POLITECNICO DI TORINO
Repository ISTITUZIONALE

Usability Assessment in Parkinson’s Disease: the Case Study of the FarmExergame

Original
Usability Assessment in Parkinson’s Disease: the Case Study of the FarmExergame / Masi, G., Tonti, S., Ferraris, C., Olmo, G., Priano, L., Amprimo, G.. - (2025), pp. 464-469. (2025 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events, PerCom Workshops 2025 Washington DC (USA) 17-21 March 2025) [10.1109/percomworkshops65533.2025.00109].

Availability:
This version is available at: 11583/3003707 since: 2025-10-07T10:01:59Z

Publisher:
Institute of Electrical and Electronics Engineers

Published
DOI:10.1109/percomworkshops65533.2025.00109

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

25 June 2026

Usability Assessment in Parkinson's Disease: the Case Study of the FarmExergame

1st Giulia Masi
Politecnico di Torino
Italy
giulia.masi@polito.it

2nd Sofia Tonti
Politecnico di Torino
Italy
sofia.tonti@studenti.polito.it

3rd Claudia Ferraris
CNR
Italy
claudia.ferraris@cnr.it

4th Gabriella Olmo
Politecnico di Torino
Italy
gabriella.olmo@polito.it

5th Lorenzo Priano
IRCCS Istituto Auxologico Italiano
Italy
lorenzo.priano@unito.it

6th Gianluca Amprimo
Politecnico di Torino
Italy
gianluca.amprimo@polito.it

***Abstract*—Cognitive and motor rehabilitation through exergaming could democratize access to continuity of care, especially when delivered via low-cost and portable systems. However, since individuals with cognitive and/or motor impairments stand to benefit most from these systems, establishing their usability for such populations is both critical and challenging. This paper presents the FarmExergame, a low-cost and lightweight solution for hand dexterity rehabilitation. In particular, the findings from a preliminary two-weeks protocol designed to evaluate several aspects of its usability are discussed. The FarmExergame gamifies tasks commonly used in Parkinson's disease assessment while incorporating cognitive stimulation elements such as challenges, rewards, and both auditory and visual feedback. Among a cohort of 25 participants (9 with Parkinson's, 4 elderly controls, and 12 young controls), the exergame achieved an overall System Usability Score of 80.2, indicating high usability. The assessment emphasized the role of learnability and memorability as attributes of an exergaming system for people with Parkinson's disease and highlighted the value of objective performance measures over subjective questionnaires.**

***Index Terms*—Usability, Parkinson's disease, Mediapipe, Exergames**

I. Introduction

The increasing prevalence of neurodegenerative conditions such as Parkinson's Disease (PD) poses significant challenges for healthcare systems worldwide, particularly in ensuring fair access to effective and continuous care [1]. As neurodegeneration progresses, cognitive and motor rehabilitation plays a crucial role in ameliorating symptoms and improving quality of life [2], yet traditional therapeutic interventions often face barriers such as high costs, lack of accessibility, and limited scalability [3]. In this context, to guarantee continuity of use and fairness, new paradigms of *telerahabilitation* should focus on some key aspects: low costs, high usability, user engagement and efficacy of treatment.

Exergames may address these needs by providing a *gamified* rehabilitation. In exergames, motor and cognitive activities are used to control video-games, introducing a reward mechanism and an acoustic and visual stimulation that reinforces traditional rehabilitation. This approach is possible thanks to new interaction methods, such as video-based Human Pose Estimation (HPE), that leverage deep learning models and low-cost hardware to track body motion of the patient.

While many applications have been proposed and validated through extensive protocols [4], their widespread adoption is still lacking. Usability assessment may still remain a significant hurdle for this type of system, and it is also often overlooked despite its key role, especially when considering subjects with both motor and cognitive impairment as people with PD (PwPD). Most of usability assessments for exergames limit to the administration of qualitative questionnaires such as the System Usability Scale (SUS), administer after a single use of the system.

Modeling the concept of usability is a complex task. According to the ISO/IEC 9126-1 standard, usability is defined as *the degree of understandability, learnability, operability, and attractiveness of a system*. This definition was later expanded in ISO 9241-11, which emphasizes efficiency, effectiveness, and satisfaction [5]. While these standards are widely used, usability remains inherently context-dependent, leading some authors to propose their own definitions. Among them, the definition by Nielsen et al. [6] is the most used. This definition includes five specific attributes:

- **Learnability**: The ease in accomplish basic tasks after a first attempt.
- **Efficiency**: How quickly and effectively users can perform tasks after learning the system.
- **Memorability**: Users' ability to regain proficiency after a period of non-use.
- **Errors**: The frequency and severity of user errors and the ease of recovery from them.
- **Satisfaction**: The overall pleasantness and confidence users feel when using the system.

This definition may better suits the case of PwPD, as memorability and learnability are key aspects to ensure continuous, possibly unsupervised usage by elderly with different levels of cognitive decline.

This paper presents the results of the usability assessment conducted for the *FarmExergame*. The game implements a suit of tasks designed for rehabilitation of hand dexterity in PD and exploits MediaPipe tracking running on webcam inputs from an off-the-shelf notebook, posing as an extremely low cost solution for the tele-rehabilitation of symptoms. To fill the gap in current usability assessment of this type of systems, a protocol spanning two weeks and all the five aspects of Nielsen's usability was defined and tested both on healthy and pathological users. Moreover, quantitative performance parameters were also investigated to provide a complement to the results obtained by the qualitative feedback provided by questionnaires and from researcher's observations during testing. Therefore, this work contributes to the current literature both by presenting a low-cost and usable solution suitable for tele-rehabilitation, as well as by proposing a systematic approach to usability assessment in a sample population involving PwPD. In addition, the current limitations of this approach and possible directions of improvement for future investigations are also discussed.

II. Material & Methods

A. The FarmExergame

The FarmExergame was designed after a careful review of current exergaming solutions, to address the specific needs of monitoring disease progression and support rehabilitation of symptoms caused by PD. Clinicians and physiotherapists with expertise in the management of the disease were also involved in the design process. The game is set in a farm environment, as an ecological and natural setting may increase user involvement and a sense of calm during game play. The player is assigned an order containing a list of products to collect and to deliver from the farm. Each order represents an iteration of the game, as several orders can be executed in a row. The hand gestures required to complete the requests of the games are mostly based on the tasks for assessing PD that are coded in Section III (Motor Examination) of the Movement Disorder Society- Unified Parkinson's disease Rating Scale (MDS-UPDRS) [7]. In particular, the following tasks must be carried out to complete an order from the game:

- **Eggs Harvesting (EH)**: pick 10 eggs and put them in a basket, throughout the hand opening-closing gesture;
- **Milk Bottling (MB)**: fill 5 milk bottles by performing frontal open hand rotations;
- **Cherry Picking (CP)**: pick 10 cherries from a tree and place them in a basket, by using the finger tapping gesture of index and thumb (Figure 1). This task, compared to the previous, stimulates fine motor control of fingers;
- **Order Delivery (OD)**: drag the car along the road showed on screen to deliver the order from the farm to the city. This task is carried out by performing finger pinching and sliding the hand across the map.

At the end of the game, a report is generated that assigns a score based on the number of errors and correct actions performed. In all the tasks, visual feedback (e.g., halos surrounding objects in collision) as well as acoustic feedback

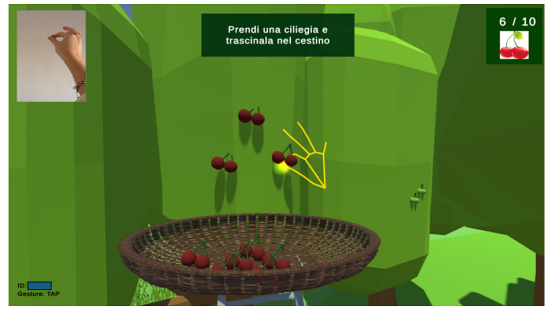

Fig. 1. The Cherry Picking task of the FarmExergame

(e.g., a noisy sound associated to wrong actions) are also present. Before playing the game, the user is provided with a tutorial training activity. The tutorial let the user practice the gestures required to complete the game, but in a different, minimal setting where the focus is specifically on learning how to interact with the system.

From a technical perspective, the game was implemented using Unity version 2022.3.24f1. The human-computer interaction is achieved through the 3D hand tracking provided by Google MediaPipe. As MediaPipe does not include a direct implementation for Unity, the tracking data are obtained by a python script running in the background, which communicates through an inter-process pipe with the game environment. The use of MediaPipe is pivotal to achieve a low-cost solution, as the model does not require GPU acceleration to run inference at real-time speed (above 30 frame per seconds). While efficient, this solutions was not previously investigated for exergaming in a PD cohort, where hand impairment is a relevant symptom. Thus, proving suitability of this hand tracking solution in this context represents a relevant secondary endeavor of the reported experiment.

Gesture recognition allows to interact with game objects according to the rehabilitative task. A simple threshold-based algorithm is employed to distinguish between gestures, using reference landmarks in the hand reconstruction (i.e., thumb and index tips landmarks for finger pinching gesture). Thresholds to recognize the gesture are defined based on an initial calibration procedure that is conducted before gameplay, where the user repeats the gesture multiple times. Considering the different profiles of impairment that PwPD may have, personalization should guarantee the highest accuracy in gesture recognition, compared for example to automatic gesture recognition models, pretrained only on healthy subjects' data.

B. Usability Assessment protocol

The experimental protocol was designed to evaluate the usability of the system, according to a standardized and repeatable process to ensure the comparability of results. While designing the protocol, particular attention was placed on addressing all the five usability attributes of Nielsen's definition.

A visual description of the protocol adopted can be seen in Figure 2. Each participant underwent two testing sessions (i.e., T1 and T2), separated by at least two weeks. This two-points evaluation was introduced to test memorability and learnability of the system. Before T1, subjects were enrolled, received explanation about the project and its objectives and signed an informed consent. All the experiments were conducted according to the Declaration of Helsinki and its amendment, with approval by the local Ethics Committee of "Istituto Auxologico Italiano" (protocol code 2020_02_18_01, approved on 28 February 2020).

Testing at T1 started with the completion of two preliminary questionnaires: a demographic questionnaire to collect basic personal data and the Cognitive Failures Questionnaire (CFQ) [8], which assesses self-perceived cognitive lapses in daily activities. The latter was collected to provide a fast and minimal estimation of possible cognitive impairment, especially in the PwPD group. Following this evaluation, the participant performed the calibration procedure for the hand gesture recognition module of the exergame and a training session under the experimenter's guidance. After the training, participants played the exergame without supervision, completing a sequence of two full orders (each from EH to OD), with the possibility of taking a short break between the repetitions. After the game play, participants were asked to complete two post-session questionnaires: the NASA TLX, and the *User Experience v1* (UX1) questionnaires. The NASA-TLX is a measure of the workload perceived during the task, and it is also employed in usability assessment to determine if a task is over-stressing and frustrating [9]. Thus, it can be considered an evaluation of efficiency and satisfaction. UX1 collected feedback on the user's experience with the system through a series of 5-points Likert questions (highest score, highest agreement with question) and provided a measure of user's satisfaction, efficiency and errors experienced during gameplay.

The second test (T2) followed the same steps as the first test, but the training phase was omitted before the gameplay. This modification allowed researchers to assess the impact of the initial training on user experience and evaluate participants' ability to recall the mechanics of the exergame without additional guidance (i.e., learnability). By comparing performance between the first and the second tests, the study also aimed to analyze the extent to which the participants had retained their familiarity with the game (i.e., memorability). Moreover, the post-gameplay questionnaires differed from those used in the first test. In this test, participants completed the NASA-TLX, an extended version of UX1, the *User Experience v2* (UX2) that included the mentioned learnability and memorability aspects. Lastly, the SUS questionnaire was administered, as it is among the most standardized and popular measures for the usability of a system.

The usability evaluations did not rely exclusively on self-assessed questionnaires but included an a-posterior analysis of quantitative performance metrics during game play such as completion times of the single tasks and players' errors, to obtain a more comprehensive and objective overview of system usability.

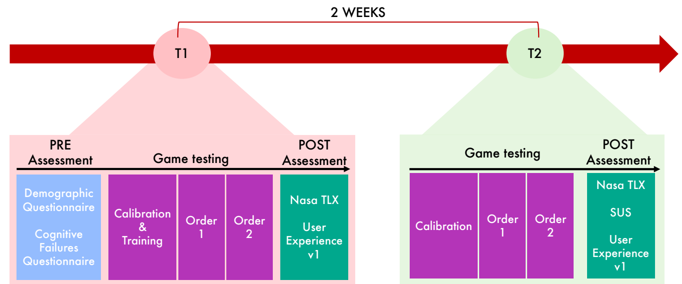

Fig. 2. Timeline of the usability experiment

C. Study participants

The usability protocol was tested on both healthy subjects (HC) and PwPD. In particular, the HC group consisted of 12 young healthy subjects (range: 24-30 years old) and 4 elderly (range: 60-67 years old), all without diagnosed cognitive impairment or motor disabilities related to hand dexterity. Due to the reduced number of older HC, the sample was considered as a unique group, without stratifying according to age. The PwPD group was composed of 9 subjects (age range: 59-81 years old) recruited among the members of a local patients' association (Associazione Amici Parkinsoniani Onlus) during their weekly session of adapted physiotherapy. Unfortunately, one subject had to drop out after T1 due to a shoulder injury, reducing the PwPD group to 8 subjects who completed the protocol. All PwPD were in the range 1-3 for the Hoehn&Yahr scale of motor impairment and performed the test while in the ON state, i.e., after having received their medication (Levodopa) for symptoms control. Stratification with respect to sex was not considered meaningful for the study but was similar among groups (HC: 4/16 females, PwPD: 2/8 female) and in line with statistics of incidence of PD. While the sample size is limited, especially for the PwPD group, five users are enough to assess usability of a system according to Nielsen's [10].

III. Results & Discussion

Regarding the pre-assessment, from CFQ, it is possible to retrieve a cognitive impairment score that ranges from 0 to 100. The HC group scored a mean of 38.8, while PwPD scored a mean of 38.7 (Table I). These scores do not highlight a strong difference in self-assessed cognitive impairment, suggesting no critical cognitive impairment in all the subjects. However some aspects must be taken into account. First, as the PwPD group is small, the mean values are not representative and an in-depth analysis shows that while no HC subject scored above 48, 4 out of 8 PwPd did. In addition, the use of a self-reported measure has inherent limitations, due to subjective biases. For instance, subjects with serious cognitive decay may not recall their own failures and elderly in general may be more reluctant than young people to report them during testing. This may be the case of 3 PwPD and of the oldest (67 years old) HC who scored a very low and somehow unexpected value (below 25). Moreover, concerning the self-reported familiarity

TABLE I
COGNITIVE FAILURE QUESTIONNAIRE AND SYSTEM USABILITY SCALE

	Mean	**Median**	**25th Perc**	**75th Perc**
CFQ_{HC}	38.8	38.5	34	42.2
CFQ_{PwPD}	38.7	44	25	50
SUS_{HC}	82.3	88.7	81.9	90.6
SUS_{PwPD}	78.2	82.5	71.25	90.6

with technological aids, PwPD scored below 3 out of 5 in the questionnaire, while HC above 4.

General observations were derived from the operators running the experimental protocol. Overall, the game presented few technical issues and was warmly welcomed by the PwPD group. While young HC subjects were more interested in the challenge provided by a new game and type of interaction (markerless hand tracking), PwPD and elderly HC demonstrated involvement and support to the introduction of new technologies in their weekly rehabilitation practice. For some PwPD, the severe impairment of hand dexterity challenged the automatic gesture recognition algorithm despite the initial calibration, but this did not prevent any of them from completing the protocol successfully and without external aid. Among the main source of errors, the shift between different gestures was challenging. In particular, some subjects did not consistently return the hand to the initial position (e.g., widening the hand in opening-closing gesture, or widening fingers in finger tapping), with resulting problems in releasing collected game objects (e.g., eggs in EH or cherries in CP). Another significant factor to report is that PwPD often had to ask for the help of their caregivers to complete the post-assessment questionnaires.

A. SUS and Satisfaction

The results of the SUS revealed a mean score of 82.3 for the control group and 78.21 for the group of PwPD, further details can be found in Table I. The SUS results for the two populations were similar, suggesting that both groups positively evaluated the usability of the exergame. The obtained scores were normalized to position the system on the SUS usability scale, which ranks systems in percentiles, indicating how a given application compares in usability against others. According to this ranking, a score of 78 on the SUS corresponds to usability rated as superior to 85% of other systems tested with this metric [11]. This suggests a positive assessment of the exergame's usability. Regarding NASA-TLX, scores ranged between 40 and 50, with higher scores for PwPD.

Scores reported place both groups in the *Some-What-High* self-assessed workload according to [12]. This result is reasonable, considering that the task should be perceived as moderately challenging since its oriented towards a rehabilitative goal. Due to small sample size of PwPD group and unbalance between groups' size, it is not possible to state significance of the differences found, but it is interesting to notice that both population gave the highest raw score to physical workload.

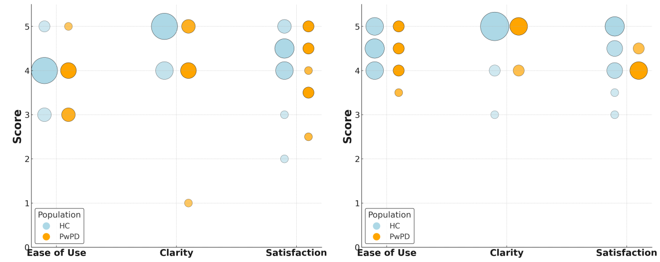

Fig. 3. Frequency of scores in UX1 (on the left) and UX2 (on the right) regarding satisfaction with the system

However, the PwPD group, who was observed to encounter physical strain in some of the gestures, felt more solicited in mental workload (higher weighted subscore). This may be related to a higher tolerance and habituation to motor impairment, that could also explain the minimum frustration scored for the PwPD group.

Regarding satisfaction measures from UX1 and UX2, questions were summarized in three main axis: clarity, easiness of use, and overall satisfaction. The results are shown in Figure 3 through bubble plots. The size of the bubbles are directly proportional to the frequency of the score on the y axis of the plot. The majority of the subjects seemed satisfied by the interaction with the game (score in the range 4-5). The clarity score increased in T2, suggesting a possible familiarization effect. General satisfaction was stable in T1 and T2, with a slight increase in T2.

B. Learnability & Memorability

Regarding learnability and memorability, some objective results can be obtained from the completion time of the game tasks by the users. The completion time between the two orders at T1 (T1_1 and T1_2) and those at T2 (T2_1,T2_1) was compared using the Raincloud plot in Figure 4. Not all the subjects were faster at completing the second order and a variety of behaviors, possibly caused also by errors during gameplay, can be observed from the connecting lines of the plot. However, on average there was a solid speed up in completing all the tasks of the second order both at T1 and at T2, suggesting a good level of learnability even for the pathological group. This mean behavior can be appreciated from the bold blue and orange lines connecting mean values of the distributions at the four time points. Due to the small sample size is difficult to compare the distribution of the two groups, but as expected the PwPD had on average higher completion times and showed a more significant effect of learning (steeper decrease) in all the tasks. This result may depend from a lower efficacy of the initial training compared to HC subjects, thus PwPD may require two or more iterations of the game itself to improve their skills with it. On the contrary, HC users' performances were quite stable, indicating that for them the training done at T1 and the completion of the first order was sufficient to achieve optimal performance. To the question: *"How do you think you have performed the second order compared to the first?"*, all subjects answered

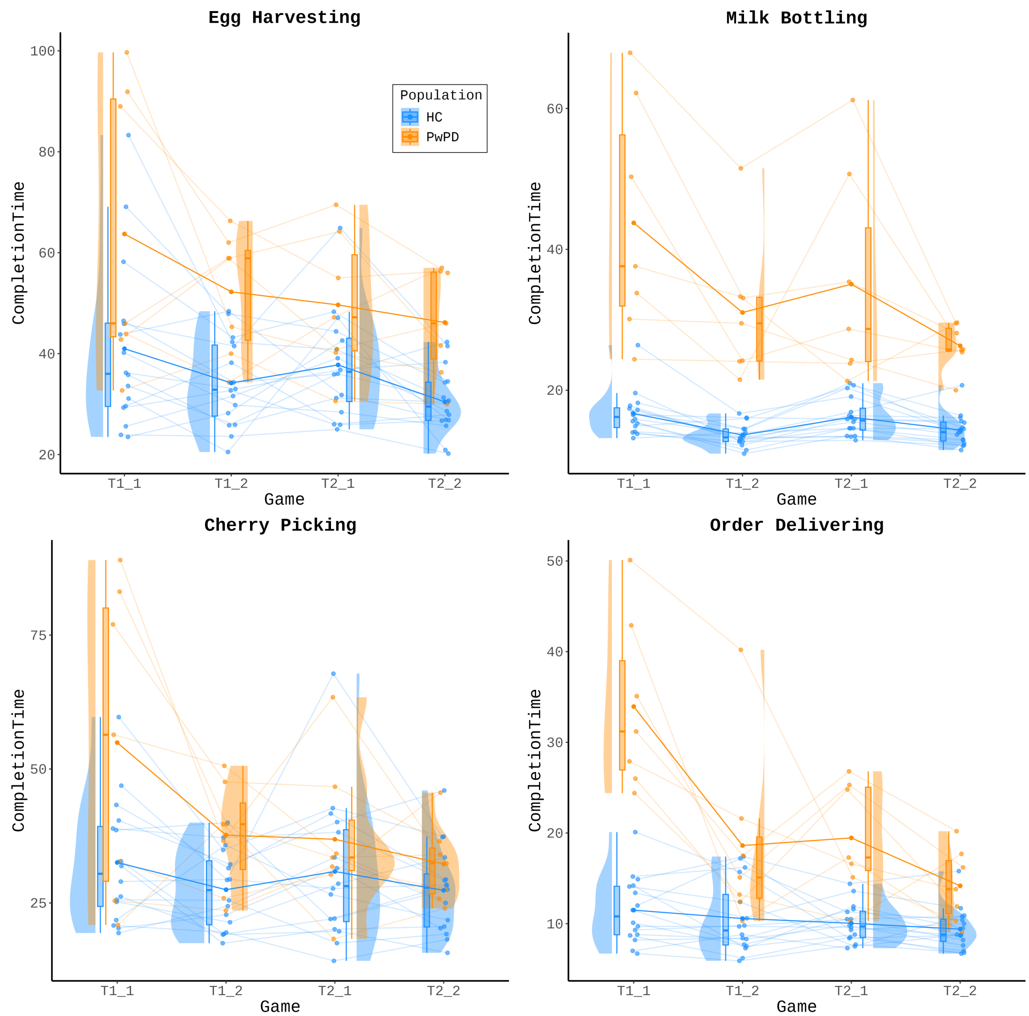

Fig. 4. Raincloud plot of completion time in seconds of Egg Harvesting (EH), Milk Bottling (MB), Cherry Picking (CP), Order Delivering (OD) games in T1 (2 orders) and T2 (2 orders). The points represent each subject performance in the 4 repetition of the games (orange for People with Parkinson Disease (PwPD) and blue for Healthy controls(HC)). The lines show the trend for each subject. Medium trends are shown by the bold lines.

positively (score above 4), both at T1_2 and at T2_2.These results are then in line with the improvement observed from the completion time, suggesting overall agreement between the self-assessment and the objective metric.

In terms of memorability, both groups and most of the subjects showed longer completion time after the two-weeks stop, but still lower at T2_1 then at T1_1. In addition, in order T2_2 they managed to go back to performance similar or lower than those achieved at T1_2. These outcomes seem to suggest that a good memorability of the FarmExergame was achieved in both groups, with an expected major recall for the HC group, also likely due to the younger mean age and their higher level of familiarity with technology. This trend was confirmed by self-assessment since in UX2, the item regarding the memory retain of T1 scored 4.7/5 for HC and 4.1/5 for PwPD. Instead, the items regarding the perception of better performance in T2 respect to T1 scored on average 3.69/5 in HC and 3.71/5 in PwPD, showing that PwPD subjects did not perceived a significant improvement compared to T1, even though their objective performance were actually better.

Finally, it is worth noticing how the usability protocol designed and analysis of the completion time through the Raincloud plot allowed for an effective and easy-to-interpret assessment of learnability and memorability in the both the control and the pathological groups. This is particularly relevant as self-assessment alone may have not provide similar results due to subjective biases and difficulty in recalling previous executions after long time gaps between sessions.

C. *Efficiency & Errors*

Analyzing the answers to UX1 and UX2, overall, the exergame was rated positively, in terms of comprehension of instruction, clarity, engagement, and will of continuity of use, for both groups of study (average score above 3.5/5). The feedback provided, both audio and visual, also received a positive score. However, the usefulness of presented textual game instructions received a rating of 3 for the PwPD, lower than all the others. In general, the results for PwPD were comparable to HC, but lower when related to the usefulness of feedback (score table in the scene, and the sound associated with picking up an object). This could indicate greater effort for PwPD subjects in performing the required gesture, resulting in a lower ability to pay attention to secondary graphic details, or could be a result of lower cognitive performance. Regarding the use of visual aids, such as the halo, both groups confirmed the usefulness of this tool in understanding how to grasp objects during play, improving spatial perception in the virtual environment.

User errors were generally sparse, making it difficult to derive quantitative conclusions. The task with the highest number of errors was CP for both groups. This result was expected for PwPD, as fine motor control is impaired in this population. For HC, the higher error rate may be attributed to limitations in the hand-tracking technology. Specifically, MediaPipe tracking of the finger-tapping gesture was found prone to errors in other studies [13], so this could have contributed to users' mistakes. Conversely, the tasks with the fewest errors were OD for HC and EH for PwPD. Subjective opinions on task complexity align well with these observations. In T1, HC participants rated MB as the easiest task and CP as the hardest. PwPD considered EH the easiest but did not consistently identify a hardest task. In T2, the easiest task remained MB for HC and EH for PwPD, while CP was reported as the hardest task by both groups. Technical issues also influenced user experience. In UX1, 4 out of 16 HC participants reported instances where their movements were performed correctly but were not classified as expected. This issue was most prevalent in CP. Similarly, in the PwPD group, gesture misclassification was highlighted by 2 out of 9 participants, primarily in OD. In UX2, gesture misclassification persisted for both groups (4/16 HC and 2/8 PwPD), again mainly in CP. Additionally, as external causes of errors unrelated to technical problems, 2 out of 16 HC and 3 out of 8 PwPD reported difficulty in executing the required hand gestures. However, as already mentioned, despite possible gesture recognition errors, all the users were able to complete the experimental sessions, demonstrating reliability of the systems even in case of failure.

IV. Conclusion and Limitations

The overall evaluation of usability was positive, with no major issue reported. Satisfactory efficiency and learnability were obtained for the PwPD involved in the study, suggesting the employability of the FarmExergame and the hand-tracking interface on a larger scale. The usability protocol seemed robust in addressing all the major aspects of usability both in HC and PwPD, who were the main challenge of the study. The systematic approach to usability provided evident results and the use of quantitative metrics of performance, along with observations from operators running the experiment proved extremely useful to address the bias introduced by self-assessed questionnaires. Moreover, the effort requested to PwPD participant to compile forms was evident to the researchers, due to difficulty in oculomotor coordination needed when using the mouse as a pointer. Thus, alternative evaluation approaches to questionnaires should become a staple in usability assessment with PwPD. Even though 8 subjects completed the study, it is clear that the robustness of the results regarding the PwPD may still be further expanded, considering the variety of cognitive and physical impairments caused by the disease. Thus, the empirical size of 5 users defined by Nielsen may not be sufficient in this scenario. Lastly, the effect of neurodegeneration on the emotional sphere is frequently overlooked, but plays a central role in user experience during exergaming, as engagement and motivation are pivotal. Technology or interaction can be overwhelming for PwPD in a way that is, at the state of the art, not clear. However, subjective measures alone showed shortcomings, as also displayed in this case study. The use of physiological signals, like heart rate variability and electrodermal activity, may provide a non-invasive and low-cost approach that will be considered in future developments.

Acknowledgment

The authors would like to thank Associazione Amici Parkinsoniani Onlus and the participants. A special mention to A.T.

References

[1] A. Fornari, M. Lanza, E. Guastafierro, A. Marcassoli, P. Sismondo, C. Curatoli, A. Raggi, and M. Leonardi, "Inequities in neurological care: Access to services, care gaps, and other barriers—a systematic review," *European Journal of Neurology*, vol. n/a, no. n/a, p. e16553. e16553 EJoN-24-1253.R2.

[2] H. Johansson, A.-K. Folkerts, I. Hammarström, E. Kalbe, and B. Leavy, "Effects of motor–cognitive training on dual-task performance in people with parkinson's disease: a systematic review and meta-analysis," *Journal of Neurology*, vol. 270, no. 6, pp. 2890–2907, 2023.

[3] W. H. Organization, *Parkinson disease: a public health approach. Technical Brief.* World Health Organization, 2022.

[4] D. L. Kappen and et al., "Older adults' physical activity and exergames: a systematic review," *International Journal of Human–Computer Interaction*, vol. 35, no. 2, pp. 140–167, 2019.

[5] B. C. Zapata, J. L. Fernández-Alemán, A. Idri, and A. Toval, "Empirical studies on usability of mhealth apps: a systematic literature review," *Journal of medical systems*, vol. 39, pp. 1–19, 2015.

[6] J. Nielsen, *Usability engineering*. Morgan Kaufmann, 1994.

[7] C. G. Goetz *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.

[8] D. E. Broadbent and et al., "The cognitive failures questionnaire (cfq) and its correlates," *British Journal of Clinical Psychology*, vol. 21, no. 1, pp. 1–16, 1982.

[9] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, pp. 904–908, Sage publications Sage CA: Los Angeles, CA, 2006.

[10] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," in *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pp. 206–213, 1993.

[11] J. Brooke, "Sus: a retrospective.," *Journal of usability studies*, vol. 8, no. 2, 2013.

[12] R. A. Grier, "How high is high? a meta-analysis of nasa-tlx global workload scores," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 59, pp. 1727–1731, Sage Publications Sage CA: Los Angeles, CA, 2015.

[13] G. Amprimo and et al., "Hand tracking for clinical applications: validation of the google mediapipe hand (gmh) and the depth-enhanced gmh-d frameworks," *Biomedical Signal Processing and Control*, vol. 96, p. 106508, 2024.