proposed to the respondents: (1) Performance-related measurement, (2) Ques-

111 tionnaire/Survey, (3) Interview, unstructured oral feedback, (4) Thinking Aloud,  
112 (5) Observation of users, (6) Document-based method, (7) Model- or simulation-  
113 based approach, (8) (Usability) Expert evaluation [27]. For each previously selected  
114 usability attribute, the respondents were asked to state which of these methods  
115 were used to assess it. If, at any point, the respondents selected the method types  
116 Performance-related measures and/or Questionnaires/Surveys, follow-up questions  
117 were generated asking the respondents to specify which exact metric or measure  
118 was used. The last few questions of the survey allowed the respondents to reflect on  
119 their usability evaluation practice. Respondents were asked to state their level of  
120 agreement from 1 (completely disagree) to 5 (completely agree) on usability eval-  
121 uation usefulness, availability of benchmarks, and preference towards certain types  
122 of usability data.

### 123 Sample

124 The target population of this study were developers of WRD with academic, indus-  
125 trial, and/or clinical backgrounds. The respondents were instructed to only partic-  
126 ipate in the survey if they work on, or develop a WRD and to focus on one specific  
127 device during survey completion. The survey started after the respondents agreed  
128 to the terms and conditions. The survey link was distributed globally using mailing  
129 lists, social media (LinkedIn, Twitter), as well as blogs featured on Exoskeleton  
130 Report [28] and on Biomed Central [29]. The targeted sample size was 100 fully  
131 completed surveys, in order to collect a meaningful and divers sample of WRD  
132 projects and their usability practices. Data were collected from June to October  
133 2020.

### 134 Data Analysis

135 The criterion for incomplete responses to be included in the analysis was the min-  
136 imum completion of survey sections on demographics information, context of use,  
137 and user involvement. Responses that did not fulfill the inclusion criteria were re-

138 moved using the data analysis functions of QuestionPro. The eligible data was  
139 exported as EXCEL data reports for post-processing and analysis. All statistical  
140 analyses and visualizations were done in MATLAB R2020a (MathWorks, MA, USA)  
141 and in RStudio Team 2021 (RStudio PBC, MA, USA). Descriptive statistics such  
142 as frequency distributions and cross-tabulations were used to analyze and visualize  
143 the survey responses. For intra-response comparisons, paired t-test were used to,  
144 e.g., investigate individual respondents' evaluation protocols and focuses. For all  
145 inter-response comparisons, a two-sample t-test to account for unequal sample sizes  
146 in the analyzed subgroups was used. All statistical tests were performed with a 5%  
147 significance level. Locally weighted regression fits (Loess regression) were used for  
148 scatter plot smoothing.

## 149 **Results**

150 From a total of 286 initiated responses, 158 fulfilled the inclusion criteria and 102  
151 were fully completed (35.6% completion rate). Due to logic-branching and par-  
152 tially missing answers, certain questions had a lower sample number than the total.  
153 Therefore, the sample number ( $n$ ) is provided for each analysis. Out of the 158  
154 respondents, 33.5% were female, and 79.1% were between 25 to 44 years old. The  
155 dataset contained WRD projects from all continents except Africa and Antarc-  
156 tica. The background of the respondents ranged from academia (71.5%) to industry  
157 (32.3%) and clinical practice (15.2%), while 28 of the 158 respondents selected more  
158 than one background. Detailed participant demographics are described in Table 1.

## 159 **Context of Use**

160 The responses covered a large diversity in applications and target user groups. A  
161 summary of the context of use information is visualized in Figure 1. The general  
162 usage purposes of the WRD (Figure 1A) are grouped into four categories: Aug-  
163 mentation (16.5%), Assistance (38.0%) Therapy (39.8%), and Other (5.7%). The  
164 grouping of these four general usage purposes was used for most analyses and vi-

Table 1: Respondent Demographics (n = 158)

	Frequency	Percentage
<b>Gender</b>		
Male	105	66.5%
Female	53	33.5%
<b>Age</b>		
18-24	13	8.2%
25-34	86	54.4%
35-44	39	24.7%
45-54	7	4.4%
55-64	10	6.3%
Above 64	3	1.9%
<b>Background*</b>		
Academia	113	71.5%
Industry	51	32.3%
Clinical practice	24	15.2%
<b>Location of WRD project</b>		
Europe	88	55.7%
North America	37	23.4%
Asia	24	15.2%
Latin America	7	4.4%
Oceania	2	1.3%
Africa & Antarctica	0	0.0%

\*More than one background could be selected

165 sualizations. The intended form of supervision visualized in Figure 1B shows that  
 166 76 out of 158 WRD aim for unsupervised, independent use. According to our re-  
 167 sponses, WRD for therapy applications are intended for use with a certain level  
 168 of aid (33.3%) or full supervision (46.0%), while the majority of applications for  
 169 augmentation and assistance are envisioned to be used independently (65.4% and  
 170 71.7% respectively)

Figure 1: **Context of Use of Wearable Robotic Devices: A) General usage purposes:** augmentation, assistance, therapy and other. \*Other usage purposes reported were: *all of the above* (n = 3), *brain computer interfaces* (n = 2), *fitness and sports tracking* (n = 1), *training and assistance for surgery* (n = 1), *benchmarking* (n = 1) and *user research* (n = 1), **B) supervision form:** We can differentiate between fully supervised, partially-supervised (aided) and unsupervised use of WRD, **C) Target body areas:** The relative frequencies of the four general usage purposes are reported for each body area **D) Usage environment:** The intended use ranges from rather controlled (laboratory, clinic, home) to more dynamic (leisure, military, work) environments **E) Target population of medical applications:** For the medical applications (therapy and assistance) respondents reported the specific target groups. \*\*Other target groups where as: *visually impaired, essential tremor, trauma, rhabdomyolysis, peripheral artery disease, first responders (emergency), physical training and exercise, and sports injuries.*

171 The most targeted body area by current WRD were the lower limbs (n = 81),  
 172 out of which 82.7% reported a use for medical applications (Figure 1C). In the  
 173 second most reported body area, the upper limbs, we can differentiate between full  
 174 arm applications (distal + proximal, 18.3%), and specifically distal applications for  
 175 the hand, wrist and/or lower arm (22.3%). The intended usage environment of the  
 176 investigated WRD (Figure 1D) varied greatly depending on the intended usage pur-  
 177 pose. While devices for augmentation mostly focus on work applications, assistive  
 178 WRD aim to also help at home, during leisure activities, and within clinical use.  
 179 Within the 63 therapy-oriented devices, 80% are designed to be used in clinics with  
 180 additional usage intentions at home or in research (both 42.8%). Across all reported  
 181 WRD, the homes of end-users users appeared to be the most targeted usage envi-

182 ronment (52.5%) followed by clinics (50.6%) and laboratory (37.3%) applications.  
183 Figure 1E depicts the targeted disorders and disabilities of all WRD for medical  
184 applications ( $n = 132$ ). Stroke survivors built the largest user group across all re-  
185 sponses, targeted by 63.6% of all medical WRD. Especially for therapy devices,  
186 rehabilitation after stroke is targeted by 85% of all devices. The second largest user  
187 group addressed by almost half of all medical applications (48.5%) are people with  
188 spinal cord injuries. Out of 132 medical WRD applications, 59.8% are intended for  
189 more than one specific population. For example, sensorimotor impairments resulting  
190 from stroke or spinal cord injury are often targeted with the same WRD.

#### 191 Technology Maturity and User Evaluation

192 The TRL distribution of 143 WRD is visualized in Figure 2. The four usage purposes  
193 showed a similar distribution across the TRL spectrum. The largest total number  
194 of WRD ( $n = 40$ , 28.0%) classified as TRL 4. Ordered by the mean TRL as usage  
195 purpose maturity indicator, the most market ready applications are Other ( $6.11$   
196  $\pm 1.62$ ) and Therapy ( $5.64 \pm 2.29$ ) followed by augmentation ( $5.23 \pm 2.43$ ) and  
197 Assistance ( $4.77 \pm 2.18$ ). Overall, 60.8% of the reported WRD were classified as  
198 prototypes, 20.3% as products, and 18.9% as concepts.

Figure 2: **Technology Maturity of Current Wearable Robotic Devices:**  
The Technology Readiness Levels (TRL) of the four general usage purposes,  
visualized as stacked histogram ( $n = 143$ ). The TRL were grouped in three  
Technology Readiness Phases (TRP): “Concept” (TRL 1 - 3), “Prototype” (TRL  
4 - 7) and “Product” (TRL 8 - 9).

199 In terms of project duration, 41.9% of WRD projects were initiated within the  
200 last two years preceding survey completion, while 23.2% have been ongoing for more  
201 than five years. Of all 158 WRD projects recorded, 108 (68.4%) tested their device  
202 with at least one target user so far. From those 108, 48.1% additionally tested with  
203 mock users. No tests with the target population, but only with mock users was

204 performed in 17.1% of the 158 WRD. Overall, 125 of 158 WRD (85.4%) reported  
205 some sort of user testing. Figure 3 shows the number of tested users, according to  
206 the TRL reported ( $n = 143$ ).

Figure 3: **Progression of User Involvement with Technology Maturation**  
The number of users (target and mock) per Technology Readiness Level (TRL)  
are shown ( $n = 143$ ). The size of the circles represents the number of responses  
within each TRL. For all respondents that only indicated testing with the one  
specific user group (target or mock), 0 users are shown in the opposite group.  
The Loess regression lines were fitted to smooth the scatter plot visualization.

### 207 Usability Evaluation Practice

208 The results of the usability evaluation efforts allocation ( $n = 117$ ) are shown in  
209 Figure 4A. The paired comparisons between the usability dimensions within each  
210 Technology Readiness Phase (TRP) showed significant differences in terms of efforts  
211 allocations: in the conceptual phase, significantly fewer evaluation effort was dedi-  
212 cated to satisfaction compared to effectiveness ( $p < 0.001$ ) and efficiency ( $p < 0.001$ ).  
213 When evaluating prototypes, effectiveness remained the most dominant usability  
214 dimension over satisfaction ( $p < 0.001$ ) and efficiency ( $p < 0.001$ ). In the product  
215 phase, only the difference of evaluation efforts split between effectiveness and ef-  
216 ficiency showed a significant difference ( $p < 0.001$ ). In the unpaired, two-sample  
217 comparison of each usability dimension evaluation, we can observe that satisfaction  
218 gained focus in the prototyping ( $p < 0.05$ ) and product ( $p < 0.05$ ) phases compared  
219 to the conceptual phase. In contrast, the relative efforts to evaluate efficiency ap-  
220 peared reduce as technology matures (concept to prototype,  $p < 0.001$ ; concept to  
221 product,  $p < 0.001$ ). The share of effectiveness in the evaluation efforts of WRD  
222 remained constantly high between 42.9% - 45.7%.

223 A frequency analysis of the top 15 selected usability attributes (ordered from most  
224 to least selected) and their respective evaluation methods is summarized in Table  
225 2. The most frequently evaluated usability attributes were functionality (37.6%),

Figure 4: **Analysis of Usability Evaluation Practice: A) Evaluation efforts allocation per Technology Readiness Phase (TRP, n = 117):** The allocated total of 100 evaluation-effort-points among the usability dimensions effectiveness, efficiency, and satisfaction, per TRP are shown. Paired comparison between the three dimension were analyzed within each TRP, while unpaired, two-sample comparisons between the TRP were calculated. Levels of significance indicated as: \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ . **B) List of reported performance-related measurements (PRM, n = 88):** \*Other PRM described as *spatiotemporal metric analysis, eye-tracking recording, number of steps*. \*\*Standardized functional measures (SFM) specified: upper limb SFM; Box and Block Test (BBT, n = 6), Jebsen-Taylor Hand Function Test (JTHFT, n = 5), Action Research Arm Test (ARAT, n = 4), Chedoke Arm and Hand Activity Inventory (CAHAI, n = 2), Southampton Hand Assessment Procedure (SHAP), Assisting Hand Assessment (AHA), Smart Pegboard, Frenchay Arm Test (FAT), lower limb SFM; 10 Meter Walk Test (10MWT, n = 2), 2 Minute Walk Test (2MWT, n = 2), 6 Minute Walk Test (6MWT, n = 2), Timed-up-and-Go (TUG), general SFM; Human-Robot Fluency Metrics, Assessment of Capacity for Myoelectric Control (ACMC), Thermography, Failure Mode and Effects Analysis (FMEA), ISO regulation, fit and tolerance assemblies

226 ease of use (36.8%), performance (32.0 %), safety (32.0 %), and comfort (29.6%).  
 227 The least selected attributes were learnability (4.0%), mental demand (3.2%), and  
 228 understandability (2.4%). The most reported usability evaluation method across  
 229 all attributes are performance-related measurements (PRM), followed by question-  
 230 naires and observation of users, while the least used appear to be Thinking Aloud  
 231 and document-based methods. PRM were reported to be used to evaluate almost all  
 232 of the listed usability attributes (94.1%) except for complexity and mental demand.  
 233 Attributes of a more subjective nature, such as comfort, ergonomics, wearability,  
 234 adaptability, and intuitiveness appear to be primarily evaluated with qualitative  
 235 measures (unstructured interviews, observations) and questionnaires/surveys. The  
 236 full table with all 34 attributes is provided in Additional file 2.

237 Figure 4B shows the selection frequencies of specific PRM, ordered from most to  
 238 least used. Task success & failure (70.5%) and time for task (67.0%) were reported  
 239 as the most popular PRM across all general usage purposes. From the provided

Table 2: Usability Attributes and Evaluation Methods (n = 125)

Attributes	Times selected (%)	Performance-related measurements	Questionnaire, Survey	Interview, unstructured oral feedback	Thinking Aloud	Observation of users	Document-based methods	Model- or simulation-based approach	(Usability) Expert evaluation
Functionality	47 (37.6%)	31	17	15	7	18	4	10	10
Ease of use	46 (36.8%)	18	27	23	12	20	7	5	11
Performance	40 (32.0%)	32	12	7	6	15	5	12	5
Safety	40 (32.0%)	14	11	11	8	17	9	8	9
Comfort	37 (29.6%)	8	24	17	4	10	3	2	7
Benefit	26 (20.8%)	18	15	9	4	9	3	4	4
Reliability	25 (20.0%)	18	5	5	2	6	8	7	7
Ergonomics	23 (18.4%)	4	9	5	3	12	7	3	5
Technical req.	22 (17.6%)	15	1	3	2	7	6	12	8
Wearability	22 (17.6%)	7	9	9	5	12	8	5	7
Adaptability	21 (16.8%)	4	6	11	7	9	3	3	3
Meet user needs	20 (16.0%)	4	14	11	3	6	3	2	7
Autonomy	16 (12.8%)	10	5	5	5	4	2	4	3
Feasibility	16 (12.8%)	9	4	5	2	5	2	3	1
Intuitiveness	16 (12.8%)	5	8	11	4	4	2	1	3
Others*		(57)	(62)	(55)	(25)	(55)	(19)	(24)	(34)
<b>Total count</b>		<b>254</b>	<b>229</b>	<b>202</b>	<b>99</b>	<b>209</b>	<b>91</b>	<b>105</b>	<b>124</b>

\*Full list of attributes in Additional file 2

240 selection of specific PRM, standardized functional measures, such as the 10 Meter  
241 Walk Test or the Box and Block Test were reported to be used the least (35.2%).  
242 The highest relative usage of standardized tools was observed for medical applica-  
243 tions. Kinetic and kinematic measures such as interaction forces, or spatiotemporal  
244 parameters appear to be most valued in the evaluation of WRD for therapy.

245 The results from 74 respondents who specified which questionnaires and scales  
246 they use are summarized in Figure 5. Custom-made questionnaires made with Lik-  
247 ert Scales (LS), Visual Analogue Scales (VAS), Numeric Rating Scales (NRS), or  
248 open text questions were most frequently reported. Overall, 73.0% of all respondents  
249 indicated the use of at least one custom-made questionnaire. Still, most evaluations  
250 protocols appear to be a mix of standardized and custom-made measures, as only  
251 25.7% of the respondents indicated the sole use of custom tools. From 74 responses,  
252 22 (29.7%) only chose one specific measure, while 38 (51.3%) selected three or more  
253 from the provided list or added additional ones specified in "Other". The System  
254 Usability Scale (SUS, 25.7%), NASA Task Load Index (TLX, 16.2%) and Quebec  
255 User Evaluation of Satisfaction with Assistive Technology 2.0 (QUEST 2.0, 16.2%)  
256 represent the most frequently selected standardized and validated questionnaires.  
257 Modifications of standardized scales are also frequently used, as the modified SUS  
258 and the raw, unweighted TLX were both selected by more than 10% of the respon-  
259 dents.

260 Lastly, the insights from the reflections on personal usability evaluation practices  
261 are summarized in Figure 6. More than half of all WRD developers showed a certain  
262 level of disagreement with the two statements "I was able to compare my evaluation  
263 data with state-of-the-art benchmarks" (52.3% disagreement,  $n = 109$ ) and "It was  
264 easy to find standardized measures for my context of use" (50.5% disagreement,  $n$   
265  $= 93$ ). From the 109 respondents that completed this last part of the survey, 72.5%  
266 agreed that usability evaluation improved their WRD. When asked if they would

Figure 5: **Questionnaires and Scales Used in Wearable Robotics Usability Evaluation:** The 20 most frequently selected questionnaires and scales are displayed (n = 74). Custom-made forms are marked in dark red, modified questionnaires are marked in light red. **Abbreviations:** LS = Likert Scale, VAS = Visual Analogue Scale, SUS = System Usability Scale, NRS = Numeric Rating Scale, NASA TLX = NASA Task Load Index, QUEST 2.0 = Quebec User Evaluation of Satisfaction with Assistive Technology, mSUS = modified System Usability Scale, RTLX = Raw Task Load Index, ASQ = After Scenario Questionnaire, PUEU = Perceived Usefulness, Perceived Ease of Use, ATD PA = Assistive Technology Device Predisposition Assessment, PSSUQ = Post-Study Usability Questionnaire, COPM = Canadian Occupational Performance Measure, USAT = Usability Scale for Assistive Technology, IMI = Intrinsic Motivation Inventory, USEQ = Usefulness, Satisfaction, Ease of Use Questionnaire, PYTHEIA = Psychometric Scale to Assess the Satisfaction of Users with Assistive Technology, QUIS = The Questionnaire for User interaction Satisfaction. **\*Other** = Borg Scale of Perceived Exertion, Michigan Hand Outcomes Questionnaire, Prosthesis Evaluation Questionnaire, Psychosocial Impact of Assistive Device, Questionnaire to Explore Human Factors and their Technical potential, Quick-DASH, SF-36, Telehealthcare Satisfaction Questionnaire – Wearable Technology, Trinity Amputation and Prosthesis Experience Scale, Usability Metric for User Experience, and Embodiment Questionnaire.

267 prefer custom-made measures over standardized tools, most participants selected  
 268 the neutral option. The respondents showed the second lowest level of disagreement  
 269 with the statement “I prefer quantitative over qualitative data” (18.9% disagree-  
 270 ment, n = 109).

Figure 6: **Reflection of Usability Evaluation Practise:** The level of agreement for each statement is shown. Disagreement is red, agreement is green. The more the entire bar (100%) is shifted to the left or right, the clearer a trend of agree- or disagreement is observable among the responses.

## 271 Discussion

272 The aim of this work was to investigate current landscape in WRD applications,  
 273 the extent of their user-centered design practice, as well as their usability evalua-  
 274 tion practices. We collected information on contexts of use, technology maturities,  
 275 and usability evaluation strategies from a heterogeneous population of 158 WRD

276 developers to provide an analysis of trends and best-practices for the development  
277 of wearable robotics.

#### 278 Contexts of Use and Maturity of Wearable Robots

279 The proposed breakdown of general usage purposes, target users, usage environ-  
280 ments, and supervision forms provides a realistic glimpse into the current applica-  
281 tion landscape of wearable robotics. From the respondents' data, we can conclude  
282 that applications for robot-aided therapy and daily assistance of people with func-  
283 tional disabilities currently build the largest use cases of WRD. In combination with  
284 the TRL information, we understand that devices for daily-life assistance are in a  
285 comparably younger maturity state compared to the other usage purposes. Robot-  
286 aided therapy appears to be more mature overall, with a larger number of devices  
287 close to - or already available on the market. In our collected sample, devices for  
288 augmentation appear to be a minority of WRD use cases, with only 16.5% of the  
289 investigated devices aiming for, e.g., prevention of work-related musculoskeletal dis-  
290 orders in industrial applications. Most reported augmentation solutions target to  
291 support the torso (e.g., support lifting tasks) or proximal upper limbs (e.g., over-  
292 head work). Combining all insights from the context of use data, we can identify  
293 the three currently most popular contexts of use in the WRD field, ranked by their  
294 current technology maturity and success in translation to daily use: (1) Robot-aided  
295 therapy applications for the upper and lower extremities in a supervised (clinical or  
296 research) environment, (2) Augmentation for labor-intensive tasks in industrial, un-  
297 supervised workplaces, and (3) Unsupervised, independent use of wearable robotic  
298 assistive technology in the home environment.

299 Even though the WRD field initially emerged from industrial, i.e., augmentation  
300 applications for the upper limbs [30], the predominant number of lower limb WRD is  
301 no surprise. A majority of WRD developments and research of the past two decades  
302 focused on the challenge of restoring or assisting human gait [1, 3]. Pioneering de-

303 vices such as the robotic-driven gait orthosis Lokomat [31] have triggered a shift  
304 from initially more augmentation-oriented towards medical use cases, demanding a  
305 more user-centered focus on human factors [15, 30]. However, WRD for upper limb,  
306 and torso augmentation have regained focus in the last years, as more lightweight  
307 and simple solutions using passive, or semi-active actuation principles have found  
308 their niche in industrial scenarios [5]. Also, robotic hand orthoses for both, assis-  
309 tance and therapy, have gained momentum in the last decade. Interestingly, a recent  
310 review on the technological maturity of such hand orthoses provided very similar  
311 insights in terms of TRL which indicates that our data might realistically represent  
312 the current WRD maturity state [32]. In both analyses, TRL 4 (= technology tested  
313 in lab environment) appears to be a bottleneck for numerous WRD developments,  
314 indicating the first implementation barrier between basic research and application,  
315 also known as “valley of death”. Developments beyond TRL 4 are likely to ex-  
316 ceed the basic research questions of academia, and the proof-of-concept required for  
317 technical publication. An unfortunate consequence is that the majority of promis-  
318 ing WRD projects never make their way out of laboratory research, which also  
319 requires additional substantial resources. By supporting more application-oriented,  
320 translational research, academia could push WRD maturity towards TRL 6 or 7,  
321 and overcome the translational gap to materialize the potential of wearable robotics  
322 [33].

323 Robotic assistive technology was reported with an overall lower technology ma-  
324 turity than all other usage purposes, which in parts contradicts the rich history  
325 of upper and lower limb prosthetics as assistive mobility devices. This observation  
326 might be explained by terminological inconsistencies, as the term WRD is often used  
327 as a synonym for exoskeletons [30]. Exoskeletons are robotic orthoses, which so far  
328 showed rather limited technology acceptance in unsupervised, assistive applications  
329 [34, 35]. Defining WRD and exoskeletons as synonyms would therefore wrongfully

330 exclude prosthetics, which in fact make up 43% of all emerging mobility assistive  
331 technologies, while exoskeletons only occupy 19% of that specific market [36]. Pros-  
332 thetics – and their advanced technology readiness - might thus be under-sampled  
333 in our dataset (22 out of 158 responses).

334 Another point worth highlighting is the form of supervision, which has been only  
335 minimally discussed in existing WRD reviews [1, 4, 5]. Here, our data indicates  
336 that, e.g., WRD for therapy might be a mature WRD technology due to the fact  
337 that their (intended) application is a controlled environment under the supervision  
338 or aid of trained personnel. Hence, WRD for therapy are mainly used and acquired  
339 by institutions such as rehabilitation facilities, while WRD for assistance and aug-  
340 mentation aim towards independent use by individuals at home, work, or for leisure  
341 activities. This implies that WRD for assistance and augmentation have to work  
342 reliably in uncontrolled and more dynamic environments. We can therefore argue  
343 that one factor limiting the usability, acceptance, and translation of WRD for aug-  
344 mentation and assistance appears to be their goal of unsupervised, independent use.  
345 Moreover, the complexity of WRDs appears to be among a strong limiting factors  
346 of technology adoption, even for a supervised use in therapeutic applications [37].  
347 Developing a robotic device with high functionality while keeping the design simple  
348 enough to be set up, used and maintained remains a central challenge in WRD. A  
349 strong development focus on UCD and usability evaluation could help overcoming  
350 this challenge.

### 351 User Involvement and Evaluation Focus

352 The first step towards UCD is active user involvement. Although users should be  
353 involved in all development phases, this survey focused on the evaluation phase.  
354 More specifically we analyzed the number of users our respondents have tested, i.e.  
355 evaluated their WRD solution with, and which methods they used in this process.  
356 Our results show that the number of target users involved increases steadily with

357 technology maturation. Testing with mock users reaches a plateau in the prototype  
358 stage, but over-weights target user involvement until TRL 4, also because almost  
359 half of all projects claiming TRL 4 maturity did not test with a single target user  
360 yet. This practice may likely come from the circumstances that target users are  
361 only involved once a WRD concept has proven effective, and that certain regula-  
362 tory, or safety measures limit the access to target users. Also, limited resources -  
363 as discussed above - might be an additional factor limiting target user involvement  
364 in early development stages. However, also a substantial amount of WRD projects  
365 in the later prototyping stages (TRL 5, 6, 7) reported to have involved less than  
366 five target users, which might explain usability issues encountered when eventually  
367 aiming for commercialization. This was somewhat confirmed by the evaluation ef-  
368 forts allocations, as we learned that effectiveness remains the primarily investigated  
369 usability dimension across the entire technology maturity continuum. Satisfaction,  
370 which is by nature a more target user-focused value, only gains focus once the  
371 WRD comes closer to the product stage, and efficiency appears to even loose focus  
372 as technology matures. These insights suggest that a large number of WRD projects  
373 involve target users at a later stage only, while first focusing on the effectiveness of  
374 their solution.

375 The practice of device-oriented development and evaluation has been previously  
376 highlighted by works of Contreras-Vidal *et al.* (2016) and Pinto-Fernandez *et al.*  
377 (2020) who both summarized that in lower limb exoskeletons studies, evaluation out-  
378 comes on comfort, ergonomics, satisfaction, and/or mental demand are drastically  
379 scarce compared to performance-related outcomes [22, 38]. In a survey exploring  
380 user involvement and device evaluation strategies, Ármannsdóttir *et al.* (2020) also  
381 reported functionality as the most reported device performance indicator [21]. An  
382 evaluation focus on WRD effectiveness seems somewhat rational - especially in early  
383 development stages - as developers first want to make sure the device does what it is

384 supposed to do, and this in a safe, reliable, and accurate way. Another reason for the  
385 higher prevalence of device-oriented developments and their reporting in literature  
386 may come from the observations that qualitative metrics are less well-accepted as  
387 outcomes of WRD usability studies. The potential reliability, reproducibility and bi-  
388 asing limitations when reporting subjective outcomes could be factors contributing  
389 to their limited selection, or under-reporting as scientific methods. Indeed, with our  
390 survey we could identify an apparently unpublished, but nonetheless frequent use of  
391 user-oriented evaluation. Attributes such as ease of use, safety and comfort appear  
392 to be often evaluated, indicating a frequent practice of qualitative user evaluation.  
393 Such user-oriented evaluation from early on is expected to enhance stakeholder in-  
394 volvement and may help meeting user needs right from the start [35, 39, 40]. We  
395 can conclude that while effectiveness may remain a core focus of development and  
396 evaluation of WRD, at least 50% of all efforts should be directed towards optimizing  
397 efficiency and satisfaction from early on. The importance of functional performance  
398 and device effectiveness is indubitable, but at the end of the day, it is the overall  
399 user experience that decides whether a device will be used in daily life or not.

#### 400 Usability Measures and Methods

401 In our detailed analysis of specific measures used in WRD usability evaluation, the  
402 surprisingly frequent use of user-oriented methods was further revealed. All eight  
403 types of usability evaluation methods as listed by ISO TR 16982-2002 were reported  
404 to be used [27]. Most projects report a combination between objective and objec-  
405 tive evaluation, with methods such as questionnaires, interviews, and observations  
406 complementing PRM as most popular methods. Simple and fast quantitative us-  
407 ability data, mostly collected with PRM such as task success & failure, appear to  
408 be preferred in the WRD field, which is in line with the effectiveness-driven focus  
409 discussed above. Standardized functional measures such as clinical-grade function  
410 tests were the least reported PRM, mainly because context-specific, standardized

411 measures are hard to find and less preferred than customized ones. A similar trend  
412 was observed for the use of questionnaires and scales, where we found that 73% of all  
413 WRD projects use custom-made measures, often in combination with standardized  
414 ones such as the SUS.

415 These findings compare well with recent works investigating robotic assistive tech-  
416 nology evaluation, which showed that custom-made usability measures are used in  
417 60-70% of all reported studies [20, 21]. Also, we could extend the findings from  
418 Koumpouros (2016), who previously listed the SUS, QUEST 2.0, and NASA TLX as  
419 the most frequently used standardized questionnaires to subjectively measure WRD  
420 usability [20]. The general preference for easy-to-administer, objective PRM such  
421 as time for task or task success & failure was also observed by Pinto-Fernandez et  
422 al. [22]. Their review focusing on lower limb exoskeleton evaluation also highlighted  
423 the popularity of kinetic and kinematic measures, which was further confirmed by a  
424 recent review on soft wearable robotics by Xiloyannis et al. [3]. Furthermore, phys-  
425 iological measures such as muscle activation, heart rate, or metabolic consumption  
426 are frequently reported measures of WRD usability, especially among the uprising  
427 soft robotic technologies [3].

428 Although certain popular measures are validated and generalizable, an overall  
429 preference towards custom tools is a clear limiting factor in the current usability  
430 evaluation practice of WRD. This appears to be a chicken-and-egg problem: WRD  
431 developers can't find validated and standardized tools for their specific context of  
432 use and thus start creating their own, customized measures. At the same time, the  
433 usage of such non-standardized and non-validated metrics inhibits the emergence  
434 of evaluation standards and benchmarks. Another factor that fuels this problem is  
435 the generally scarce availability of WRD usability studies. As highlighted in a 15  
436 years reflection on the papers published in the Journal of Neuroengineering and  
437 Rehabilitation, only 2.4% of the works published included the term "usability" [2].

438 Of those studies, only very few report qualitative data, which is further limiting the  
439 body of literature on this topic. What the WRD field therefore might need, are not  
440 only guidelines and benchmarks for usability evaluation practice, but an agreement  
441 among peers to also value qualitative research as scientific practice worth publishing.

#### 442 Limitations and Implications

443 The insights and conclusions generated from this survey data should be taken with  
444 care, as a sample size of 158 responses may not fully capture the diversity of the  
445 WRD field. Also, the understanding and interpretation of characteristics such as  
446 TRL or number of users tested with, might be hard to estimate. Despite our efforts  
447 to distribute the survey on interdisciplinary, global platforms, the outreach and  
448 thus data sample might have been influenced by the authors' network. It can thus  
449 be argued that specific context of use, or usability evaluation practices might be  
450 somewhat underrepresented in this study. Nevertheless, our survey data allowed a  
451 novel understanding and comprehensive review of WRD applications, their develop-  
452 ment stages and current usability evaluation practices. We could further elucidate  
453 the need for evaluation benchmarks and development guidelines, as it has been de-  
454 tailed and called for in recent works [19, 25]. Our analysis may even support the  
455 benchmarking endeavors of initiatives such as the EUROBENCH project [41], the  
456 Exo Technology Center of Excellence from ASTM International [42] or the CY-  
457 BATHLON [43].

#### 458 Evaluation Recommendations

459 From the insights generated through this survey, we can propose recommendations  
460 on how to organize and set up evaluation protocols to support the UCD of wearable  
461 robots:

- 462 1 We recommend to evenly allocate the evaluation efforts between the usability  
463 dimensions effectiveness, efficiency, and satisfaction. This implies that one  
464 should start including target users in the development as soon as possible, also

465 the early stages of WRD development (TRL 1 - 3). A simple walk-through,  
466 interviews, or focus groups with target users can shape your concept or low-  
467 fidelity prototype in the right direction from the start.

468 2 A well-balanced evaluation protocol should include both, quantitative and  
469 qualitative measures. While quantitative data allows for comparison to previ-  
470 ous design iterations and the state of the art, qualitative data helps identifying  
471 usability issues that may not be obvious to WRD developers when only collect-  
472 ing and comparing numbers. Also, qualitative evaluation by nature calls for  
473 more personal interaction with the target users and their satisfaction with the  
474 device, thus further promoting UCD and improving technology acceptance.

475 3 The often necessary usage of custom-made or modified measures to investigate  
476 specific research questions should optimally be complemented with standard-  
477 ized tools to allow a fair outside comparison and to limit evaluation bias.

478 Our recommendation is to compose a WRD usability evaluation protocol in  
479 a 2-to-1 fashion, meaning that a protocol consisting of two standardized, or  
480 well-known and validated measures can be enhanced with one custom-made,  
481 internal measure. This may likely increase result comparability as well as data  
482 credibility and may help the field to establish evaluation benchmarks.

483 These recommendations, together with the detailed breakdown of methods and  
484 measures used, may help WRD developers to define their evaluation protocols and  
485 to understand which measurement tools could fit their own context of use. This  
486 survey data will further be used for the development and data-driven training of  
487 a usability evaluation toolbox for WRD [44], which will allow the users to simply  
488 enter their context of use to find a list of validated and relevant measures together  
489 with our evaluation guidelines.

## 490 **Conclusions**

491 This study provided insights into the current evaluation practices and specified  
492 the contexts of use of wearable robots for human augmentation, assistance and/or  
493 therapy. Evaluation protocols across the various applications are similar, but the  
494 lack of applicable guidelines restricts the validation and benchmarking of usability  
495 standards in the field of wearable robotics. The individual adaption of evaluation  
496 protocols to specific contexts of use remains a fundamental barrier and challenge  
497 to get more reproducible evaluation data to push forward user-centered designs of  
498 wearable robotic devices. The insights generated by this survey might serve as a  
499 data-driven basis for evaluation protocol recommendations and help researchers in  
500 defining or comparing their evaluation protocols and data. We believe that more  
501 structured and comparable usability evaluations can tackle the technology accep-  
502 tance limitations of WRD and can help making first steps towards WRD evaluation  
503 benchmarking to eventually fill the translational gap of wearable robotics.

## 504 **Abbreviations**

505 WRD: Wearable Robotic Devices, UCD: User-centered Design, TRL: Technology Readiness Level, TRP: Technology  
506 Readiness Phase, PRM: Performance-related Measures, SFM: Standardized Functional Measure, LS: Likert Scale,  
507 VAS: Visual Analogue Scale, SUS: System Usability Scale, NRS: Numeric Rating Scale, NASA TLX: NASA Task  
508 Load Index, QUEST 2.0: Quebec User Evaluation of Satisfaction with Assistive Technology, mSUS: modified System  
509 Usability Scale, RTLX: Raw Task Load Index, ASQ: After Scenario Questionnaire, PUEU: Perceived Usefulness,  
510 Perceived Ease of Use, ATD PA: Assistive Technology Device Predisposition Assessment, PSSUQ: Post-Study  
511 Usability Questionnaire, COPM: Canadian Occupational Performance Measure, USAT: Usability Scale for Assistive  
512 Technology, IMI: Intrinsic Motivation Inventory, USEQ: Usefulness, Satisfaction, Ease of Use Questionnaire,  
513 PYTHEIA: Psychometric Scale to Assess the Satisfaction of Users with Assistive Technology, QUIS: The  
514 Questionnaire for User interaction Satisfaction, MHQ: Michigan Hand Outcomes Questionnaire, PEQ: Prosthesis  
515 Evaluation Questionnaire, PIADS: Psychosocial Impact of Assistive Device, QEFTH: Questionnaire to Explore  
516 Human Factors and their Technical potential, TSQ-WT: Telehealthcare Satisfaction Questionnaire – Wearable  
517 Technology, PIADS: The Psychosocial Impact of Assistive Devices Scale, TAPES: Trinity Amputation and  
518 Prosthesis Experience Scale, UMUX: Usability Metric for User Experience;

## 519 **Ethics approval and consent to participate**

520 The respondents were presented with written informed consent on the landing page of the online survey. By starting  
521 the survey, the respondents agreed to all terms and conditions.

## 522 **Consent for publication**

523 Informed consent for publication was obtained from the respondents before starting the survey.

**524 Availability of data and materials**

525 Data and materials can be made available upon reasonable request to the authors.

**526 Competing interests**

527 The authors declare that they have no competing interests.

**528 Author's contributions**

529 JTM, RG and OL designed the survey. JTM, RG and OL distributed and promoted the survey link and purpose.

530 JTM analyzed and interpreted the data. JTM, RG and OL wrote the manuscript. All authors provided critical

531 feedback on the manuscript. All authors read and approved the final manuscript.

**532 Funding**

533 This work was supported by Vontobel Foundation, the Swiss National Science Foundation through the National

534 Centre of Competence in Research on Robotics, the ETH Zurich Foundation in collaboration with Hocoma AG, and

535 the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and

536 Technological Enterprise (CREATE) program.

**537 Acknowledgements**

538 The authors would like to thank all respondents for taking the time to complete our survey. We also thank the Cost

539 Action 16116, the Exoskeleton Report, Springer Nature and David Reinkensmeyer for their invaluable support in

540 distributing and promoting the survey. Special thanks to Jessica Gantenbein, Giada Devittori, Jan Dittli and the

541 entire Rehabilitation Engineering Laboratory for their help with ideating survey questions and reviewing the

542 manuscript.

**543 Author details**

544 <sup>1</sup>Rehabilitation Engineering Laboratory, Department of Health Sciences and Technology, ETH Zurich, Switzerland.

545 <sup>2</sup>Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence And Technological Enterprise  
546 (CREATE), Singapore.

**547 References**

- 548 1. Bao, G., Pan, L., Fang, H., Wu, X., Yu, H., Cai, S., Yu, B., Wan, Y.: Academic Review and Perspectives on  
549 Robotic Exoskeletons. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**(11), 2294–2304  
550 (2019). doi:10.1109/TNSRE.2019.2944655
- 551 2. Reinkensmeyer, D.J.: JNER at 15 years: analysis of the state of neuroengineering and rehabilitation. *Journal of*  
552 *NeuroEngineering and Rehabilitation* **16**(1), 144 (2019). doi:10.1186/s12984-019-0610-0
- 553 3. Xiloyannis, M., Alicea, R., Georarakis, A.-M., Haufe, F.L., Wolf, P., Masia, L., Riener, R.: Soft robotic suits:  
554 State of the art, core technologies, and open challenges. *IEEE Transactions on Robotics*, 1–20 (2021).  
555 doi:10.1109/TRO.2021.3084466
- 556 4. Molteni, F., Gasperini, G., Cannaviello, G., Guanziroli, E.: Exoskeleton and End-Effector Robots for Upper and  
557 Lower Limbs Rehabilitation: Narrative Review. *PM&R* **10**, 174–188 (2018). doi:10.1016/j.pmrj.2018.06.005
- 558 5. McFarland, T., Fischer, S.: Considerations for Industrial Use: A Systematic Review of the Impact of Active and  
559 Passive Upper Limb Exoskeletons on Physical Exposures. *IIEE Transactions on Occupational Ergonomics and*  
560 *Human Factors* **7**(3–4), 322–347 (2019). doi:10.1080/24725838.2019.1684399
- 561 6. Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. Standard,  
562 International Organization for Standardization, Geneva, CH (March 2018)

- 563 7. Scherer, M.J., Sax, C., Vanbiervliet, A., Cushman, L.A., Scherer, J.V.: Predictors of assistive technology use:  
564 The importance of personal and psychosocial factors. *Disability and Rehabilitation* **27**(21), 1321–1331 (2005).  
565 doi:10.1080/09638280500164800
- 566 8. Biddiss, E.A., Chau, T.T.: Upper limb prosthesis use and abandonment: A survey of the last 25 years.  
567 *Prosthetics and Orthotics International* **31**(3), 236–257 (2007). doi:10.1080/03093640600994581
- 568 9. van Ommeren, A.L., Smulders, L.C., Prange-Lasonder, G.B., Buurke, J.H., Veltink, P.H., Rietman, J.S.:  
569 Assistive Technology for the Upper Extremities After Stroke: Systematic Review of Users' Needs. *JMIR*  
570 *Rehabilitation and Assistive Technologies* **5**(2), 10510 (2018). doi:10.2196/10510
- 571 10. Sugawara, A.T., Ramos, V.D., Alfieri, F.M., Battistella, L.R.: Abandonment of assistive products: assessing  
572 abandonment levels and factors that impact on it. *Disability and Rehabilitation: Assistive Technology* **13**(7),  
573 716–723 (2018). doi:10.1080/17483107.2018.1425748
- 574 11. Shah, S.G.S., Robinson, I.: Benefits of and barriers to involving users in medical device technology development  
575 and evaluation. *International Journal of Technology Assessment in Health Care* **23**(1), 131–137 (2007).  
576 doi:10.1017/S0266462307051677
- 577 12. Blanco, T., Berbegal, A., Blasco, R., Casas, R.: Xassess: crossdisciplinary framework in user-centred design of  
578 assistive products. *Journal of Engineering Design* **27**(9), 636–664 (2016). doi:10.1080/09544828.2016.1200717
- 579 13. Fischer, B., Peine, A., Östlund, B.: The Importance of User Involvement: A Systematic Review of Involving  
580 Older Users in Technology Design. *The Gerontologist* **60**(7), 513–523 (2020). doi:10.1093/geront/gnz163
- 581 14. Poulson, D., Richardson, S.: USERfit – a framework for user centred design in assistive technology. *Technology*  
582 *and Disability* **9**(3), 163–171 (1998). doi:10.3233/TAD-1998-9307
- 583 15. Shah, S.G.S., Robinson, I., AlShawi, S.: Developing medical device technologies from users' perspectives: A  
584 theoretical framework for involving users in the development process. *International Journal of Technology*  
585 *Assessment in Health Care* **25**(4), 514–521 (2009). doi:10.1017/S0266462309990328
- 586 16. Christ, O., Jokisch, M., Preller, J., Beckerle, P., Wojtusch, J., Rinderknecht, S., von Stryk, O., Vogt, J.:  
587 User-Centered Prosthetic Development: Comprehension of Amputees' Needs. *Biomedical Engineering /*  
588 *Biomedizinische Technik* **57**(SI-1 Track-R) (2012). doi:10.1515/bmt-2012-4306
- 589 17. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. Standard,  
590 International Organization for Standardization, Geneva, CH (July 2019)
- 591 18. Kapeller, A., Felzmann, H., Fosch-Villaronga, E., Hughes, A.-M.: A Taxonomy of Ethical, Legal and Social  
592 Implications of Wearable Robots: An Expert Perspective. *Science and Engineering Ethics* **26**(6), 3229–3247  
593 (2020)
- 594 19. Torricelli, D., Rodríguez-Guerrero, C., Veneman, J.F., Crea, S., Briem, K., Lenggenhager, B., Beckerle, P.:  
595 Benchmarking Wearable Robots: Challenges and Recommendations From Functional, User Experience, and  
596 Methodological Perspectives. *Frontiers in Robotics and AI* **7** (2020). doi:10.3389/frobt.2020.561774
- 597 20. Koumpouros, Y.: A Systematic Review on Existing Measures for the Subjective Assessment of Rehabilitation  
598 and Assistive Robot Devices. *Journal of Healthcare Engineering* **2016**, 1–10 (2016). doi:10.1155/2016/1048964
- 599 21. Ármannsdóttir, A.L., Beckerle, P., Moreno, J.C., van Asseldonk, E.H.F., Manrique-Sancho, M.-T., Del-Ama,  
600 A.J., Veneman, J.F., Briem, K.: Assessing the Involvement of Users During Development of Lower Limb  
601 Wearable Robotic Exoskeletons: A Survey Study. *Human Factors: The Journal of the Human Factors and*  
602 *Ergonomics Society* **62**(3), 351–364 (2020). doi:10.1177/0018720819883500
- 603 22. Pinto-Fernandez, D., Torricelli, D., Sanchez-Villamanan, M.D.C., Aller, F., Mombaur, K., Conti, R., Vitiello,  
604 N., Moreno, J.C., Pons, J.L.: Performance Evaluation of Lower Limb Exoskeletons: A Systematic Review. *IEEE*  
605 *Transactions on Neural Systems and Rehabilitation Engineering* **28**(7), 1573–1583 (2020).

- 606 doi:10.1109/TNSRE.2020.2989481
- 607 23. Bryce, T.N., Dijkers, M.P., Kozlowski, A.J.: Framework for Assessment of the Usability of Lower-Extremity  
608 Robotic Exoskeletal Orthoses. *American Journal of Physical Medicine & Rehabilitation* **94**(11), 1000–1014  
609 (2015). doi:10.1097/PHM.0000000000000321
- 610 24. Shore, L., Power, V., Hartigan, B., Schülein, S., Graf, E., de Eyto, A., O’Sullivan, L.: Exoscore: A Design Tool  
611 to Evaluate Factors Associated With Technology Acceptance of Soft Lower Limb Exosuits by Older Adults.  
612 *Human Factors: The Journal of the Human Factors and Ergonomics Society* **62**(3), 391–410 (2020).  
613 doi:10.1177/0018720819868122
- 614 25. Lenker, J.A., Koester, H.H., Smith, R.O.: Toward a national system of assistive technology outcomes  
615 measurement. *Assistive Technology* (2019). doi:10.1080/10400435.2019.1567620
- 616 26. Horizon 2020 Work Programme(s) General Annex G “Technology readiness levels (TRL)”. Standard, European  
617 Commission, Brussels, Belgium (April 2014)
- 618 27. Ergonomics of human-system interaction — Usability methods supporting human-centred design. Standard,  
619 International Organization for Standardization, Geneva, CH (June 2002)
- 620 28. Meyer, J.T., Lambercy, O., Gassert, R.: Survey: How do you evaluate the usability of wearable robots?  
621 Exoskeleton Report (2020).  
622 <https://exoskeletonreport.com/2020/07/survey-how-do-you-evaluate-the-usability-of-wearable-robots/>  
623 Accessed 2021-06-06
- 624 29. Meyer, J.T., Lambercy, O., Gassert, R.: Can a systematic usability evaluation help us develop better wearable  
625 robots? *Biomed Central* (2020). [https://blogs.biomedcentral.com/on-medicine/2020/07/30/can-a-systematic-](https://blogs.biomedcentral.com/on-medicine/2020/07/30/can-a-systematic-usability-evaluation-help-us-develop-better-wearable-robots/)  
626 [usability-evaluation-help-us-develop-better-wearable-robots/](https://blogs.biomedcentral.com/on-medicine/2020/07/30/can-a-systematic-usability-evaluation-help-us-develop-better-wearable-robots/) Accessed 2021-06-06
- 627 30. Pons, J.L.: *Wearable robots: biomechatronic exoskeletons*. John Wiley & Sons (2008)
- 628 31. Colombo, G., Jorg, M., Dietz, V.: Driven gait orthosis to do locomotor training of paraplegic patients. In:  
629 *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of*  
630 *the IEEE*, vol. 4, pp. 3159–3163 (2000). IEEE
- 631 32. Noronha, B., Accoto, D.: Exoskeletal Devices for Hand Assistance and Rehabilitation: A Comprehensive  
632 Analysis of State-of-the-Art Technologies. *IEEE Transactions on Medical Robotics and Bionics* (2021).  
633 doi:10.1109/tmr.2021.3064412
- 634 33. Tolikas, M., Antoniou, A., Ingber, D.E.: The wyss institute: A new model for medical technology innovation  
635 and translation across the academic-industrial interface. *Bioengineering & translational medicine* **2**(3), 247–257  
636 (2017)
- 637 34. Hall, A.K., Backonja, U., Painter, I., Cakmak, M., Sung, M., Lau, T., Thompson, H.J., Demiris, G.:  
638 Acceptance and perceived usefulness of robots to assist with activities of daily living and healthcare tasks.  
639 *Assistive Technology* (2019). doi:10.1080/10400435.2017.1396565
- 640 35. Wolff, J., Parker, C., Borisoff, J., Mortenson, W., Mattie, J.: A survey of stakeholder perspectives on  
641 exoskeleton technology. *Journal of NeuroEngineering and Rehabilitation* **11**(1), 169 (2014).  
642 doi:10.1186/1743-0003-11-169
- 643 36. WIPO Technology Trends 2021- Assistive Technology. Annual report, World Intellectual Property Organization,  
644 Geneva, Switzerland (2021)
- 645 37. Celian, C., Swanson, V., Shah, M., Newman, C., Fowler-King, B., Gallik, S., Reilly, K., Reinkensmeyer, D.J.,  
646 Patton, J., Rafferty, M.R.: Real-time decision-making on whether to incorporate neurorehabilitation technology:  
647 Informing developers on a day in the life of a therapist. *Journal of NeuroEngineering and Rehabilitation* **18**(121)  
648 (2021)

- 649 38. Contreras-Vidal, J.L., A Bhagat, N., Brantley, J., Cruz-Garza, J.G., He, Y., Manley, Q., Nakagome, S., Nathan,  
650 K., Tan, S.H., Zhu, F., Pons, J.L.: Powered exoskeletons for bipedal locomotion after spinal cord injury. *Journal*  
651 *of Neural Engineering* **13**(3) (2016). doi:10.1088/1741-2560/13/3/031001
- 652 39. Biddiss, E., Beaton, D., Chau, T.: Consumer design priorities for upper limb prosthetics. *Disability and*  
653 *Rehabilitation: Assistive Technology* **2**(6), 346–357 (2007). doi:10.1080/17483100701714733
- 654 40. van Dijksseldonk, R.B., van Nes, I.J.W., Geurts, A.C.H., Keijsers, N.L.W.: Exoskeleton home and community use  
655 in people with complete spinal cord injury. *Scientific Reports* (2020). doi:10.1038/s41598-020-72397-6
- 656 41. Torricelli, D., Pons, J.L.: EUROBENCH: Preparing Robots for the Real World. In: *Biosystems and Biorobotics*,  
657 pp. 375–378 (2019)
- 658 42. ASTM International: Exo Technology Center of Excellence (2021). <https://www.etcoc.org/> Accessed  
659 2021-06-06
- 660 43. CYBATHLON ETH Zurich: Official website of the CYBATHLON (2021). <https://cybathlon.ethz.ch/en>  
661 Accessed 2021-06-06
- 662 44. Meyer, J.T., Lamercy, O., Gassert, R.: Proceedings Part 11: User Participation - Enabling User-centered  
663 Design and Evaluation to Increase Acceptance of Wearable Robotic Assistive Technologies. *Technology and*  
664 *Disability* **31**(s1), 189–193 (2019). doi:10.3233/TAD-190014

#### 665 **Figures**

666 attached in separate file as PDFs.

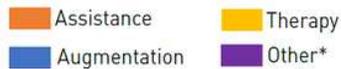
#### 667 **Additional Files**

668 Additional file 1 - Print of of full survey

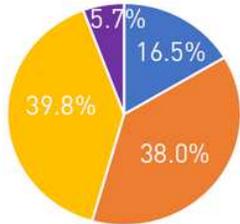
669 Additional file 2 - Extension of table 2 with all usability attributes

# Figures

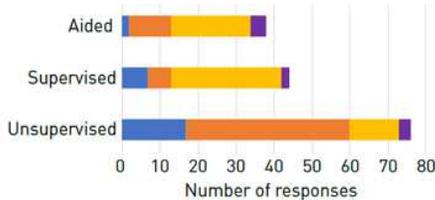
## LEGEND



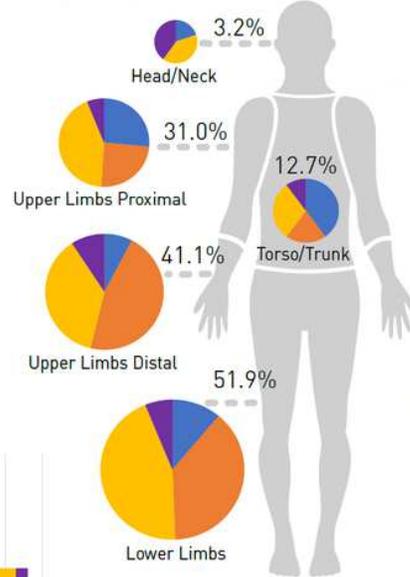
### A) General Usage Purpose (n = 158)



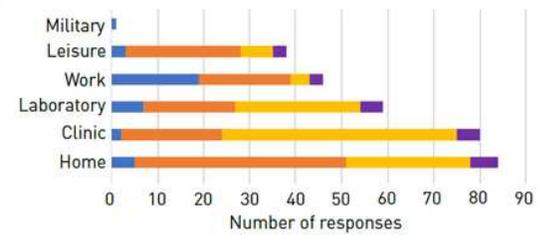
### B) Supervision Form (n = 158)



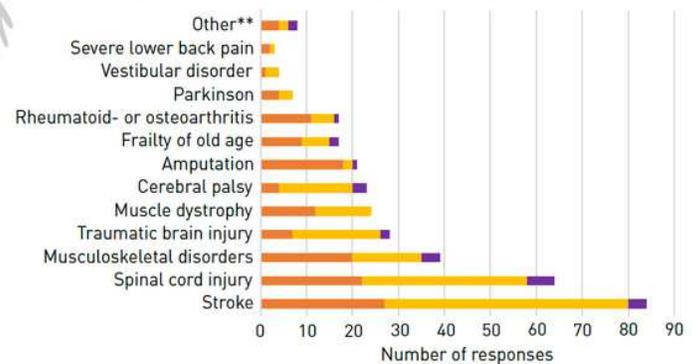
### C) Target Body Area (n = 158)



### D) Usage Environment (n = 158)

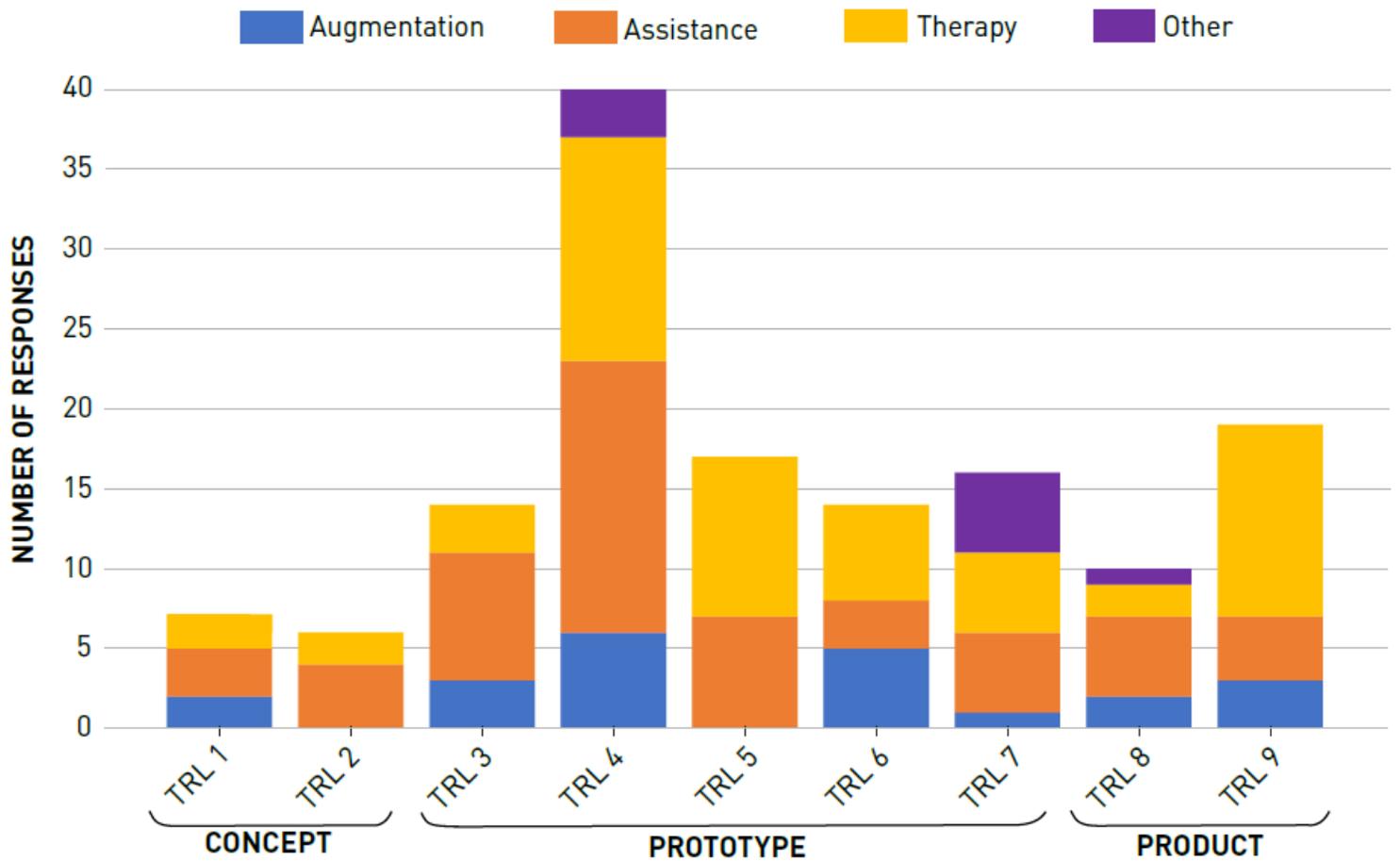


### E) Target Population of Medical Applications (n = 132)



**Figure 1**

Context of Use of Wearable Robotic Devices: A) General usage purposes: augmentation, assistance, therapy and other. \*Other usage purposes reported were: all of the above (n = 3), brain computer interfaces (n = 2), fitness and sports tracking (n = 1), training and assistance for surgery (n = 1), benchmarking (n = 1) and user research (n = 1), B) supervision form: We can differentiate between fully supervised, partially-supervised (aided) and unsupervised use of WRD, C) Target body areas: The relative frequencies of the four general usage purposes are reported for each body area D) Usage environment: The intended use ranges from rather controlled (laboratory, clinic, home) to more dynamic (leisure, military, work) environments E) Target population of medical applications: For the medical applications (therapy and assistance) respondents reported the specific target groups. \*\*Other target groups where as: visually impaired, essential tremor, trauma, rhabdomyolysis, peripheral artery disease, first responders (emergency), physical training and exercise, and sports injuries.



**Figure 2**

Technology Maturity of Current Wearable Robotic Devices: The Technology Readiness Levels (TRL) of the four general usage purposes, visualized as stacked histogram (n = 143). The TRL were grouped in three Technology Readiness Phases (TRP): "Concept" (TRL 1 - 3), "Prototype" (TRL 4 - 7) and "Product" (TRL 8 - 9).

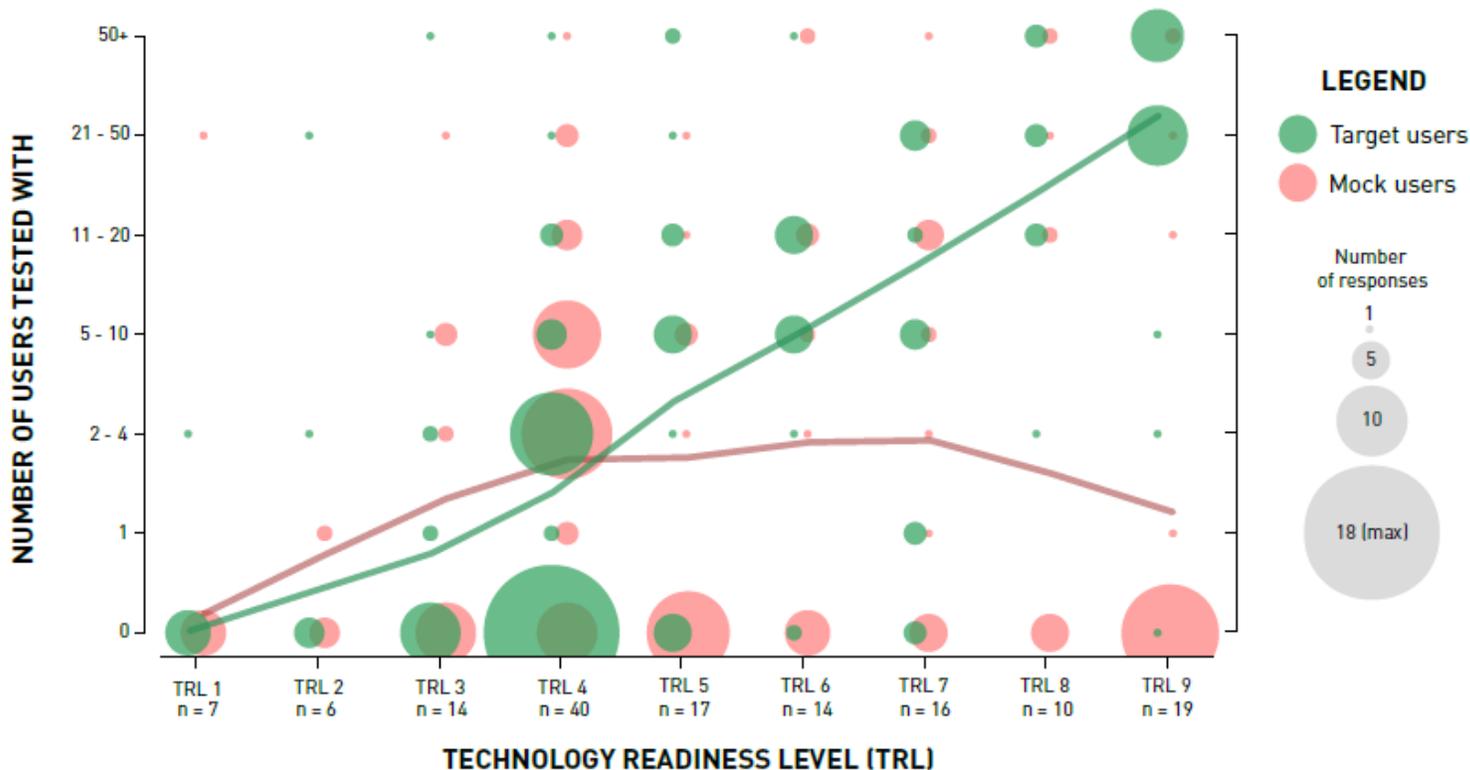


Figure 3

Progression of User Involvement with Technology Maturation The number of users (target and mock) per Technology Readiness Level (TRL) are shown (n = 143). The size of the circles represents the number of responses within each TRL. For all respondents that only indicated testing with the one specific user group (target or mock), 0 users are shown in the opposite group. The Loess regression lines were fitted to smooth the scatter plot visualization.

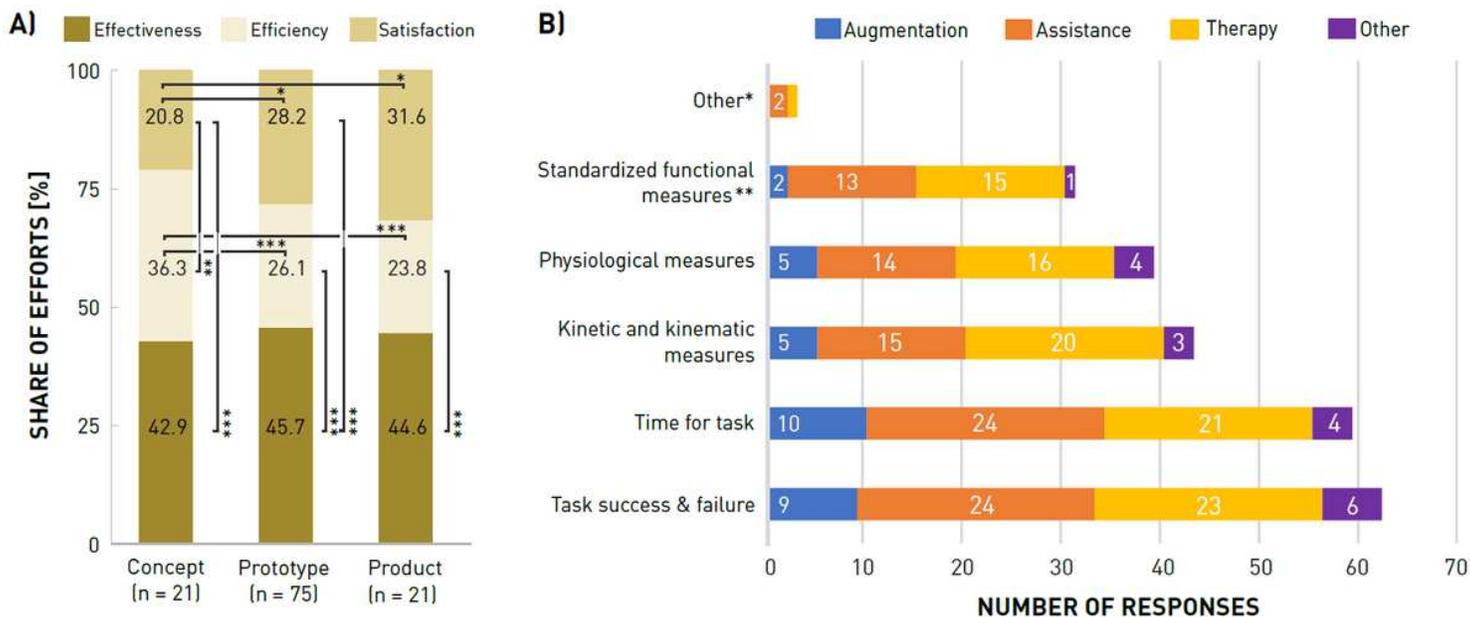
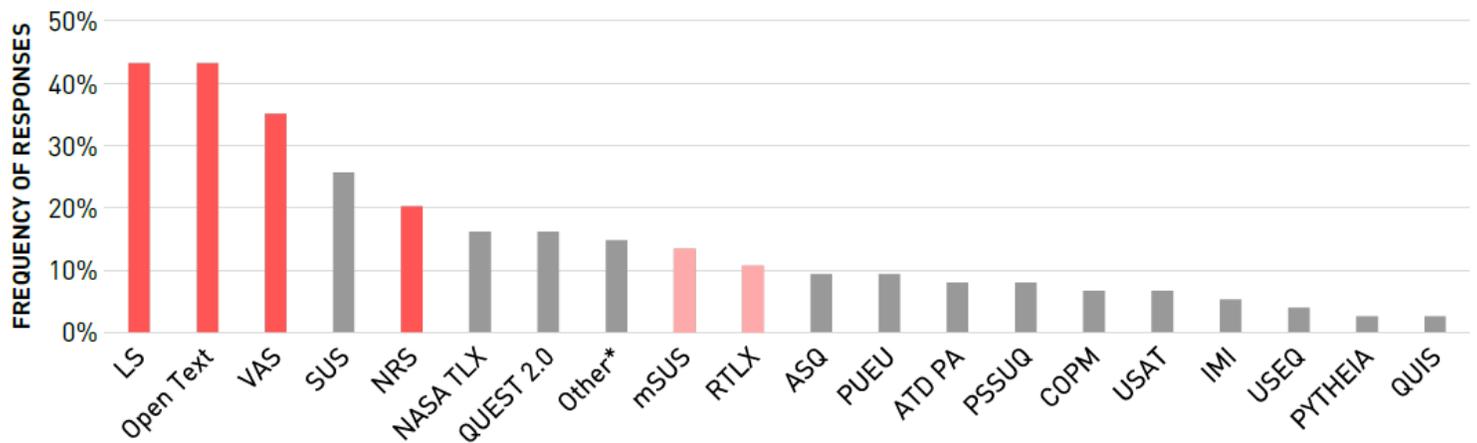


Figure 4

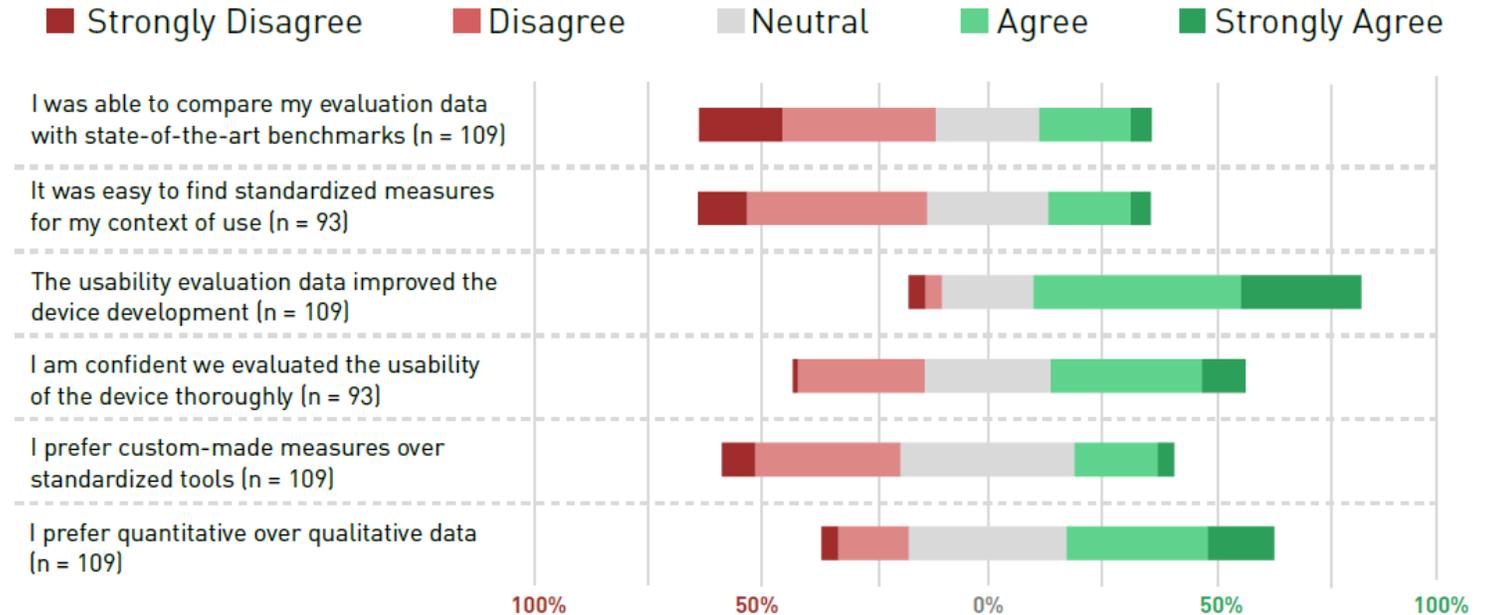
Analysis of Usability Evaluation Practice: A) Evaluation efforts allocation per Technology Readiness Phase (TRP, n = 117): The allocated total of 100 evaluation-effort-points among the usability dimensions effectiveness, efficiency, and satisfaction, per TRP are shown. Paired comparison between the three dimension were analyzed within each TRP, while unpaired, two-sample comparisons between the TRP were calculated. Levels of significance indicated as: \* = p <0.05, \*\* = p <0.01, \*\*\* = p <0.001. B) List of reported performance-related measurements (PRM, n = 88): \*Other PRM described as spatiotemporal metric analysis, eye-tracking recording, number of steps. \*\*Standardized functional measures (SFM) specified: upper limb SFM; Box and Block Test (BBT, n = 6), Jebsen-Taylor Hand Function Test (JTHFT, n = 5), Action Research Arm Test (ARAT, n = 4), Chedoke Arm and Hand Activity Inventory (CAHAI, n = 2), Southampton Hand Assessment Procedure (SHAP), Assisting Hand Assessment (AHA), Smart Pegboard, Frenchay Arm Test (FAT), lower limb SFM; 10 Meter Walk Test (10MWT, n = 2), 2 Minute Walk Test (2MWT, n = 2), 6 Minute Walk Test (6MWT, n = 2), Timed-up- and-Go (TUG), general SFM; Human-Robot Fluency Metrics, Assessment of Capacity for Myoelectric Control (ACMC), Thermography, Failure Mode and Effects Analysis (FMEA), ISO regulation, t and tolerance assemblies



**Figure 5**

Questionnaires and Scales Used in Wearable Robotics Usability Evaluation: The 20 most frequently selected questionnaires and scales are displayed (n = 74). Custom-made forms are marked in dark red, modified questionnaires are marked in light red. Abbreviations: LS = Likert Scale, VAS = Visual Analogue Scale, SUS = System Usability Scale, NRS = Numeric Rating Scale, NASA TLX = NASA Task Load Index, QUEST 2.0 = Quebec User Evaluation of Satisfaction with Assistive Technology, mSUS = modified System Usability Scale, RTLX = Raw Task Load Index, ASQ = After Scenario Questionnaire, PUEU = Perceived Usefulness, Perceived Ease of Use, ATD PA = Assistive Technology Device Predisposition Assessment, PSSUQ = Post-Study Usability Questionnaire, COPM = Canadian Occupational Performance Measure, USAT = Usability Scale for Assistive Technology, IMI = Intrinsic Motivation Inventory, USEQ = Usefulness, Satisfaction, Ease of Use Questionnaire, PYTHEIA = Psychometric Scale to Assess the Satisfaction of Users with Assistive Technology, QUIS = The Questionnaire for User interaction Satisfaction. \*Other = Borg Scale of Perceived Exertion, Michigan Hand Outcomes Questionnaire, Prosthesis Evaluation Questionnaire, Psychosocial Impact of Assistive Device, Questionnaire to Explore Human Factors and

their Technical potential, Quick- DASH, SF-36, Telehealthcare Satisfaction Questionnaire { Wearable Technology, Trinity Amputation and Prosthesis Experience Scale, Usability Metric for User Experience, and Embodiment Questionnaire.



**Figure 6**

Reflection of Usability Evaluation Practise: The level of agreement for each statement is shown. Disagreement is red, agreement is green. The more the entire bar (100%) is shifted to the left or right, the clearer a trend of agree- or disagreement is observable among the responses.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalMaterialsMeyerJNER2021.pdf](#)