applied sciences

MDPI

Article

How Familiarization Impacts Construction Workers' Acceptance of an Active Exoskeleton: Insights from a Case Study

Vasco Fanti, Sergio Leggieri, Darwin G. Caldwell and Christian Di Natali *

Department of Advanced Robotics (ADVR), Istituto Italiano di Tecnologia (IIT), Center for Robotics and Intelligent Systems (CRIS), Via San Quirico, 19D, 16163 Genova, GE, Italy; vasco.fanti@iit.it (V.F.); sergio.leggieri@iit.it (S.L.); darwin.caldwell@iit.it (D.G.C.)
* Correspondence: christian.dinatali@iit.it

Abstract

To promote the adoption of exoskeletons in industries, they must be effective in providing physical assistance; however, other factors, such as familiarity with their use and comfort, contribute to device acceptance and workers' approval. During on-site testing, the use of common assessment tools (e.g., EMGs, MoCap systems, etc.) and the pool of available subjects are usually limited. Thus, evaluating the reliability of qualitative results (e.g., from questionnaires) is critical. This paper investigates the impact of introducing a back-support exoskeleton to five construction workers in operative scenarios. Key factors, including perceived physical exertion, the task workload, and overall exoskeleton acceptance, were evaluated. The results were compared before and after the workers were instructed on the proposed familiarization approach. This analysis aims to investigate the potential influence of novelty bias by assessing the consistency of users' feedback, which shows increased agreement among workers after familiarization. The relative dispersion of responses is assessed using the Coefficient of Variation (CoV) index. The evidence suggests a potential bias and overestimation in the results obtained during the first test day (before familiarization). After familiarization, the results indicate incremental reliability for the selected questionnaires. Physical exertion perceived in the trunk (before and after familiarization) is further reduced by 0.5 out of 5, while acceptance of the device increases from 5.7 to 6.3 out of 7. The proposed method can support the adoption of industrial exoskeletons by providing a tool for assessing the reliability of qualitative data.

Keywords: occupational exoskeletons; device familiarization; acceptance improvement; robotic assistance; occupational disease prevention

Check for updates

Academic Editors: Lun Lou, Zhanxiao Kang, Qiuqiong Shi and Yadie Yang

Received: 28 September 2026
Revised: 21 October 2026
Accepted: 26 October 2026
Published: 29 October 2026

Copyright: © 2026 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.

1. Introduction

In recent years, the growing interest in wearable robotics has led to an exponential growth in publications on exoskeletons [1–3]. These exoskeletons can be grouped according to their physical characteristics, such as structure (rigid or soft) and actuation technology (active or passive), as well as their purpose, including physical motion recovery, performance enhancement, or injury prevention [4]. Exoskeletons used by industry are called Occupational Exoskeletons (OEs) and they are intended to help workers reduce their physical workload during the execution of work tasks [5]. This is because prolonged exposure of the body to strenuous workloads is associated with an increased risk of developing occupational diseases such as musculoskeletal disorders (MSD) [6,7]. OEs are seen as a technology to reduce these occurrences. Work-related MSD mainly affect the back and

upper limbs (with an incidence of 43% and 41%, respectively, in Europe [8]) and are the leading cause of absences due to illness in countries such as Europe, the United States, and Canada [9,10].

The development of new OEs to assist the back and upper extremities, namely, back-support exoskeletons (BSE) and upper-limb exoskeletons (ULE), respectively, leads to the need for field testing to truly evaluate their effect in the workplace [11]. Testing in a real work environment enables the identification and monitoring of potential benefits or side effects associated with exoskeleton use in real-life work situations [12]. Often, these effects are not visible or clearly identifiable in the laboratory [13]. Since performing field tests with the equipment typically used in the laboratory (e.g., EMG [14,15] and motion capture systems [16–18]) is challenging, surveys are the most widely used evaluation method for on-site testing [11]. Moreover, according to [12], surveys are also the method that permits investigation of the greatest number of dimensions (i.e., human, machine, and task) and research objectives (i.e., acceptability, usability, job performance, workload, etc.).

In recent years, many research papers have focused on the acceptance and usefulness of OEs as perceived by workers [19–23]. Kim et al. [24] investigated the acceptance and usability of a passive ULE that was tested by workers in the automotive sector over an 18-month period. The results showed positive effects on workers' performance, with reduced strain and fatigue in the assisted body regions. Comfort proved to be a crucial factor in device acceptance [25]. This study evaluated exoskeleton performance over long periods (i.e., feedback was collected at 1, 6, 12, and 18 months) and short-term effects were neglected. Although long-term tests are crucial to evaluate the acceptability of exoskeletons, research activities often have stringent time constraints, usually limited to one or a few weeks and a small pool of workers [11].

In a short time and with few subjects, both positive and negative physical and psychological aspects (such as expectations, prejudice, stigma, or frustration) can affect data quality [26,27], leading to hardly reliable or biased results [28]. Thus, assessing the reliability of these outcomes becomes critical. Statistical indices for calculating the reliability of results obtained from questionnaires, such as Cronbach's alpha, require populations of at least 20 subjects [29]. The Coefficient of Variation (CoV) [30], on the other hand, can be applied to small populations and used in conjunction with thresholds to assess the reliability of results [31].

In previous studies, Sposito et al. [32] report an example of fluctuating behaviors in workers' perception and acceptance of the exoskeleton during a three-day experimental field campaign with an active BSE. Specifically, on the first day of testing, overall exoskeleton evaluation scores and impact are high. In contrast, on the second day, the scores worsen and then readjust on the third day of testing. The results suggested that the oscillation of qualitative results was determined by a combination of both physical and psychological aspects that evolve and change while practicing with the device. In fact, qualitative metrics are subjective and, given the same stimuli, the result is influenced by the attitude of individuals that is affected by emotions [33] and different learning methods [34]. Moreover, the performance of repetitive tasks, especially with an exoskeleton that could alter the worker's biomechanics, leads to muscle fatigue [35], which could affect acceptability and comfort [32].

To improve exoskeleton acceptance, increase their beneficial effects, and accelerate users' adaptation to new motor patterns [36], familiarization periods have been introduced to train novice users during exoskeleton testing [37]. Familiarization periods are designed to educate users on the proper use of the exoskeleton, providing them with tips to maximize assistance, practicing different control strategies, and personalizing their use of the device [38]. Although the benefits of the familiarization period are clear and demonstrated

by metrics such as reduced metabolic cost and muscle activities [39–41], its implementation has not yet been standardized. Therefore, predefined circuits, video explanations, or verbal briefings are used according to the final user's needs, and the duration of familiarization can range from minutes to hours or even months, depending on the task and the user's learning curve [37]. Furthermore, familiarization is device-specific, as its duration and effectiveness depend on the exoskeleton's mechanical characteristics, control strategies, and interaction with the user.

In this work, an active BSE, StreamEXO, was tested for one week in field trials with building construction workers. We proposed a familiarization approach and assessed its benefits by evaluating incremental consent on workers' opinions. Feedback was collected through questionnaires to evaluate physical exertion, the task workload, and technology acceptance before and after the familiarization period. Given the small sample size, as is common in relevant environments and operational assessment [11], and because our objective was not to assess the psychometric properties of the questionnaire but rather the degree of agreement among participants' subjective evaluations, we quantified the relative dispersion of responses using the Coefficient of Variation (CoV). The CoV provides an indication of the level of consensus among users, whereas psychometric reliability measures such as Cronbach's alpha evaluate the internal consistency of questionnaire items and were therefore outside the scope of this study. The objectives of this study are to (i) assess how a familiarization period impacts the workers; (ii) demonstrate a method for assessing the reliability of qualitative data for a small experimental data set. The first objective is investigated through surveys that evaluate changes in workers' perceived exertion, perceived workload, and acceptance of the device. For the second objective, we introduce and evaluate the application of the CoV, assuming that shared opinions are more reliable after familiarization.

The impact that the familiarization process can generate on reliability of the qualitative data acquired during trials is paramount for the community. This provides researchers with an additional tool to systematically evaluate and compare users' opinions on different exoskeletons, potentially paving the way for informed adoption of exoskeletons across industries. Additionally, the results of this study represent an important step to provide insight from qualitative data collected in worksites and with small subject pools.

2. Methods

2.1. Exoskeleton

The exoskeleton used in this series of experiments is the StreamEXO back-support exoskeleton, Figure 1a. StreamEXO is an active exoskeleton developed to assist construction workers in building, maintenance, and inspection activities [42]. The frame and wearable components are designed to maximize workers' comfort and mobility, and its control strategies are specifically tailored to assist in lifting/lowering and carrying tasks during heavy-duty activities [43]. The control settings were tuned based on the task performed, the load lifted (10 kg and 25 kg), and the subjects' average anthropometry. Max torque was set to 30 Nm per motor during lifting and lowering, while during carrying, the maximum torque during stance was 16 Nm. Since the torque generated by the exoskeleton's control is a function of the task and the subject's dynamics, there is some variability in the torque delivered. Figure 2 shows, for example, the trends in right- and left-motor activity for workers 3 and 5 during lifting and carrying tasks. The first part of the signal (when the blue and red actuator trends are symmetric and opposite) represents a lift task, then the carrying task starts until the end of the signal portion. It is worth noting that the signals from the right and left actuators are opposite because the motors' spatial orientation points toward each other. From the figure, it is evident that in Worker 3, the maximum absolute

force of both motors was about 24 Nm, whereas in Worker 5, it even reached 28 Nm. In both cases, during carrying, the forces spanned from 12 Nm to 16 Nm.

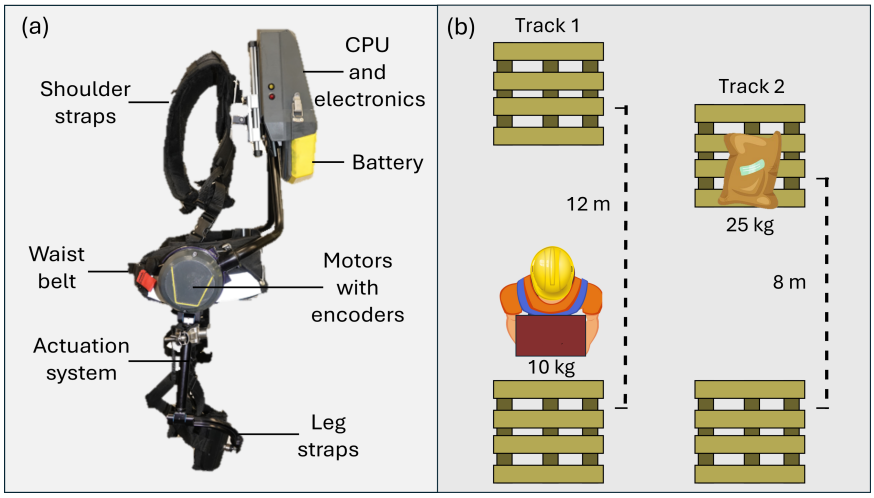


Figure 1. (**a**) The active back-support exoskeleton StreamEXO and its components. (**b**) Overview of the experimental task performed at the construction site. On the left is the track used with the stack of bricks, on the right is the track used with the cement bag.

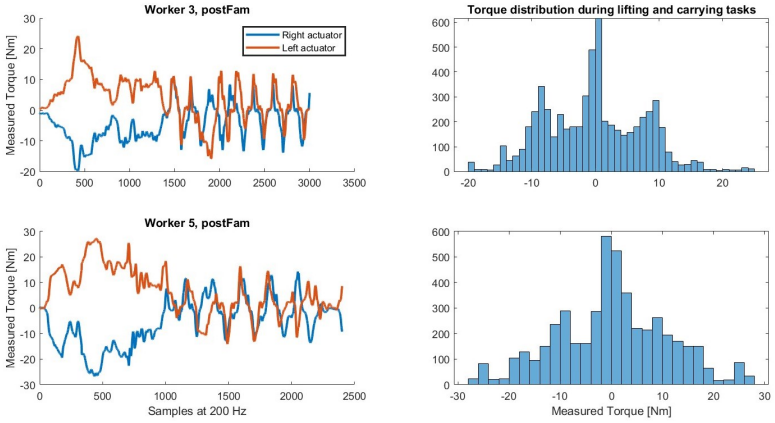


Figure 2. Example of typical trends in right- and left-motor activity for workers 3 and 5 during lifting and carrying tasks.

2.2. Experimental Protocol

Field tests were conducted to evaluate the efficacy of the exoskeleton in assisting construction workers. The test was performed in accordance with the experimental protocol approved by the Ethics Committee of Liguria, Italy (protocol number 001/2019) and in compliance with the Declaration of Helsinki. A pool of professional construction workers, characteristic of the industrial sector, was selected by a partner company FIRA, which also provided us with the site for testing. The subjects were 5 male workers with the following characteristics: age 30 $\pm$ 17 years (18–60 years), height 176 $\pm$ 6 cm (170–183 cm), with an average of 10 years of construction work experience, recruited on-site. Given the small sample size ($n = 5$), the statistical analyses had limited sensitivity to detect anything other than large and consistent effects. Accordingly, non-significant findings should be interpreted with caution, as they may reflect insufficient statistical power rather than the absence of an effect.

The experiment was carried out inside an under-construction building, where workers reproduced the typical hauling activity (composed of the three tasks: load lifting, carrying, and lowering) by moving common materials daily handled in the site. To be as faithful

as possible to realistic working environments and in ergonomically safe conditions, tasks, movement frequencies, distances, and weights were chosen as follows. Based on the space constraints and to reduce hindrance to the other workers in the site, two couples of pallets were placed 12 m and 8 m apart, defining two tracks as in Figure 1b. A stack of three bricks (10 kg in total) was assigned to the 12 m track, while a bag of cement (25 kg) was assigned to the 8 m track. To replicate the frequency proportions of a typical MMH task, the workers were asked to perform a complete hauling cycle, which means moving the brick stack from one pallet to the other (12 m) five times and then moving the concrete bag one time (8 m), Figure 1b. To limit ergonomic risks associated with the activity, workers were asked to recover by walking freely along the 12 m track four times after performing five consecutive hauling cycles. The workers were also advised to perform the hauling task and walk at the pace they would use in real-world scenarios.

In this contribution, each worker was tested under two conditions, called NoExo (without wearing the exoskeleton) and Exo (wearing the exoskeleton with assistance provided). The tests were conducted on two different non-consecutive days, and the familiarization period was performed between the two testing days. The first day of tests is called preFam, while the second day is postFam. The NoExo, Exo-preFam, and Exo-postFam conditions were alternated as reported in Table 1. Specifically, to avoid attributing muscle fatigue to previously performed activities, three out of five workers used the exoskeleton in the preFam condition in the morning, and two in the afternoon. The opposite order was used for post-Fam. Furthermore, to ensure the comparability of results between different test days and subjects, the exoskeleton assistance settings were kept constant between workers and the Exo-preFam and Exo-postFam conditions.

Table 1. Work shifts of the five workers tested in NoExo, Exo-preFam, and Exo-postFam conditions.

	First Day (preFam)		Second Day (postFam)	
	Morning	**Afternoon**	**Morning**	**Afternoon**
Worker 1	Exo-preFam	NoExo	NoExo	Exo-postFam
Worker 2	NoExo	Exo-preFam	Exo-postFam	NoExo
Worker 3	Exo-preFam	NoExo	NoExo	Exo-postFam
Worker 4	NoExo	Exo-preFam	Exo-postFam	NoExo
Worker 5	Exo-preFam	NoExo	NoExo	Exo-postFam

2.3. Familiarization Period

With the StreamEXO, the user's body and movements act as controllers for the exoskeleton. Thus, it is essential to fully and comfortably engage with the exoskeleton to understand the interaction forces, the system's response time, and the specific movements required to generate the device's desired effect. To master proper control of the exoskeleton, the familiarization period is a learning phase during which subjects can wear and use the exoskeleton to become accustomed to it and take full advantage of its control strategies. The duration of the familiarization period varies according to the individual's specific needs. Investigators use it to address workers' concerns and provide recommendations aimed at maximizing the benefits of interaction with the exoskeleton. The familiarization process applied in this work is shown in Figure 3. The preFam condition provides a general introduction to how the exoskeleton is designed and operates. After this introduction, the worker performs the test independently to learn how to control the device better. The postFam condition, on the other hand, is based on a series of processes that revolve around the users and adapt to their requirements. Following the diagram in Figure 3, the general introduction is followed by a task and its corresponding control strategy. The user practices the assigned task by performing the activity repeatedly until they feel confident they have truly understood how

the exoskeleton responds to their specific movements. During practice, they must evaluate the experience by describing their sensations and providing feedback to the operator. The operator's role is to coach users and instruct them on maximizing the exoskeleton's benefits. Based on the user's needs and feedback, the operator can assist by providing theoretical guidance, adjusting the exoskeleton's fit to improve wearability, or tuning the assistive strategy's settings. When the user is satisfied, the same process is applied to all other tasks. Once the process is complete, the worker has developed an awareness of the technology and is ready to perform the test in the postFam condition. This phase can last as long as needed to instruct the user. From this pool of workers, each user spent about 90 min on familiarization.

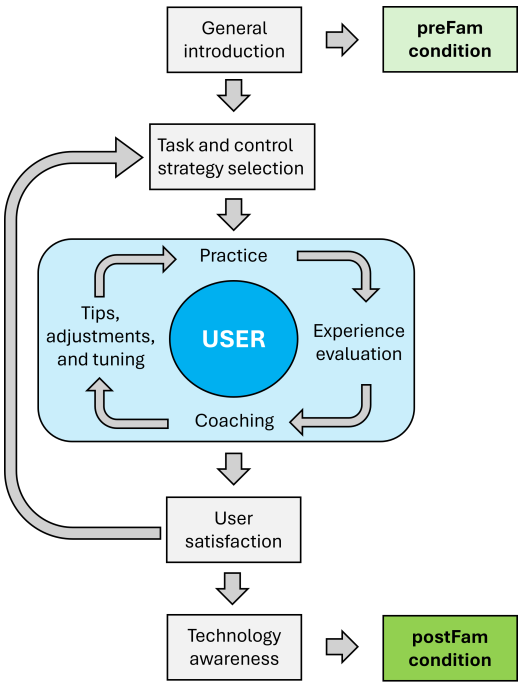


Figure 3. Flow chart describing the familiarization period with the exoskeleton. The chart illustrates the processes that characterize familiarization, distinguishing between preFam and postFam conditions. The preFam condition only relies on a general introduction. The postFam condition relies on a user-centered process where the user practices with the exoskeleton and its control strategies until they are fully aware of how the technology operates.

The familiarization period is a subjective process that depends on the device's nature, such as whether it is a passive or active exoskeleton. Also, for an active exoskeleton as in this work, the familiarization process would depend on the device complexity of use, control performance, user interface, and human factors. In fact, every subject has a different learning curve. Therefore, at this stage, it was difficult to establish a standardization process for familiarizing our exoskeleton. Instead, we rely on user feedback to assess the effectiveness of the familiarization process. Everyone who is trying to learn to use a new device applies a process of intuition [44]. In particular, such an intuitive process implies a rapid onset of perceptual processes, provides mechanisms for learning, and incorporates processes that show how perception is linked to action and emotion [45]. Therefore, in nature, it is a subjective process that can be characterized or standardized for each device or user interface, as evidenced by the statistical results from dedicated experiments. Moreover, such standardization of the process was out of the scope of this work.

Sections 2.1 and 2.2 report the selected tasks for this study are lifting, carrying, and lowering with the corresponding control strategies. To prevent introducing possible bias

among workers and in preFam and postFam conditions, we decided to keep all the settings of the control strategies constant throughout the testing activities.

2.4. Surveys

To evaluate the effects of the familiarization period, we collected data using several questionnaires: the Rating of Perceived Exertion (RPE), the NASA TLX, and a custom Acceptance Survey. A custom questionnaire is required because there are no standard questionnaires for evaluating exoskeletons for device acceptance, as reported by [32]. It includes the following constructs: assistance, comfort, stability, usability, and acceptability (questions are listed in the Appendix A).

- The *Rating of Perceived Exertion (RPE)* was used to evaluate perceived effort in the Upper Limbs, Trunk, and Lower Limbs during task execution. The survey relies on the Borg CR-10 scale [46] expressed through five verbal anchor points [47]. To strengthen results and evaluate their evolution over time at the Beginning, Middle, and End of the experiment, workers were asked whether the perceived exertion on their trunk was "Low", "Normal", "Moderate", "High", or "Very High". The RPE questionnaire was administered under NoExo, Exo-preFam, and Exo-postFam conditions.
- The raw *NASA TLX* questionnaire was used to evaluate workers' perceptions of the task. This investigates the workload dimensions (mental/physical/temporal demand, effort performance, and frustration) associated with the activity. The NASA TLX was administered at the end of the experiment under NoExo, Exo-preFam, and Exo-postFam conditions.
- The *Acceptance Survey* is a customized questionnaire developed and validated [48] to collect qualitative data on exoskeleton use [42]. It has been modified for these experiments to include questions that assess the effects of the new control strategies and the familiarization period. Questions are either open-ended or 7-point Likert-type scales, grouped into five constructs to investigate the exoskeleton's Acceptability, Assistance, Comfort, Stability, and Usability. The Acceptance Survey was administered only under the Exo-preFam and Exo-postFam conditions at the end of the experiments.

2.5. Data Analysis

The RPE was evaluated by calculating the average value and standard error of each body section (Upper Limbs, Trunk, and Lower Limbs) at the phases of Beginning, Middle, and End of each working task (Lifting, Carrying, and Lowering) in each of the three test conditions (NoExo, Exo-preFam, and Exo-postFam). Each condition included five workers, with RPE measurements collected at the Beginning, Middle, and End of each task, corresponding to three repeated observations per worker rather than independent samples. The distributions of NASA TLX and Acceptance Surveys (both administered once per worker after the Exo-preFam and Exo-postFam tests) were analyzed and visualized using bar plots, highlighting the first, second (median), and third quartiles for each condition. The spread (or deviation) of results is calculated as the difference between the 75th and 25th percentiles. Since the RPE used in this study has five anchor points, "Low" was assigned a value of 1, and "Very High" a value of 5. Similarly, for the NASA TLX, "Very Low" corresponds to 1 and "Very High" to 20, and for the Acceptance Survey, "Disagree" corresponds to 1 and "Agree" to 7. To compare the results obtained under the different conditions and evaluate the reliability [31] of the qualitative scores, we calculated the CoV [30] as in Equation (1):

$$CoV = \frac{\sigma}{\mu} \quad (1)$$

where σ and μ are, respectively, the standard deviation and mean value of the scores collected for each question. Considering questionnaires that use scales where higher scores

indicate better results, the CoV index decreases as the dispersion of scores decreases and their average increases. In these conditions, low CoV values correspond to positive results, as indicated by high scores that are consistent across subjects and can therefore be considered reliable. On the other hand, if the scores have a high dispersion and a low average, the CoV is high, and the results are poorly reliable. To meet these conditions (i.e., having high scores correspond to positive results and vice versa), the rating scales of some questionnaires were reversed to calculate the CoV. In this specific instance, the scale was reversed for the entire RPE survey, all NASA TLX questions except Performance, and for Q10, Q11, and Q12 of the Acceptance Survey.

The proposed CoV-based approach does not assess the psychometric reliability of the questionnaires, but rather the consistency of users' opinions in terms of their relative dispersion and degree of consensus. Lower CoV values indicate greater agreement among participants, which is used here as an indicator of the reliability of the collected subjective feedback. Psychometric reliability is measured using statistical techniques to quantify the internal and temporal consistency of a psychological assessment, which is outside the scope of this work.

To more easily interpret the meaning of CoV and thus the reliability of the results, we defined five quality ranges (i.e., very high, high, good, medium, and low) based on the ratio between σ and μ. Their definition and interpretation are given in Table 2. The "Low" reliability was set considering the maximum ratio between σ and μ of 1/3 (i.e., 33.33%), as in [31], but approximated by excess. Subsequently, the ratio was reduced to define "Very High" reliability, which is indicated by the traditional ratio of σ to μ of 1/20 (i.e., 5%) [49].

Table 2. Confidence intervals to evaluate the quality of the CoV index. Confidence intervals are based on the σ to μ ratio. Very High reliability is represented in purple, High reliability in green, Good in yellow, Medium in orange, and Low in red.

Condition	Quality Range	Reliability
$if \ \ (0 \leq \sigma \leq \frac{1}{20}\,\mu)$	$0 \leq \text{CoV} \leq 0.05$	Very High
$if \ \ (\frac{1}{20}\,\mu < \sigma \leq \frac{1}{8}\,\mu)$	$0.05 < CoV \leq 0.125$	High
$if \ \ (\frac{1}{8}\,\mu < \sigma \leq \frac{1}{4}\,\mu)$	$0.125 < \text{CoV} \leq 0.25$	Good
$if \ \ (\frac{1}{4}\,\mu < \sigma \leq \frac{1}{3}\,\mu)$	$0.25 < CoV \leq 0.33$	Medium
$if \ \ (\sigma > \frac{1}{3}\,\mu)$	$CoV > 0.33$	Low

Finally, the questionnaire results were tested for statistical significance. The RPE was analyzed separately for each working task (Lifting, Carrying, and Lowering) using a linear mixed-effects model to assess the effects of experimental condition (NoExo, Exo-preFam, and Exo-postFam), measurement time (Beginning, Middle, and End), and their interaction. Worker identity was included as a random effect to account for repeated measurements within participants ($n = 5$). Statistical significance was evaluated using ANOVA on the fitted model, followed by Bonferroni-adjusted post hoc comparisons. The NASA-TLX was evaluated using the Friedman test to examine the effects of the three modalities (NoExo, Exo-preFam, and Exo-postFam) in a non-parametric repeated-measures design for ordinal outcomes. The test relies on the null hypothesis stating that the distribution of ordinal outcomes is identical across all three modalities. p-values below significance thresholds reject the null hypothesis, concluding that a statistically significant difference exists among the three conditions. The Acceptance Survey, which compares only Exo-preFam and Exo-postFam, was evaluated using the Wilcoxon test. Considering the reduced number of subjects (i.e., five workers), the p-value could not be less than $p = 0.0625$. Therefore, no statistically significant results were observed in this survey. The CoV scores were analyzed using a parametric repeated-measures analysis of variance (ANOVA) to compare the relative

dispersion of workers' subjective responses across experimental conditions, treating each task-phase combination or questionnaire item as a matched observation. This analysis was intended to assess differences in CoV patterns across experimental conditions rather than infer differences between individual workers. The Bonferroni correction was applied to the results to account for multiple comparisons. The reported post hoc p-values are Bonferroni-adjusted and are therefore compared against the conventional significance thresholds of $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***). Statistical results are reported only when they are significant.

3. Results

3.1. Rating of Perceived Exertion (RPE)

3.1.1. Upper Limbs

The RPE for the Upper Limbs is shown in Figure 4a. The statistical analysis showed no significant differences between modalities during Lifting and Carrying. During Lowering, a significant difference was observed between NoExo and Exo-preFam (Bonferroni-adjusted $p = 0.0135$). NoExo and Exo-postFam conditions always report a positive trend of Perceived Exertion (PE) from Beginning to End of each working task. In contrast, Exo-preFam reports growth of RPE from Begin to Middle and a decrease from Middle to End, as in Lifting and Lowering. The average RPEs across all working tasks, calculated at the Beginning, Middle, and End, are always below 3.1. At the Beginning, the average NoExo is 2.0 ± 0.1 and increases to 2.9 ± 0.2 at the End. Exo-preFam starts at 1.8 ± 0.3 and increases to 2.6 ± 0.5 at the End, while Exo-postFam starts at 1.5 ± 0.0 and finishes at 3.1 ± 0.1.

Considering the CoV calculated for the Upper Limbs, shown in Table 3, High and Good reliability characterizes the Beginning of each task. The Middle and End phases, in contrast, report CoV values constantly greater than 0.25, denoting Medium to Low reliability. The average CoV denotes Medium reliability in all three modalities.

Table 3. Coefficient of variation (CoV) calculated from the RPE for each working task (Lifting, Carrying, and Lowering) in three working phases (Beginning, Middle, and End) in the NoExo, Exo-preFam, and Exo-postFam conditions. The confidence intervals and the color legend are reported in Table 2. To represent high reliability with low values, the scale of the questionnaire was reversed for all the questions. In the last column, the results of the statistical analysis performed with the CoV values are shown. Statistical significance between experimental conditions is indicated by * (Bonferroni-adjusted $p < 0.05$), ** ($p < 0.01$), and *** ($p < 0.001$). Further details are in Section 2.5.

	Upper Limbs										
	Lifting			Carrying			Lowering			Average	Statistical Significance
	Begin	Middle	End	Begin	Middle	End	Begin	Middle	End		
CoV NoExo	0.20	0.30	0.36	0.19	0.26	0.34	0.19	0.36	0.30	0.28	-
CoV Exo-preFam	0.18	0.29	0.29	0.18	0.39	0.27	0.12	0.37	0.35	0.27	-
CoV Exo-postFam	0.13	0.29	0.35	0.13	0.29	0.38	0.13	0.34	0.47	0.28	-
	Trunk										
	Lifting			Carrying			Lowering			Average	Statistical Significance
	Begin	Middle	End	Begin	Middle	End	Begin	Middle	End		
CoV NoExo	0.31	0.30	0.30	0.27	0.30	0.32	0.27	0.37	0.30	0.30	** (NoExo & Exo-preFam)
CoV Exo-preFam	0.12	0.23	0.20	0.31	0.20	0.16	0.12	0.23	0.20	0.20	-
CoV Exo-postFam	0.11	0.23	0.12	0.11	0.23	0.12	0.11	0.23	0.00	0.14	*** (NoExo & Exo-postFam)
	Lower Limbs										
	Lifting			Carrying			Lowering			Average	Statistical Significance
	Begin	Middle	End	Begin	Middle	End	Begin	Middle	End		
CoV NoExo	0.12	0.15	0.15	0.16	0.16	0.18	0.16	0.15	0.18	0.16	* (NoExo & Exo-preFam)
CoV Exo-preFam	0.20	0.23	0.13	0.20	0.34	0.34	0.20	0.35	0.23	0.25	-
CoV Exo-postFam	0.11	0.13	0.26	0.11	0.13	0.20	0.11	0.13	0.16	0.15	-

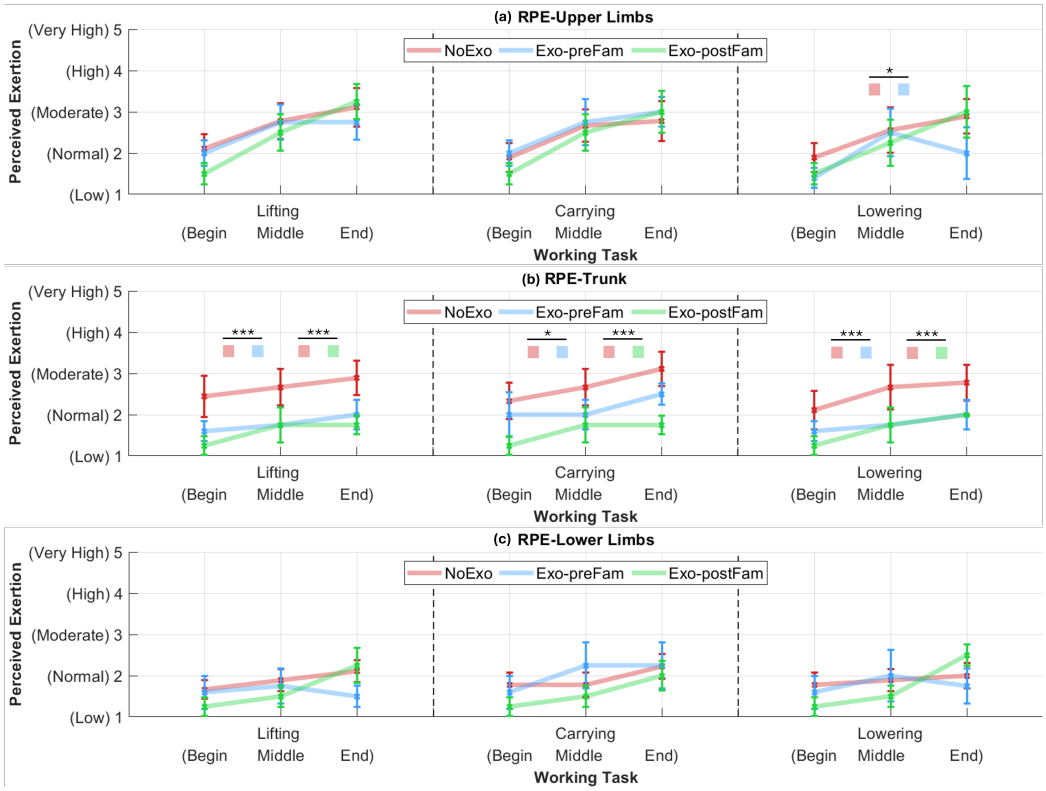


Figure 4. Rate of perceived exertion calculated for each working task (Lifting, Carrying, and Lowering) in three working phases (Beginning, Middle, and End). (**a**) refers to the Upper Limbs, (**b**) to the Trunk, and (**c**) to the Lower Limbs. The color red is associated with the NoExo condition, blue with Exo-preFam, and green with Exo-postFam. Statistical significance between experimental conditions is indicated by * (Bonferroni-adjusted $p < 0.05$) and *** ($p < 0.001$).

3.1.2. Trunk

The Trunk RPE trends are shown in Figure 4b. The statistical analysis showed significant differences during Lifting between NoExo and Exo-preFam ($p = 0.000016$) and between NoExo and Exo-postFam ($p = 0.000005$). During Carrying, significant differences were found between NoExo and Exo-preFam ($p = 0.0111$) and between NoExo and Exo-postFam ($p = 0.000045$). Similarly, during Lowering, significant differences were observed between NoExo and Exo-preFam ($p = 0.000947$) and between NoExo and Exo-postFam ($p = 0.000384$). All reported p-values are Bonferroni-adjusted. No significant differences were found between Exo-preFam and Exo-postFam across the three tasks. Considering the graphs, all data show a progressive increase in exertion from the Beginning to the End of each working task. The NoExo condition always has the highest RPE values, Exo-postFam has the lowest, and Exo-preFam is in between. On average, NoExo at the Beginning is 2.3 ± 0.2, and at the End is 2.9 ± 0.2. Exo-preFam starts from 1.7 ± 0.2 and finishes at 2.2 ± 0.3, while Exo-postFam reports 1.2 ± 0.0 and 1.8 ± 0.1, respectively.

The CoV values in Table 3 suggest that Exo-postFam is the modality with the highest reliability, with CoV scores consistently Good to High. Moreover, at the End of Lowering, the reliability is Very High with a CoV score of 0, indicating unanimity among all workers. Considering the Exo-preFam, the data reliability ranges from High to Medium, whereas in NoExo, it ranges from Medium to Low. The average CoV in NoExo modality denotes Medium reliability, while Exo-preFam and Exo-postFam are considered Good. The statistical analysis performed on the CoV calculated for the Trunk shows high statistical significance (**, Bonferroni-adjusted $p = 0.0039$) between NoExo and Exo-preFam, and

very high statistical significance (***, Bonferroni-adjusted $p = 0.00034$) between NoExo and Exo-postFam.

3.1.3. Lower Limbs

The results concerning the RPE of Lower Limbs show heterogeneous behaviors across working tasks and phases (Figure 4c). The statistical analysis showed no significant differences between experimental conditions after Bonferroni correction across Lifting, Carrying, and Lowering. At the Beginning, the NoExo condition consistently reports the highest rate, averaging 1.7 ± 0.1, while Exo-postFam reports the lowest rate, averaging 1.2 ± 0. Exo-PreFam at the Beginning has an average value of 1.6 ± 0.0. On the contrary, considering the End, the NoExo condition reports an average value of 2.1 ± 0.1 and is always between Exo-preFam and Exo-postFam, whose average values are 1.8 ± 0.4 and 2.2 ± 0.2, respectively.

The CoV values calculated for the Lower Limbs (Table 3) show the highest reliability in the Exo-postFam condition, with scores consistently Good to High, except at the End of Lifting. In contrast, the lowest reliability is in Exo-preFam, with scores consistently ranging from Good to Medium, and one Low score at the Middle of the Lowering. NoExo reports always Good results with one case of High reliability at the Beginning of Lifting. The average CoV reports Good reliability in all three modalities. The statistical analysis of these CoV values reveals a significant difference between NoExo and Exo-preFam (*, Bonferroni-adjusted $p = 0.0197$).

3.2. NASA TLX, Perceived Workload

The results of the NASA TLX survey indicate the perceived workload associated with the task, as shown in Figure 5. Comparing the NoExo, Exo-preFam, and Exo-postFam conditions, NoExo generally reports the worse feedback, as measured by median score and spread, while Exo-postFam is the most appreciated. Specifically, considering Mental Demand: both the NoExo and Exo-preFam conditions report a median of 5, while the Exo-postFam condition reports a median of 4. The spreads, calculated as the difference between the 75th and 25th percentiles, are 8.5, 6, and 3, respectively, for NoExo, Exo-preFam, and Exo-postFam. The highest Physical Demand is 13 ± 2.75 in the NoExo condition, while the lowest score is 8 ± 7 in the Exo-preFam condition, and the Exo-postFam is in between with 11 ± 7.5. Temporal Demand and Effort exhibit similar behaviors with the highest median values in the NoExo condition, the lowest in Exo-postFam, and Exo-preFam in between, with the widest standard deviation. The median values and deviation associated with Temporal Demand is 11 ± 2.75 for NoExo, 8.5 ± 7 for Exo-preFam, and 6 ± 3.75 for Exo-postFam. Considering Effort, the NoExo condition reports 14 ± 3.5, Exo-preFam 9 ± 6, and Exo-postFam 9 ± 1.5. The Performance reports 16 ± 7.25 in the NoExo condition, 15 ± 1.5 in Exo-preFam, and 17 ± 2.25 in Exo-postFam. To conclude, the graphs in Figure 5 show that the Frustration dimension exhibits a wide variation in all three conditions. NoExo has a median value of 10 ± 13.25, Exo-preFam is 9 ± 11, and Exo-postFam reports a median value of 14 ± 10.5. When averaging across all NASA-TLX answers, the average value for each testing condition is calculated as described in Section 2.5, except that only the Performance score is reversed. Thus, low values correspond to a lower workload. NoExo reports an average workload score of 9.5 ± 4.1, Exo-preFam reports 7.4 ± 1.9, and Exo-postFam reports 7.8 ± 4.2.

Considering the CoV indices calculated from this survey (Table 4), NoExo has the lowest reliability. The scores in NoExo are consistently Medium to Low in the Performance dimension, which reports Good reliability. Exo-preFam always reports Low reliability in Mental Demand, which scores well, and Performance, which scores High. Exo-postFam

reports Low reliability in Performance, Medium and Good in Physical and Temporal demand, and High in the three remaining dimensions. The average CoVs in the NoExo and Exo-preFam conditions are 0.40 and 0.36, respectively, indicating Low reliability. In contrast, the average CoV in Exo-postFam is 0.25, suggesting Good reliability.

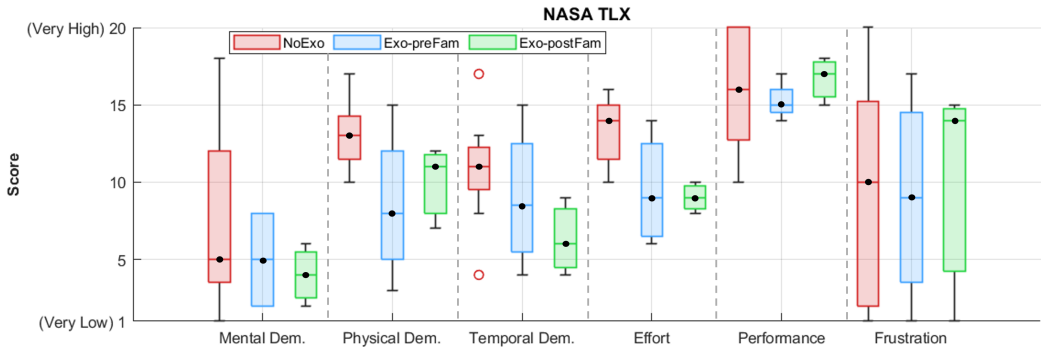


Figure 5. Results of the NASA TLX survey to evaluate the Perceived Workload associated with the working activity. Red represents the NoExo condition, blue the Exo-preFam, and green the Exo-postFam. Median values are marked with a black dot.

Table 4. Coefficient of variation (CoV) calculated for each question of the NASA TLX in NoExo, Exo-preFam, and Exo-postFam conditions. The confidence intervals and the color legend are reported in Table 2. To represent high reliability with low values, the scale of the questionnaire was reversed for all the questions excluding Performance (see Section 2.5).

	Mental Dem.	Physical Dem.	Temporal Dem.	Effort	Performance	Frustration	Average
CoV NoExo	0.44	0.32	0.39	0.32	0.24	0.72	0.40
CoV Exo-preFam	0.23	0.43	0.43	0.35	0.08	0.63	0.36
CoV Exo-postFam	0.12	0.26	0.18	0.09	0.09	0.78	0.25

3.3. Acceptance Survey

Acceptability, Figure 6a reports an overall median value of 5.3 ± 1.8 in the Exo-preFam condition and a median value of 6 ± 1.4 in the Exo-postFam. Considering Q1 and Q2, workers are satisfied with the device and believe it is useful in both preFam and postFam conditions. In Q1, the median value in Exo-preFam is 6 ± 2, while in Exo-postFam, it is 6.5 ± 1.5. In Q2, the median value in Exo-preFam is 6 ± 1.25, while in Exo-postFam, it is 6.5 ± 1.5. Considering Q4, workers would regularly use the exoskeleton in both preFam and postFam conditions. The associated scores are 6 ± 1.25 and 7 ± 1, respectively. The score attributed to Q5 (i.e., daily continuous use of the device), however, drops from 5 ± 2.75 in Exo-preFam to 4.5 ± 1.5 in the Exo-postFam. The Q6 overall rating in Exo-preFam has a median value of 5 ± 1.25, while in Exo-postFam, the median score is 6.5 ± 1.5.

Assistance is reported in Figure 6b. High score values correspond to benefits in all fields except when associated with Q10, Q11, and Q12. Therefore, to calculate the overall median value of the Assistance Survey and the CoVs, the scores for these fields were reversed (1 corresponds to 7, 2 to 6, 3 to 5, and so on). The overall median score and spread associated with Exo-preFam are 5.1 ± 2.2, while those for Exo-postFam are 6 ± 1.4. Considering fatigue, as reported in Q7, the workers experience less fatigue while using the exoskeleton. The median value is 6 ± 1.25 in preFam and 7 ± 1 in postFam. The comparison between Exo-preFam and Exo-postFam on the Q9 item shows a large difference between the two conditions. The synchronism between the exoskeleton and worker movements increases from 3 ± 0.5 in preFam to 4.5 ± 1.5 in postFam. In both conditions, the assistance fits the worker's needs (Q14), with a score of 6, but the spread decreases from 3 in preFam to 2 in postFam. To conclude, the need for a familiarization period in Q13 was scored 4 ± 3 in Exo-preFam and 6 ± 2 in Exo-postFam.

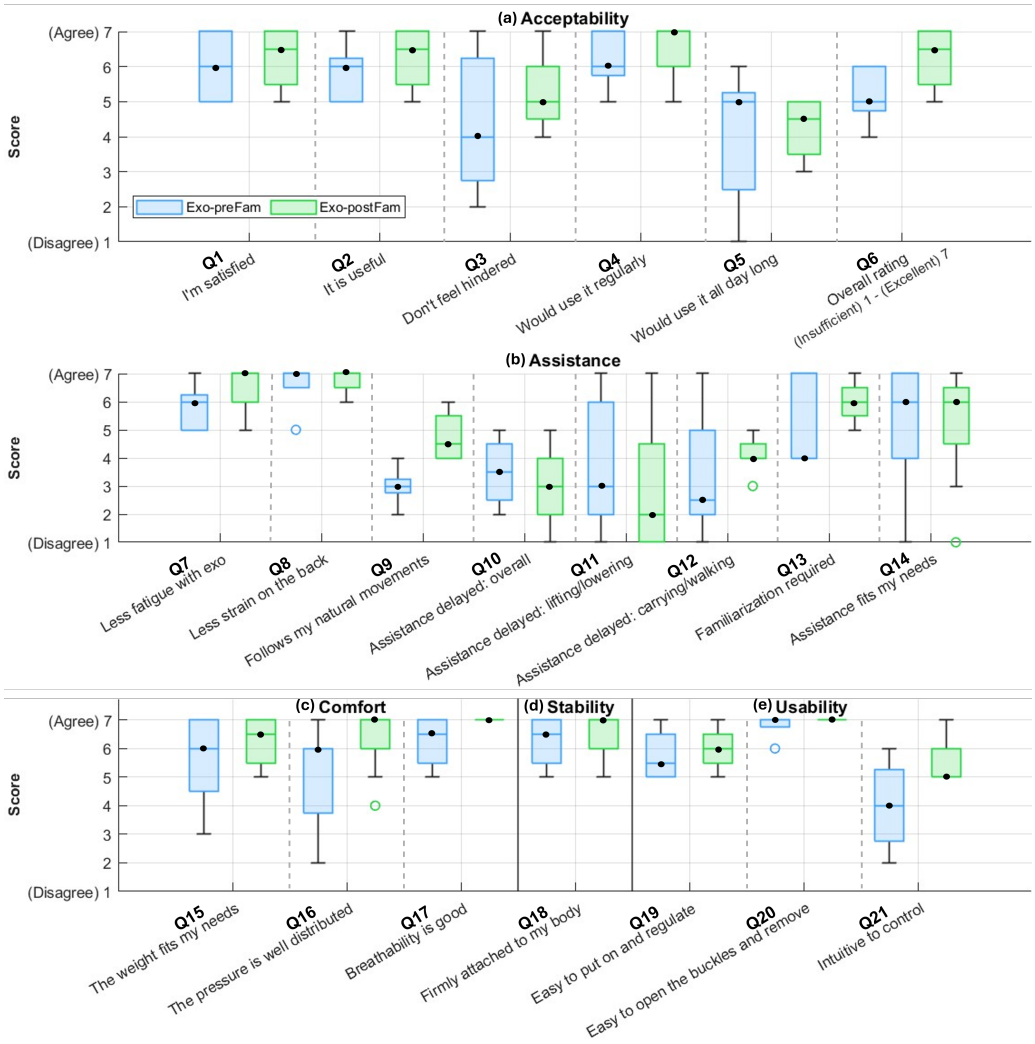


Figure 6. Results of the Acceptance Survey divided by construct: (**a**) Acceptability, (**b**) Assistance, (**c**) Comfort, (**d**) Stability, and (**e**) Usability. The color blue refers to Exo-preFam and the green to Exo-postFam. Median values are marked with a black dot.

The Comfort results (Figure 6c), consistently report high scores, with median values consistently above 6. The maximum spreads are 2.5 in Exo-preFam and 1.5 in Exo-postFam. Exo-postFam median values are always higher than Exo-preFam, and Q17 (i.e., breathability) unanimously scores 7 for all the workers in the postFam condition. The results analysis for Q16 (i.e., pressure distribution over the body) shows a significant improvement over Exo-preFam and Exo-postFam, as shown in Figure 6. The overall median value associated with Comfort in Exo-preFam is 6.2 ± 2.1, and in Exo-postFam is 6.8 ± 0.8.

Stability is shown in Figure 6d. Exo-preFam scores 6.5 ± 1.5. Exo-postFam scores 7 ± 1.

Finally, the Usability results are shown in Figure 6e. The median values for donning and regulation (Q19) are 5.5 ± 1.5 and 6 ± 1, respectively, for the preFam and postFam conditions, while the doffing scores (Q20) are 7 in both conditions. Q21 (i.e., the system is easy to control) reports a median value of 4 ± 2.5 in Exo-preFam and 5 ± 1 in Exo-postFam. Overall, the Usability score in the Exo-preFam condition is 5.5 ± 1.4; in the Exo-postFam condition, the score is 6 ± 0.7.

By averaging the overall results reported for Acceptability, Assistance, Comfort, Stability, and Usability, the final scores and spreads representative of exoskeleton acceptance under both conditions are extracted. Q10, Q11, and Q12 were reversed as for the CoV calculation. The average scores in the Exo-preFam condition are in order: 5.3 ± 0.8, 4.9 ± 1.3,

6.2 ± 0.3, 6.5 ± 1.5, and 5.5 ± 1.5. The average scores in the Exo-postFam condition are in order: 6 ± 1, 5.7 ± 1.1, 6.8 ± 0.3, 7 ± 1, and 6 ± 1. In Exo-preFam, the final and total score are 5.7 ± 1.8. In Exo-postFam, the final value is 6.3 ± 0.6.

To conclude, the CoV scores calculated from the Acceptance Survey responses are in Table 5. In general, Exo-preFam reports higher CoV scores than Exo-postFam. Both the Exo-preFam and Exo-postFam conditions report the lowest score in Q20 and the highest in Q11. Specifically, their CoV values in the Exo-preFam condition are 0.07 and 0.53, respectively, while in the Exo-postFam condition, they are 0 and 0.43, respectively. Exo-postFam also reports unanimous values for Q17, resulting in a CoV equal to 0 again. Moreover, the average CoV for the Exo-preFam condition denotes Medium reliability, while Exo-postFam reports an average Good reliability. Finally, the statistical analysis of the CoV values calculated for the Acceptance Survey reports very high statistical significance (***, $p < 0.001$) between the Exo-preFam and Exo-postFam conditions.

Table 5. Coefficient of variation (CoV) calculated for each question of the Acceptance Survey in Exo-preFam and Exo-postFam conditions. The confidence intervals and the color legend are reported in Table 2. To represent high reliability with low values, the scale of the questionnaire was reversed for the questions Q10, Q11, and Q12. In the last column, the results of the statistical analysis were performed with the CoV values. Statistical significance between experimental conditions is indicated by ***, meaning Bonferroni-adjusted $p < 0.001$. Further details are in Section 2.5.

	Acceptability								**Assistance**					
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q10**	**Q11**	**Q12**	**Q13**	**Q14**
CoV Exo-preFam	0.17	0.14	0.47	0.13	0.50	0.16	0.14	0.14	0.24	0.29	0.53	0.44	0.32	0.39
CoV Exo-postFam	0.15	0.15	0.24	0.15	0.23	0.15	0.15	0.07	0.20	0.33	0.43	0.16	0.14	0.32

	Comfort			**Stability**	**Usability**			**Average**	**Statistical Significance**
	Q15	**Q16**	**Q17**	**Q18**	**Q19**	**Q20**	**Q21**		
CoV Exo-preFam	0.30	0.33	0.13	0.14	0.17	0.07	0.40	0.27	***
CoV Exo-postFam	0.15	0.19	0.00	0.10	0.14	0.00	0.18	0.17	(Exo-preFam & Exo-postFam)

4. Discussion

4.1. Impact of the Familiarization on Workers

The impact of the familiarization period is assessed by comparing the preFam and postFam conditions while analyzing the evolution of workers' perceived exertion and workload, and their acceptance of StreamEXO.

4.1.1. Rating of Perceived Exertion

Results of the RPE on the Trunk in the Exo condition are consistently lower than in the NoExo condition. Specifically, Exo-postFam consistently reports the lowest average and standard deviation values. This is evident from both the graphs in Figure 4b and the values reported in the text. Comparing the preFam and postFam conditions, the difference is clear in the Lifting task and becomes more pronounced in the Carrying task. The results suggest that with a proper understanding of the specific control strategies, the perceived exertion rate can decrease by an average of 0.5 points out of 5, and workers' opinions converge. The average CoVs suggest Good reliability of the results in both conditions, empowering these outcomes. The benefits of understanding exoskeleton operation and control strategies are also evident in the results regarding the Lower Limbs' performance during Carrying, Figure 4c. During Carrying, the trend in preFam condition is higher than in NoExo, while postFam is consistently lower and exhibits a narrower spread. The Lifting and Lowering in the postFam condition report low values that progressively increase from the Beginning to the End of the task. Interestingly, the preFam condition shows a different trend, starting higher than postFam and decreasing sharply by the End of the

shift. This outcome suggests a possible bias caused by the use of a new technology for the first time [50]. Thus, emotional arousal or prejudices about using a new technology may lead workers in the preFam condition to perceive an exertion trend that does not follow the logical progression of fatigue during the task, which should increase from the beginning to the end of task execution, as in the NoExo and Exo postFam conditions. In some cases, such as Lifting, the exertion at the End of the shift is even less than at the Beginning. A similar effect is visible in the Upper Limbs trends, which are not even supported by the exoskeleton (Figure 4a). Specifically for this body section, the trends in the preFaM condition report lower exertion at the End of the shift than at the Middle, and it is also below the NoExo condition rather than equal to it. Therefore, the worker's expectation of the new technology leads to an overestimation of the device's effectiveness and a consequent reduction in RPE in body regions not directly supported by the device. In contrast, more coherent trends in RPE emerge once the worker is fully aware of the exoskeleton's performance, as evidenced by lower CoV values (Table 3).

4.1.2. Perceived Workload

The graphs in Figure 5 show positive results in both the Exo-preFam and Exo-postFam conditions, outperforming the scores collected in the NoExo modality. Considering Mental Demand, for instance, NoExo and Exo-preFam report the same median values, and Exo-postFam is one score lower. Interestingly, the dispersion of these results decreases as familiarity with the exoskeleton increases. The same positive outcomes, characterized by better or equal scores and lower spreads in favor of the postFam modality compared to preFam, are observed for Temporal Demand and Effort. Thus, after becoming familiar with the device and learning its operability, the Mental and Temporal dimensions, as well as the Effort associated with the task, are reduced and converge to lower values. The Performance dimension remains stable across the conditions. Similarly, the Frustration score showed only a marginal increase (7.4 ± 1.9 in Exo-preFam and 7.8 ± 4.2 in Exo-postFam). However, its variability increased substantially, indicating heterogeneous responses among participants rather than a consistent increase in frustration. This trend is difficult to interpret. One possible explanation is an order effect, as the Exo-postFam condition was always administered at the end of the testing campaign (as shown in Table 1), after participants had repeated the experimental tasks several times. Nevertheless, because the testing order was not counterbalanced, this interpretation remains speculative, and other factors, such as accumulated fatigue or individual variability, cannot be excluded. The Physical Demand dimension also shows interesting behavior. In Exo-preFam, the score is higher than both the NoExo and Exo-postFam conditions. However, the preFam score shows a high dispersion, which does not indicate a shared perception among workers, as confirmed by the corresponding CoV. On the other hand, all the results in the postFam condition, except Frustration, suggest that awareness of the exoskeleton's functioning aligns with user opinion reducing data spread and providing greater robustness to the results. The average CoVs associated with this survey indicates Low reliability in the results for NoExo and Exo-preFam, while Exo-postFam shows a Good level of reliability.

4.1.3. Acceptance

Most results from the Acceptance Survey show an improvement in the Exo-postFam condition, highlighting the benefits of the familiarization period for workers' acceptance of the exoskeleton. The overall Acceptability, which increases by 0.7 out of 7 from Exo-preFam to Exo-postFam, underscores the importance of explaining to workers how the device operates. The improvement is evident in the rates of worker satisfaction (Q1) and perceived usefulness (Q2). It becomes prominent when considering the Overall rating in

Q6, which improves by 1.5 out of 7 in the Exo-postFam condition. Interestingly, Q5 reports a countertrend reducing the score in posFam. Workers reported that they would use it only for the specific tasks that are assisted (i.e., hauling activities), and not for other tasks that characterize their work. Also, the Assistance construct of the questionnaire reveals the benefit of the familiarization period, showing an overall improvement of 0.8 out of 7 from Exo-preFam to Exo-postFam. Despite the delay in assistance during carrying/walking (Q12), whose rates worsen in the Exo-postFam condition, all the other items show improvement, which is particularly clear given the reduction in fatigue (Q7) and the synchrony with workers' natural movements (Q9).

The CoV associated with Q12, however, decreases from 0.44 in preFam to 0.16 in postFam, suggesting greater reliability of the result obtained after familiarization. Finally, the need for a familiarization period has a significant impact on the specific Q13 item, whose median value increases from 4 to 6 out of 7 in Exo-postFam and whose associated CoV drops from 0.32 in Exo-preFam to 0.14 in Exo-postFam. Thus, after experiencing the effects of the familiarization period, workers themselves attest to its crucial importance. Considering the Comfort, Stability, and Usability constructs, a regular improvement in the Exo-postFam condition is again visible through an increase in the median value and/or a reduction in the data spread. Among other questions, after the familiarization period, the exoskeleton's ease of control (Q21) improves by 1 score out of 7 compared to Exo-preFam, and the associated CoV drops from 0.40 in Exo-preFam to 0.18 in Exo-postFam. To conclude, it is interesting to note that the statistical analysis applied to the CoV scores calculated for the Assistance Survey reports a very high level of significance (***, $p < 0.001$), indicating a substantial difference in results between Exo-preFam and Exo-postFam.

4.1.4. Effects of the Familiarization Period

The results in Figure 4a–c show that the familiarization period reduces perceived exertion across all body districts, with particular effectiveness in the assisted trunk area, where the average gap between Exo-preFam and Exo-postFam is 0.5 out of 5, supporting the familiarization effect. Considering the NASA TLX, Figure 5, and the Acceptance Survey, Figure 6a–e, there is a convergence in the results characterized by a lower dispersion of scores, and this insight becomes prominent considering the CoV scores in Tables 3–5. Thus, awareness of how the exoskeleton operates and how to leverage all its qualities aligns workers' opinions toward more consistent, reliable results. The same outcome, finally, is confirmed by considering the overall ratings extracted from each of the Acceptance Survey constructs, where the Exo-preFam condition gets a mean value of 5.7 ± 1.8, and the Exo-postFam gets a score of 6.3 ± 0.6. In conclusion, the results from the second day (Exo-postFam) outperform those from the first day (Exo-preFam). The importance of the familiarization is also confirmed by the open question Q14.2 (see Appendix A), where all the workers agree on the importance of the experience and offer no suggestions for modifying it. Thus, a familiarization period in which the user is instructed on the exoskeleton's functioning proves to be a reliable and useful method for fully exploiting the device's potential, converging users' opinions, and improving their acceptance. Moreover, the familiarization proves to be a tool that can compensate for the drop in scores characteristic of the second day of the in-field test reported in [32].

4.1.5. Bias in the First Use of the Technology

Generally, the results presented in the Exo-preFam condition indicate good perceptions of the effects of exoskeletons. Considering the RPE, exertion in the Exo-preFam condition is globally lower than in the NoExo condition. On the Upper Limbs (where the exoskeleton does not directly assist) and Lower Limbs, the perceived exertion at the End of the task can

be even lower than at the Beginning or Middle (i.e., when the worker should be fresher and more rested). However, as suggested by the CoV scores in Table 3 and their statistical analysis, the results concerning Exo-preFam in the Upper Limbs and Lower Limbs exhibit greater dispersion among workers' responses. According to the NASA TLX, the perceived task load scores in the Exo-preFam condition are generally good but exhibit substantial deviations. For instance, Physical Demand in Exo-preFam is drastically lower than in NoExo and Exo-postFam. At the same time, Effort reports the same median score as Exo-postFam but with a much higher deviation in the preFam condition. These good but fluctuating results that characterize the physical assistance in Exo-preFam condition, along with the CoV values in Table 4, and the considerations reported in Sections 4.1.1 and 4.1.2, are consistent with the observations reported in [32], suggesting that initial exposure to an exoskeleton may influence workers' subjective evaluations. Specifically, this study's results suggest a potential overestimation of the perceived beneficial effects of the exoskeleton, with workers reporting lower exertion in their arms and legs at the End of the work shift than at the beginning. A plausible explanation is that high expectations for a new technology can lead to worker emotional arousal [50], which may influence the perception of the device's beneficial effects and contribute to greater variability in the collected feedback. Indeed, the CoV scores in Exo-preFam for the Upper Limbs and Lower Limbs are higher than those in postFam (Table 3). In addition, we report two workers' comments while trying the exoskeleton for the first time: "With the exoskeleton, I feel like I've stepped out of Terminator", and "I feel like a warrior of the future".

4.2. Reliability of the Results

To evaluate the the consistency of users' opinions, we calculated the CoV of the questionnaire results (averaged results are reported in Table 6). Its interpretation is based on the assumption that, when analyzing subjective evaluations of a technology by several inexperienced users, greater consensus indicates more consistent feedback. Conversely, poorly shared opinions indicate greater variability among users. The subjectivity of the opinions and the heterogeneity of the responses are confirmed by the results collected in the NoExo modality, which show the greatest dispersion across most of the questionnaire items. In addition, the NoExo modality shows high CoVs and, therefore, Medium to Low reliability in all but one of the questionnaires. The use of a technology, even for the first time, polarizes users' opinions. In fact, when using the exoskeleton in the preFam condition, workers share common perceptions that reduce dispersion and stabilize or improve average CoVs, whose reliability becomes Medium to Good. Finally, after familiarization, during which users learn to operate the exoskeleton and its control strategies, the average scores improve, and the dispersion further reduces. The average CoVs stabilize at lower values and report a general Good reliability. Specific cases also show High and Very High reliability.

Therefore, given the deviation in the results and the derived CoVs, it seems possible to discriminate between the results obtained across the different modalities. The results in the Exo-postFam condition show greater agreement among workers and lower CoVs, whereas Exo-preFam and NoExo exhibit greater dispersion of responses, reflected by higher CoVs. Furthermore, the CoV analysis highlights differences in the relative dispersion of workers' responses across experimental conditions, providing complementary information to the analysis of questionnaire scores.

Table 6. Average scores and CoVs calculated for each survey in the NoExo, Exo-preFam, and Exo-postFam conditions. From the top: the exertion perceived on the trunk at the beginning and the end of the task, the workload associated with the task, and the acceptance score of the exoskeleton.

RPE-Trunk			
	NoExo	Exo-preFam	Exo-postFam
Average (Begin)	2.3 ± 0.2	1.7 ± 0.2	1.2 ± 0.0
Average (End)	2.9 ± 0.2	2.2 ± 0.3	1.8 ± 0.1
CoV	0.30	0.20	0.14
NASA-TLX			
	NoExo	Exo-preFam	Exo-postFam
Average	9.5 ± 4.1	7.4 ± 1.9	7.8 ± 4.2
CoV	0.40	0.36	0.25
Acceptance			
	NoExo	Exo-preFam	Exo-postFam
Average	-	5.7 ± 1.8	6.3 ± 0.6
CoV	-	0.27	0.17

4.3. Limitations of the Study

A principal limitation of this study is the small sample size ($n = 5$), comprising only five professional workers, due to recruitment, safety, and access constraints associated with testing in a real operational environment. The findings should therefore be considered exploratory and require confirmation in studies with larger sample sizes.

Although the RPE, NASA-TLX, and Acceptance Survey provide complementary information on perceived exertion, workload, and acceptance, their different measurement constructs do not allow direct comparison of scores. A dedicated cross-scale consistency analysis was not performed and remains an aspect for future investigation.

Furthermore, although familiarization may influence task-execution efficiency, task-completion time was not specifically evaluated in this study. Future investigations should assess its potential influence on work productivity.

In addition, future studies will place particular care on protocol design, employing randomized or counterbalanced testing to isolate better effects correlated with accumulated fatigue and individual variability.

5. Conclusions

Recent years have seen an increase in exoskeletons testing to introduce OEs into industries. Tests are often conducted on a limited subject pool, within a restricted period, and using qualitative evaluation methods such as questionnaires. Therefore, it is critical to verify whether the experimental protocol provides reliable results [11]. In this work, we investigated whether the first use of the occupational exoskeleton affects workers' feedback and whether a dedicated familiarization period impacts the exoskeleton's performance, perception, and acceptance.

A dedicated experimental test investigated the impact of a familiarization process, from an Exo-preFam condition, in which the exoskeleton was worn for the first time and used with minimal knowledge of its functionalities. In addition, we have introduced a CoV index to assess the the relative dispersion and degree of agreement among workers' subjective responses, even for a small sample.

The perceived physical exertion in the trunk after familiarization showed greater consistency among workers' responses, as indicated by lower CoV values.

Also, the exoskeleton's acceptance rate shows greater agreement among workers after the familiarization period. The reported evidence suggests that device familiarization may contribute to improving workers' evaluations of the device's performance and reducing fluctuations in their subjective responses [32].

The outcomes of this work, particularly the CoV index, can support the industrial adoption of exoskeletons. At the same time, an adequate familiarization may contribute to improving workers' acceptance of the exoskeleton.

In the future, we will be interested in applying this method to a larger number of subjects while also evaluating the impact of familiarization on qualitative results. In addition, we will develop a standard approach for streamEXO familiarization.

Author Contributions: Conceptualization, V.F., S.L. and C.D.N.; methodology, V.F., S.L. and C.D.N.; software, V.F. and S.L.; validation, V.F., S.L. and C.D.N.; formal analysis, V.F.; investigation, V.F. and S.L.; resources, C.D.N.; data curation, V.F. and C.D.N.; writing—original draft preparation, V.F.; writing—review and editing, V.F., S.L., D.G.C. and C.D.N.; visualization, V.F.; supervision, D.G.C. and C.D.N.; project administration, C.D.N.; funding acquisition, C.D.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the BEEYONDERS project funded by the European Union's Horizon Europe research and innovation programme under grant agreement N° 101058548. Responsibility for the information and views expressed in the paper/article lies entirely with the authors.

Institutional Review Board Statement: Tests were performed in accordance with the experimental protocol approved by the Ethics Committee of Liguria, Italy (protocol number 001/2019) and complied with the Helsinki Declaration.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data will be made available upon request to the corresponding author.

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

Appendix A

This section reports the questions of the custom Acceptance Survey for each construct: assistance, comfort, stability, usability, and acceptability. Each question is open-ended (open) or uses a 7-point Likert-type scale.

ACCEPTABILITY

Q1. According to my expectations, I am satisfied with the use of the device

Q2. I think the device is really useful for my activities

Q3. I don't feel hindered or slowed down by the device while working

Q4. I would use it regularly if it was available

Q5. I think I would use the device all day long

Q6. Give the device an overall rating

Q6.1. Would you suggest the use of the system to other colleagues with the same or different job task? What tasks? (open)

ASSISTANCE

Q7. I feel less fatigue when using the exoskeleton

Q8. I feel less strain on my back

Q9. I feel that the system follows my natural movements

Q10. The assistance is delayed with my movement: Overall

Q11. The assistance is delayed with my movement: Lifting, Lowering

Q12. The assistance is delayed with my movement: Carrying, Walking

Q13. A familiarization period is required to take full advantage of the device

Q14. The assistance level fits my needs

Q14.1. Do you feel any other part of your body overloaded? If yes, which one? (open)

Q14.2. Would you change anything in the familiarization phase? Provide any comments about familiarization. (open)

Q14.3. Please, describe how to improve the exoskeleton's control strategy (open)

COMFORT

Q15. Does the exoskeleton's weight fit my needs

Q16. The exoskeleton's pressure is well distributed on my body

Q17. I feel that the breathability is good

Q17.1. Describe other parts of the device that created discomfort, if any. (open)

STABILITY

Q18. The exoskeleton is firmly attached to my body

Q18.1. Do you feel any parts of the exoskeleton that are loose on your body? If yes, which one? (open)

USABILITY

Q19. It is easy to put on and to regulate for my size

Q20. It is easy to open the buckles and remove the exoskeleton

Q21. The device is intuitive to control

Q21.1. Do you think you will be able to wear it without any assistance? (open)

Q21.2. Would you rather wear the system continuously or doff and don when moving between tasks? (open)

References

1. De Looze, M.P.; Bosch, T.; Krause, F.; Stadler, K.S.; O'sullivan, L.W. Exoskeletons for industrial application and their potential effects on physical work load. *Ergonomics* **2016**, *59*, 671–681. [CrossRef] [PubMed]
2. Bao, G.; Pan, L.; Fang, H.; Wu, X.; Yu, H.; Cai, S.; Yu, B.; Wan, Y. Academic Review and Perspectives on Robotic Exoskeletons. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 2294–2304. [CrossRef] [PubMed]
3. Tian, J.; Wei, B.; Luo, S.; Yang, C.; Chen, C.; Liu, Y.; Feng, J.; Li, P.; Zhu, H.; Yi, C. A systematic review of occupational shoulder exoskeletons for industrial use: Mechanism design, actuators, control, and evaluation aspects. *Actuators* **2024**, *13*, 501. [CrossRef]
4. de la Tejera, J.A.; Bustamante-Bello, R.; Ramirez-Mendoza, R.A.; Izquierdo-Reyes, J. Systematic Review of Exoskeletons towards a General Categorization Model Proposal. *Appl. Sci.* **2020**, *11*, 76. [CrossRef]
5. Theurel, J.; Desbrosses, K. Occupational Exoskeletons: Overview of Their Benefits and Limitations in Preventing Work-Related Musculoskeletal Disorders. *IISE Trans. Occup. Ergon. Hum. Factors* **2019**, *7*, 264–280. [CrossRef]
6. Hales, T.R.; Bernard, B.P. Epidemiology of work-related musculoskeletal disorders. *Orthop. Clin. N. Am.* **1996**, *27*, 679–709. [CrossRef]
7. Wang, X.; Dong, X.S.; Choi, S.D.; Dement, J. Work-related musculoskeletal disorders among construction workers in the United States from 1992 to 2014. *Occup. Environ. Med.* **2017**, *74*, 374–380. [CrossRef] [PubMed]
8. De Kok, J.; Vroonhof, P.; Snijders, J.; Roullis, G.; Clarke, M.; Peereboom, K.; van Dorst, P.; Isusi, I. *Work-Related Musculoskeletal Disorders: Prevalence, Costs and Demographics in the EU*; Research Report; European Agency for Safety and Health at Work (EU-OSHA): Bilbao, Spain, 2019.
9. Murray, C.J.L.; Vos, T.; Lozano, R.; Naghavi, M.; Flaxman, A.D.; Michaud, C.; Ezzati, M.; Shibuya, K.; Salomon, J.A.; Abdalla, S.; et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380*, 2197–2223. [CrossRef] [PubMed]
10. Punnett, L.; Wegman, D.H. Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. *J. Electromyogr. Kinesiol.* **2004**, *14*, 13–23. [CrossRef] [PubMed]
11. Ahmad, J.; Fanti, V.; Caldwell, D.G.; Di Natali, C. Framework for the adoption, evaluation and impact of occupational Exoskeletons at different technology readiness levels: A systematic review. *Robot. Auton. Syst.* **2024**, *179*, 104743. [CrossRef]
12. Hoffmann, N.; Prokop, G.; Weidner, R. Methodologies for evaluating exoskeletons with industrial applications. *Ergonomics* **2021**, *65*, 276–295. [CrossRef] [PubMed]
13. Crea, S.; Beckerle, P.; De Looze, M.; De Pauw, K.; Grazi, L.; Kermavnar, T.; Masood, J.; O'Sullivan, L.W.; Pacifico, I.; Rodriguez-Guerrero, C.; et al. Occupational exoskeletons: A roadmap toward large-scale adoption. Methodology and challenges of bringing exoskeletons to workplaces. *Wearable Technol.* **2021**, *2*, e11. [CrossRef] [PubMed]
14. Olmo, M.d.; Domingo, R. EMG characterization and processing in production engineering. *Materials* **2020**, *13*, 5815. [CrossRef] [PubMed]

15. Pesenti, M.; Antonietti, A.; Gandolla, M.; Pedrocchi, A. Towards a functional performance validation standard for industrial low-back exoskeletons: State of the art review. *Sensors* **2021**, *21*, 808. [CrossRef] [PubMed]
16. Shalash, O.; Rowe, P. Computer-assisted robotic system for autonomous unicompartmental knee arthroplasty. *Alex. Eng. J.* **2023**, *70*, 441–451. [CrossRef]
17. Fattori, G.; Lomax, A.J.; Weber, D.C.; Safai, S. Technical assessment of the NDI Polaris Vega optical tracking system. *Radiat. Oncol.* **2021**, *16*, 87. [CrossRef] [PubMed]
18. Merriaux, P.; Dupuis, Y.; Boutteau, R.; Vasseur, P.; Savatier, X. A study of vicon system positioning performance. *Sensors* **2017**, *17*, 1591. [CrossRef] [PubMed]
19. Baltrusch, S.J.; Houdijk, H.; van Dieën, J.H.; Kruif, J.T.C.M.d. Passive Trunk Exoskeleton Acceptability and Effects on Self-efficacy in Employees with Low-Back Pain: A Mixed Method Approach. *J. Occup. Rehabil.* **2020**, *31*, 129–141. [CrossRef] [PubMed]
20. Wioland, L.; Atain Kouadio, J.J.; Bréard, H.; Clerc-Urmès, I.; Paty, B. The Adoption of Occupational Exoskeletons: From Acceptability to Situated Acceptance, Questionnaire Surveys. *Int. J. Hum.–Comput. Interact.* **2024**, *41*, 1446–1458. [CrossRef]
21. Gilotta, S.; Spada, S.; Ghibaudo, L.; Isoardi, M.; Mosso, C.O. Acceptability Beyond Usability: A Manufacturing Case Study. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 922–934. [CrossRef]
22. Elprama, S.A.; Vanderborght, B.; Jacobs, A. An industrial exoskeleton user acceptance framework based on a literature review of empirical studies. *Appl. Ergon.* **2022**, *100*, 103615. [CrossRef] [PubMed]
23. Siedl, S.M.; Mara, M. Exoskeleton acceptance and its relationship to self-efficacy enhancement, perceived usefulness, and physical relief: A field study among logistics workers. *Wearable Technol.* **2021**, *2*, e10. [CrossRef] [PubMed]
24. Kim, S.; Nussbaum, M.A.; Smets, M. Usability, User Acceptance, and Health Outcomes of Arm-Support Exoskeleton Use in Automotive Assembly: An 18-month Field Study. *J. Occup. Environ. Med.* **2021**, *64*, 202–211. [CrossRef] [PubMed]
25. Power, V.; O'Sullivan, L.; de Eyto, A.; Schülein, S.; Nikamp, C.; Bauer, C.; Mueller, J.; Ortiz, J. Exploring user requirements for a lower body soft exoskeleton to assist mobility. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*; ACM: New York, NY, USA, 2016; pp. 1–6.
26. Shore, L.; De Eyto, A.; O'Sullivan, L. Investigating Perceptions Related to Technology Acceptance & Stigma of Wearable Robotic Assistive Devices by Older Adults—Preliminary Findings. In *Proceedings of the DRS2018: Design as a Catalyst for Change*; DRS2018; Design Research Society: London, UK, 2018. [CrossRef]
27. Shore, L.; de Eyto, A.; O'Sullivan, L. Technology acceptance and perceptions of robotic assistive devices by older adults—Implications for exoskeleton design. *Disabil. Rehabil. Assist. Technol.* **2020**, *17*, 782–790. [CrossRef] [PubMed]
28. Deming, W.E. On Errors in Surveys. In *Sociological Methods*; Routledge: Abingdon, UK, 2017; pp. 317–319. [CrossRef]
29. Hobart, J.C.; Cano, S.J.; Warner, T.T.; Thompson, A.J. What sample sizes for reliability and validity studies in neurology? *J. Neurol.* **2012**, *259*, 2681–2694. [CrossRef] [PubMed]
30. Abdi, H. Coefficient of Variation. In *Encyclopedia of Research Design*; Salkind, N.J., Ed.; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2010; Volume 1, pp. 169–171. [CrossRef]
31. Statistics Canada. *Guide to the Labour Force Survey 2018*; Catalogue no. 71-543-g; Statistics Canada: Ottawa, ON, Canada, 2018.
32. Sposito, M.; Fanti, V.; Poliero, T.; Caldwell, D.G.; Di Natali, C. Field Assessment of Active BSE: Trends over Test Days of Subjective Indicators and Self-Reported Fatigue for Railway Construction Workers. *Heliyon* **2024**, *10*, e33055. [CrossRef] [PubMed]
33. Sugi, M.; Sakuraba, S.; Saito, H.; Miyazaki, M.; Yoshida, S.; Kamada, T.; Sakai, S.; Sawamura, D. Personality Traits Modulate the Impact of Emotional Stimuli During a Working Memory Task: A Near-Infrared Spectroscopy Study. *Front. Behav. Neurosci.* **2020**, *14*, 514414. [CrossRef] [PubMed]
34. He, Z.; Cassaday, H.J.; Bonardi, C.; Bibby, P.A. Do Personality Traits Predict Individual Differences in Excitatory and Inhibitory Learning? *Front. Psychol.* **2013**, *4*, 00245. [CrossRef] [PubMed]
35. Bonato, P.; Ebenbichler, G.R.; Roy, S.H.; Lehr, S.; Posch, M.; Kollmitzer, J.; Della Croce, U. Muscle Fatigue and Fatigue-Related Biomechanical Changes During a Cyclic Lifting Task. *Spine* **2003**, *28*, 1810–1820. [CrossRef] [PubMed]
36. Shadmehr, R.; Mussa-Ivaldi, F. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **1994**, *14*, 3208–3224. [CrossRef] [PubMed]
37. Kuber, P.M.; Rashedi, E. Training and Familiarization with Industrial Exoskeletons: A Review of Considerations, Protocols, and Approaches for Effective Implementation. *Biomimetics* **2024**, *9*, 520. [CrossRef] [PubMed]
38. Dufraisse, M.; Cegarra, J.; Atain Kouadio, J.J.; Clerc-Urmès, I.; Wioland, L. From unknown to familiar: An exploratory longitudinal field study on occupational exoskeletons adoption. *Appl. Ergon.* **2025**, *122*, 104393. [CrossRef] [PubMed]
39. Park, S.; Jung, M.K.; Kim, K.; Lim, H.; Yoon, J.; Hyun, D.J. Correlations Between Biomechanical Variables and Subjective Measures of Satisfaction While Using a Passive Upper-Limb Exoskeleton for Overhead Tasks in the Field. *IEEE Trans. Hum.-Mach. Syst.* **2025**, *55*, 176–184. [CrossRef]

40. Diamond-Ouellette, G.; Telonio, A.; Karakolis, T.; Leblond, J.; Bouyer, L.J.; Best, K.L. Exploring the Change in Metabolic Cost of Walking before and after Familiarization with a Passive Load-Bearing Exoskeleton: A Case Series. *IISE Trans. Occup. Ergon. Hum. Factors* **2022**, *10*, 161–172. [CrossRef] [PubMed]
41. Favennec, A.; Frère, J.; Mornieux, G. Changes in Human Motor Behavior during the Familiarization with a Soft Back-Support Occupational Exoskeleton. *Appl. Sci.* **2024**, *14*, 1160. [CrossRef]
42. Di Natali, C.; Poliero, T.; Sposito, M.; Fanti, V.; Leggieri, S.; Caldwell, D.G. From the idea to the user: A pragmatic multifaceted approach to testing occupational exoskeletons. *Wearable Technol.* **2025**, *6*, e28. [CrossRef] [PubMed]
43. Di Natali, C.; Poliero, T.; Fanti, V.; Sposito, M.; Caldwell, D.G. Dynamic and Static Assistive Strategies for a Tailored Occupational Back-Support Exoskeleton: Assessment on Real Tasks Carried Out by Railway Workers. *Bioengineering* **2024**, *11*, 172. [CrossRef] [PubMed]
44. Blackler, A.; Popovic, V.; Mahar, D.P. Designing for intuitive use of products: An investigation. In Proceedings of the 2003 Asian Design International Conference, Tsukuba, Japan, 14–17 October 2003; pp. 1–16.
45. Gobet, F.; Chassy, P. Expertise and intuition: A tale of three theories. *Minds Mach.* **2009**, *19*, 151–180. [CrossRef]
46. Williams, N. The Borg Rating of Perceived Exertion (RPE) scale. *Occup. Med.* **2017**, *67*, 404–405. [CrossRef]
47. Dawes, H.N.; Barker, K.L.; Cockburn, J.; Roach, N.; Scott, O.; Wade, D. Borg's Rating of Perceived Exertion Scales: Do the Verbal Anchors Mean the Same for Different Clinical Groups? *Arch. Phys. Med. Rehabil.* **2005**, *86*, 912–916. [CrossRef] [PubMed]
48. Sposito, M.; Poliero, T.; Di Natali, C.; Toxiri, S.; Anastasi, S.; Draicchio, F.; Monica, L.; Caldwell, D.; De Momi, E.; Ortiz, J. Physical comfort of occupational exoskeletons: Influence of static fit on subjective scores. In *Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference*; ACM: New York, NY, USA, 2021; pp. 60–68.
49. Spence, J.R.; Stanley, D.J. Concise, Simple, and Not Wrong: In Search of a Short-Hand Interpretation of Statistical Significance. *Front. Psychol.* **2018**, *9*, 02185. [CrossRef] [PubMed]
50. Korunka, C.; Huemer, K.; Litschauer, B.; Karetta, B.; Kafka-Lützow, A. Working with new technologies: Hormone excretion as an indicator for sustained arousal. A pilot study. *Biol. Psychol.* **1996**, *42*, 439–452. [CrossRef] [PubMed]

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