Additional file 1

Table S1 CONSORT 2025 reporting checklist

Section/topic	No	CONSORT 2025 checklist item description	Reported on page no.
Title and abstract			
Title and structured abstract	1a	Identification as a randomised trial	1
	1b	Structured summary of the trial design, methods, results, and conclusions	2-3
Open science			
Trial registration	2	Name of trial registry, identifying number (with URL) and date of registration	3, 6, 23
Protocol and statistical analysis plan	3	Where the trial protocol and statistical analysis plan can be accessed	3, 6, 23
Data sharing	4	Where and how the individual de-identified participant data (including data dictionary), statistical code and any other materials can be accessed	24
Funding and conflicts of interest	5a	Sources of funding and other support (eg, supply of drugs), and role of funders in the design, conduct, analysis and reporting of the trial	24
	5b	Financial and other conflicts of interest of the manuscript authors	24
Introduction			
Background and rationale	6	Scientific background and rationale	4-6
Objectives	7	Specific objectives related to benefits and harms	6
Methods			
Patient and public involvement	8	Details of patient or public involvement in the design, conduct and reporting of the trial	NA
Trial design	9	Description of trial design including type of trial (eg, parallel group, crossover), allocation ratio, and framework (eg, superiority, equivalence, non-inferiority, exploratory)	6
Changes to trial protocol	10	Important changes to the trial after it commenced including any outcomes or analyses that were not prespecified, with reason	NA
Trial setting	11	Settings (eg, community, hospital) and locations (eg, countries, sites) where the trial was conducted	6
Eligibility criteria	12a	Eligibility criteria for participants	6-7

Section/topic	No	CONSORT 2025 checklist item description	Page no.
Eligibility criteria	12b	If applicable, eligibility criteria for sites and for individuals delivering the interventions (eg, surgeons, physiotherapists)	10
Intervention and comparator	13	Intervention and comparator with sufficient details to allow replication. If relevant, where additional materials describing the intervention and comparator (eg, intervention manual) can be accessed	8-10
Outcomes	14	Prespecified primary and secondary outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome	10-11; File S1
Harms	15	How harms were defined and assessed (eg, systematically, non-systematically)	11
Sample size	16a	How sample size was determined, including all assumptions supporting the sample size calculation	7-8
	16b	Explanation of any interim analyses and stopping guidelines	6
Randomisation:			
Sequence generation	17a	Who generated the random allocation sequence and the method used	7
	17b	Type of randomisation and details of any restriction (eg, stratification, blocking and block size)	7
Allocation concealment mechanism	18	Mechanism used to implement the random allocation sequence (eg, central computer/telephone; sequentially numbered, opaque, sealed containers), describing any steps to conceal the sequence until interventions were assigned	8
Implementation	19	Whether the personnel who enrolled and those who assigned participants to the interventions had access to the random allocation sequence	8
Blinding	20a	Who was blinded after assignment to interventions (eg, participants, care providers, outcome assessors, data analysts)	2, 7, 10
	20b	If blinded, how blinding was achieved and description of the similarity of interventions	7-10
Statistical methods	21a	Statistical methods used to compare groups for primary and secondary outcomes, including harms	11-12
	21b	Definition of who is included in each analysis (eg, all randomised participants), and in which group	11-12
	21c	How missing data were handled in the analysis	11

Section/topic	No	CONSORT 2025 checklist item description	Page no.
Statistical methods	21d	Methods for any additional analyses (eg, subgroup and sensitivity analyses), distinguishing prespecified from post hoc	NA
Results			
Participant flow, including flow diagram	22a	For each group, the numbers of participants who were randomly assigned, received intended intervention, and were analysed for the primary outcome	12, Fig. 1
	22b	For each group, losses and exclusions after randomisation, together with reasons	12, Fig. 1
Recruitment	23a	Dates defining the periods of recruitment and follow-up for outcomes of benefits and harms	6
	23b	If relevant, why the trial ended or was stopped	NA
Intervention and comparator delivery	24a	Intervention and comparator as they were actually administered (eg, where appropriate, who delivered the intervention/comparator, how participants adhered, whether they were delivered as intended (fidelity))	17
	24b	Concomitant care received during the trial for each group	8
Baseline data	25	A table showing baseline demographic and clinical characteristics for each group	Table 1
Numbers analysed, outcomes and estimation	26	For each primary and secondary outcome, by group: ● the number of participants included in the analysis ● the number of participants with available data at the outcome time point ● result for each group, and the estimated effect size and its precision (such as 95% confidence interval) ● for binary outcomes, presentation of both absolute and relative effect size	12-17; Tables 2-3; Tables S2-S5; Fig. 3; Fig. S1
Harms	27	All harms or unintended events in each group	12
Ancillary analyses	28	Any other analyses performed, including subgroup and sensitivity analyses, distinguishing pre-specified from post hoc	NA
Discussion			
Interpretation	29	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	17-22
Limitations	30	Trial limitations, addressing sources of potential bias, imprecision, generalisability, and, if relevant, multiplicity of analyses	21

Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 Statement: updated guideline for reporting randomised trials. BMJ. 2025; 388:e081123. https://dx.doi.org/10.1136/bmj-2024-081123

Table S2 Conventional upper-limb training (CON-ULT)

Training category	Exercises and tasks	Therapeutic focus	Progression, grading and feedback strategies
Functional upper-limb task training	Reaching in multiple directions; grasping, gripping, holding and releasing; manipulation of cups, blocks and tools; sorting, picking up and placing objects; key and lock tasks; bimanual activities such as opening lids and stabilising objects.	To improve movement speed, amplitude, accuracy, coordination and task-oriented upper-limb performance; to promote trunk activation, eye–hand coordination, reaction ability, movement synchronisation and expansion of the physical workspace.	Adjust task complexity; vary object size, weight, texture and shape; increase reach distance, movement amplitude and speed; add precision demands, multi-step sequences or dual-task conditions; modify posture and environmental layout; gradually reduce verbal/visual cues; apply knowledge of performance and knowledge of results feedback; use rhythmic cueing then fade; introduce external-focus cues.
Routine-based functional actions	Handwriting practice; cutting tasks using paper or food items with utensils; eating-related movements such as using cutlery or handling cups; paper-based and container transfer tasks; simple activities of daily living.	To enhance everyday functional components, fine motor control, functional sequencing, bilateral integration and motor planning.	Increase accuracy and speed requirements; introduce smaller or more complex utensils; reduce visual cues; add rule-based modifications or multitasking demands; use external-focus cues; apply structured knowledge of performance and knowledge of results feedback with gradual fading.
Dexterity training	Small-object manipulation including coins, beads, paper clips, and buttons; card-based manipulation tasks; tool-use precision tasks; fastening and unfastening; bilateral dexterity activities; tweezers/tongs tasks; coin placement or stacking tasks.	To improve finger individuation, fine motor control, grip formation, bilateral and eye–hand coordination and refined sensorimotor processing.	Reduce object size; increase placement precision; add timed constraints; introduce multi-step manipulation; alternate hands; incorporate external-focus cues; apply knowledge of performance and knowledge of results feedback; add dual-task elements; progressively fade cues.
Sensory discrimination and	Texture matching using fabrics or textured dowels; weight discrimination with weighted items; object identification without vision including	To improve tactile and proprioceptive discrimination, sensory integration and object	Progress from visible to occluded tasks; increase similarity between textures, shapes or weights; add time pressure or greater variability; promote active sensory exploration;

stereognosis training	beans in putty or hidden familiar objects; stereognosis activities using puzzle pieces, dominoes or letters; sorting items by size or shape.	recognition without visual input.	gradually reduce tactile or verbal guidance; introduce rule-based sensory sorting.
Strength training	Grip strengthening with putty or grippers; wrist and forearm strengthening; elbow flexion and extension; shoulder strengthening using resistance bands or weights.	To improve functional strength and endurance required for grasping, lifting, holding and manipulating objects.	Increase resistance and repetitions; adjust lever arm or posture; integrate strengthening into functional tasks; apply intermittent of results feedback; introduce endurance challenges; increase movement amplitude.
Flexibility and range-of-motion exercises	Finger, wrist, and elbow stretching; forearm pronation and supination; shoulder and scapular mobility techniques.	To reduce rigidity, improve joint mobility and support smoother, more efficient movement patterns.	Increase stretch duration; combine stretching with functional reach tasks; provide corrective knowledge of performance feedback; increase precision and smoothness demands.
Coordination and movement control training	Diadochokinesis involving rapid alternating movements; finger and hand tapping; thumb opposition sequences; finger rolls; isolated finger bends and finger lifts.	To improve timing, rhythmicity, synchronisation, reaction ability, movement smoothness and interjoint coordination.	Increase movement frequency or alternating-speed patterns; add accuracy or timing targets; use rhythmic cues and fade; integrate bilateral tasks; apply summary feedback.
Cognitive movement strategy training	Plan–do–check sequence; stepwise task breakdown; use of visual markers & auditory pacing; guided attention to movement amplitude.	To support motor planning, initiation, sequencing and reduction of bradykinesia.	Fade cues systematically; increase sequencing complexity; add dual-task cognitive load; provide knowledge of performance and knowledge of results feedback on strategy use; use delayed feedback to promote autonomy.
Individual tailoring and environmental setup	Task selection based on neurological and motor examination; arrangement of environment to provide a ‘just-right challenge’; rest breaks when needed.	To ensure feasibility, safety, engagement and individualised challenge.	Modify workspace height and reach distance; adjust object characteristics; scale tasks from simple to complex; increase duration or complexity; adapt support surfaces or task direction.

Table S3 Immersive virtual reality upper-limb training (iVR-ULT)

Training module	**Exercises and tasks**	**Therapeutic focus**	**Progression, grading and feedback strategies**
Mobility and motor control training (moto module)	Guided unilateral and bilateral arm movements; bilateral mirror-arm training; reaching along visualised virtual paths; arm movements against gravity with preset directional changes; trunk activation through movement prompts; colour and shape matching using virtual objects; multisensory and rhythmic–acoustic stimulation; meteor avoidance or interception activities.	To promote passive and active mobility; initiate movement; extend range of motion; improve eye–hand coordination, general coordination, reaction ability, postural control, trunk activation, synchronised movement patterns, tone regulation and visual field exploration.	Increase trajectory complexity; reduce visual guidance; add amplitude, speed and precision demands; expand the virtual workspace; introduce additional directional changes; gradually fade rhythmic–acoustic cues; apply knowledge of performance and knowledge of results feedback.
Fine motor and grip training (finger module)	Controller-free finger and hand tracking; selective finger activation tasks; grasping, gripping, crushing, holding and placing virtual objects using pincer and power grips; unilateral and alternating bilateral tasks; mirror therapy–inspired finger exercises.	To promote selective finger movement, fine motor dexterity, grip formation and strength, bilateral hand coordination and sensorimotor refinement.	Reduce virtual object size or increase object complexity; increase accuracy or precision requirements; introduce multi-step manipulation tasks; alternate hands; increase tempo; apply external focus cues and structured knowledge of performance and knowledge of results feedback; fade cues progressively.
Task-oriented everyday skills (daily module)	Simulated everyday actions in virtual kitchen or room environments; ADL-cube; tool-use tasks; object transport, placement and balancing; multi-stage activities of daily living-relevant sequences with verbal and visual guidance; tasks requiring focusing and multitasking.	To promote everyday skills and independence; foster independent task recognition and active assistance initiation; develop action competence; coordinate fine and gross motor actions; improve sequencing, multitasking and tool use; increase efficiency of	Increase number, duration or complexity of steps; introduce multitasking or divided-attention demands; reduce visual prompts; add rule-based task requirements; increase cognitive load; provide external focus cues and fade feedback over time.

		activities of daily living performance; enhance participation and engagement.	
Gamified motor activity training (activities module)	Kite flying through virtual cloud rings; water polo–style throwing; archery.	To increase activity level, intrinsic motivation and adherence and to promote repetitive large-amplitude, goal-directed movements.	Increase game speed and difficulty; narrow scoring zones; add distractors; incorporate dual-task elements; provide intermittent knowledge of performance and knowledge of results feedback rather than continuous guidance.
Multisensory cue integration (across modules)	Real-time visual guidance; auditory rhythmic cues; pallaesthetic or haptic-like virtual reality effects.	To enhance movement timing, accuracy, synchronisation, movement smoothness and motor learning.	Gradually reduce cue dependence; shift from continuous to summary or delayed feedback; progress sensory complexity by modulating cue combinations.
Individual tailoring and safety procedures	Therapist-led module selection; real-time adjustment of task difficulty, workspace and object features; rest breaks and fatigue monitoring.	To ensure safe, feasible and individually optimised challenge.	Increase task load; adjust virtual reach distance and object characteristics; modify timing constraints and sequence complexity according to participant performance.

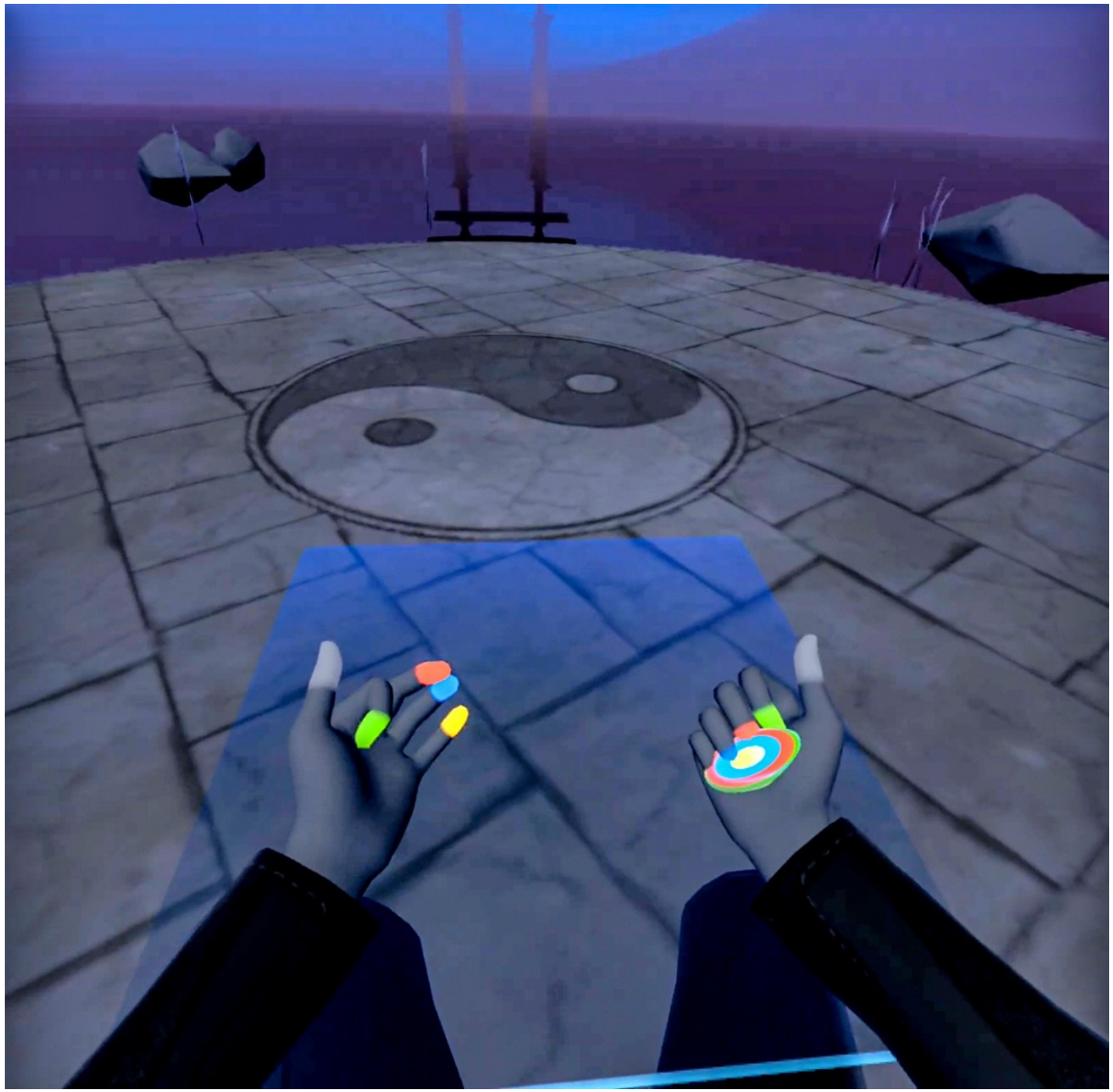

Fig. S1 Balloons game

File S1 Detailed description of assessments used in this study

Primary outcomes

The Action Research Arm Test (ARAT) is a standardised, widely used assessment of upper-limb function in individuals with neurological conditions such as Parkinson's Disease (PD) (1). It evaluates goal-directed tasks reflecting daily activities, comprising 19 items across four subscales: grasp, grip, pinch and gross movement. Each item is scored on a four-point ordinal scale (0 = unable to perform, 3 = normal performance), with tasks including lifting blocks, placing objects on shelves and manipulating small items such as marbles. Total scores range from 0 to 57, with higher scores indicating better function. The ARAT is frequently used in clinical trials to assess arm and hand function and to monitor rehabilitation outcomes over time. It has shown strong psychometric properties, yielding reliable data and demonstrating suitability for clinical application (2).

The Jebsen–Taylor Hand Function Test (JTHFT) is a standardised assessment of hand function in everyday activities (3). It comprises seven timed subtests simulating daily tasks: writing a short sentence, turning cards, picking up small objects, simulated feeding, stacking checkers, and lifting large light and heavy objects. Each task is performed with both hands, although in clinical practice the emphasis is usually on the more affected hand. Performance is measured by completion time in seconds, with shorter times indicating better function. The JTHFT does not use a rating scale but instead relies on task times to evaluate hand function, monitor rehabilitation progress and assess intervention effectiveness. It represents a reliable, valid, and respondent measure of upper hand function in individuals with PD (4).

Secondary outcomes

The Box and Block Test (BBT) is a widely used, standardised measure of gross manual dexterity (5). It assesses the ability to move small objects quickly and efficiently from one compartment of a box to another within a fixed time. The apparatus consists of a rectangular box divided into two equal compartments and 150 wooden cubes, each 2.5 cm in size. Using one hand at a time, the participant transfers as many blocks as possible, one by one, over the partition within 60 seconds. The number of blocks successfully transferred is recorded as the score for that hand. Each hand is tested separately, typically beginning with the dominant or less affected hand. The BBT provides an objective, time-efficient assessment of unilateral gross manual dexterity and is applicable across populations, including individuals with neurological conditions such as PD. The BBT has demonstrated good psychometric properties, producing robust data and proving suitable for clinical use (2).

The Nine Hole Peg Test (NHPT) was used to assess fine manual dexterity (6). Participants were instructed to place nine pegs, one at a time, into the holes of a standardised pegboard and then remove them, using one hand only. Each hand was tested separately, and completion time was recorded in seconds, with shorter times indicating better performance. The NHPT has demonstrated good reliability and validity in individuals with PD (7).

The smartphone-based Finger Tapping Test (FTT) was used to assess finger coordination and motor speed (8). Participants tapped a response button as rapidly as possible with the index finger of the tested hand for 10 seconds. The test was performed three times per hand, and the mean number

of taps was calculated. Assessments were conducted separately for each hand at baseline and post-intervention, with higher scores indicating greater motor speed and coordination. The smartphone-based FTT has been shown to be practicable, valid and reliable, and comparable to conventional methods for assessing motor dysfunction in individuals with PD (8).

The Unified Parkinsons Disease Rating Scale (UPDRS), Parts II and III, was used to assess activities of daily living (ADL) and motor function in participants with PD (9). Part II captures self-reported difficulties in daily tasks such as speech, hygiene, dressing and walking, reflecting the impact of motor symptoms on daily life. Part III is a clinician-rated motor examination assessing tremor, rigidity, bradykinesia, postural stability and gait. In the present study, assessments were conducted by trained raters at baseline and post-intervention. Higher scores on both parts indicate greater impairment. The UPDRS is widely used to monitor disease progression and treatment effects in clinical and research settings.

The Hoehn and Yahr (H&Y) scale was used to classify the severity of PD in participants (10). It categorises disease progression into five stages (I–V), with higher stages reflecting more advanced motor impairment and disability. In clinical practice and research, the H&Y is widely employed as a simple and practical measure of overall disease severity and progression. Although it provides only a global staging rather than a detailed assessment of motor or non-motor features, the scale has demonstrated good validity and reliability and remains a standard outcome measure in PD studies (11).

The Short Scale of Intrinsic Motivation (Kurzskala Intrinsischer Motivation, KIM) was used at post-intervention to assess participants' intrinsic motivation in relation to the intervention (12). It comprises 12 items rated on a 5-point Likert scale (1 = "not at all true" to 5 = "completely true"), with higher scores indicating greater intrinsic motivation. The KIM is a brief, reliable instrument that captures key aspects of intrinsic motivation, such as interest, enjoyment and internal drive, and is suitable for use in therapeutic and research contexts (Wilde et al., 2009).

The Parkinson's Disease Questionnaire-39 (PDQ-39) was used to assess health-related quality of life specific to PD (13). It is a 39-item self-report questionnaire covering eight dimensions: mobility, activities of daily living (ADL), emotional well-being, stigma, social support, cognition, communication and bodily discomfort. Scores are calculated for each dimension and as an overall summary index, with higher scores indicating greater impairment and poorer quality of life. The PDQ-39 is the most extensively tested and widely applied instrument for evaluating health-related quality of life in PD and has been recommended by the Movement Disorder Society Task Force (14).

The Smiley Likert Scale is a pictorial rating tool based on the universality of emotional expression (15, 16). In the present study, a five-point black-and-white version was used to evaluate intervention acceptability (17). Participants rated four items: (1) enjoyment of the training, (2) comfort during training, (3) satisfaction with performance, and (4) willingness to continue training at home. Evidence supports the validity of pictorial Likert scales (18, 19). A large Dutch survey (n = 7096) found that smiley scales produced responses comparable to traditional formats and were rated more positively than alternative pictorial designs such as stars or hearts (18). To minimise colour bias, black-and-white smileys were recommended and have also been applied in clinical research (20, 21).

The System Usability Scale (SUS) was used to assess usability of the CUREO® iVR therapy system (22). It is a 10-item questionnaire providing a standardised measure of perceived ease of use, satisfaction and overall user experience. Items are rated on a 5-point Likert scale (1 = “strongly disagree” to 5 = “strongly agree”), and responses are converted to a score from 0 to 100, with ≥51.7 indicating minimal and ≥71.1 good usability (23, 24). In the present study, the SUS was administered post-intervention.

References

1. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int J Rehabil Res. 1981;4(4):483–92.
2. Connell LA, Tyson SF. Clinical reality of measuring upper-limb ability in neurologic conditions: a systematic review. Arch Phys Med Rehabil. 2012;93(2):221–8.
3. Jebsen RH, Taylor N, Trieschmann RB, Trotter MJ, Howard LA. An objective and standardized test of hand function. Arch Phys Med Rehabil. 1969;50(6):311–9.
4. Galeoto G, Berardi A, Simeon R, Panuccio F, Fabbrini G, Belvisi D, et al. Psychometric Properties of Jebsen Taylor Hand Function Test in an Italian Population with Parkinson's Disease. Healthcare (Basel). 2024;12(13).
5. Cromwell F. Occupational therapist manual for basic skills assessment: primary prevocational evaluation. Pasadena, CA: Fair Oaks Printing; 1965. 29–31 p.
6. Kellor M, Frost J, Silberberg N, Iversen I, Cummings R. Hand strength and dexterity. Am J Occup Ther. 1971;25(2):77–83.
7. Earhart GM, Cavanaugh JT, Ellis T, Ford MP, Foreman KB, Dibble L. The 9-hole PEG test of upper extremity function: average values, test-retest reliability, and factors contributing to performance in people with Parkinson disease. J Neurol Phys Ther. 2011;35(4):157–63.
8. Lee CY, Kang SJ, Hong SK, Ma HI, Lee U, Kim YJ. A Validation Study of a Smartphone-Based Finger Tapping Application for Quantitative Assessment of Bradykinesia in Parkinson's Disease. PloS one. 2016;11(7):e0158852.
9. Fahn S, Elton RL, UPDRS pm. Unified Parkinson’s disease rating scale. In: Fahn S, Goldstein M, Marsden D, Calne DB, editors. Recent developments in Parkinson’s disease. 2. Florham Park, New Jersey: MacMillan Healthcare Information; 1987. p. 153–63.
10. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology. 1967;17(5):427–42.
11. Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. Mov Disord. 2004;19(9):1020–8.
12. Wilde M, Bätz K, Kovaleva A, Urhahne D. Überprüfung einer Kurzskala intrinsischer Motivation (KIM). Zeitschrift für Didaktik der Naturwissenschaften. 2009;15:31–45
13. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score. Age Ageing. 1997;26(5):353–7.
14. Martinez-Martin P, Jeukens-Visser M, Lyons KE, Rodriguez-Blazquez C, Selai C, Siderowf A, et al. Health-related quality-of-life scales in Parkinson's disease: critique and recommendations. Mov Disord. 2011;26(13):2371–80.

15. Ekman P, Friesen WV. Constants across cultures in the face and emotion. Journal of personality and social psychology. 1971;17(2):124–9.
16. Izard CE. The face of emotion. New York: Appleton-Century-Crofts; 1971.
17. Jäger R. Konstruktion einer Ratingskala mit Smilies als symbolische Marken. Diagnostica. 2004;50(1):31–8.
18. Toepoel V, Vermeeren B, Metin B. Smileys, Stars, Hearts, Buttons, Tiles or Grids: Influence of Response Format on Substantive Response, Questionnaire Experience and Response Time. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique. 2019;142(1):57–74.
19. Jäger R. Konstruktion einer Ratingskala mit Smilies als symbolische Marken. Diagnostica, Hogrefe. 2004;50(1):31–8.
20. Rosenberg F. The MoMA Alzheimer's Project: Programming and resources for making art accessible to people with Alzheimer's disease and their caregivers. Arts & Health. 2009;1(1):93–7.
21. Pérez-Sáez E, Cabrero-Montes EM, Llorente-Cano M, González-Ingelmo E. A pilot study on the impact of a pottery workshop on the well-being of people with dementia. Dementia (London). 2020;19(6):2056–72.
22. Brooke JB. SUS: A 'Quick and Dirty' Usability Scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. 1st ed. Bristol, PA: Taylor & Francis Inc.; 1996. p. 6.
23. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. 2nd ed. Cambridge, MA: Morgan Kaufmann, Elsevier; 2016. 350 p.
24. Sauro J, Lewis JR. When designing usability questionnaires, does it hurt to be positive? Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Vancouver, BC, Canada: Association for Computing Machinery; 2011. p. 2215–24. https://doi.org/10.1145/1978942.9266.

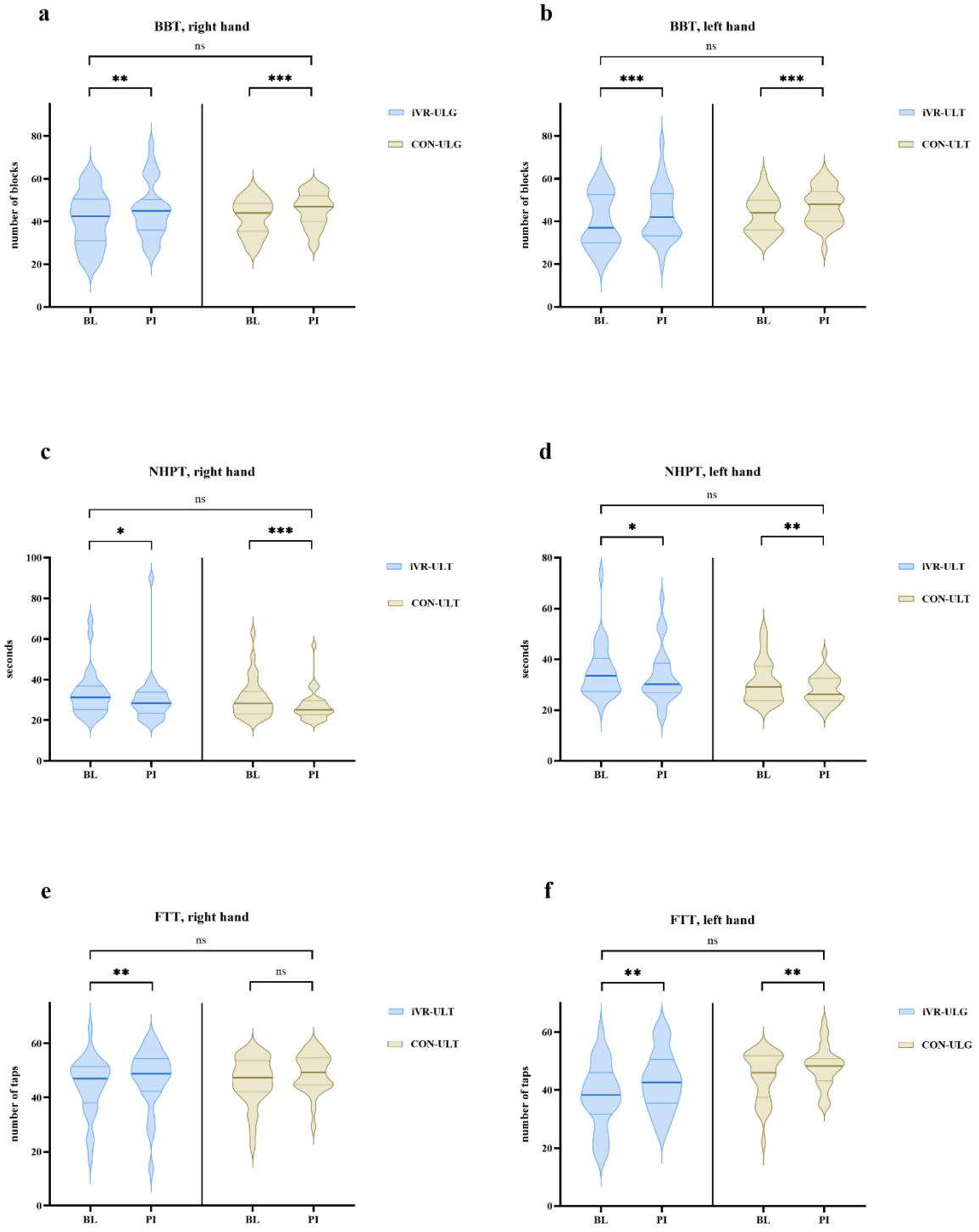

Fig. S2 Changes in gross and fine manual dexterity, finger coordination and motor speed across groups

Data distribution is illustrated using violin plots for each group and time point: **a** BBT scores for the right hand, **b** BBT scores for the left hand, **c** NHPT scores for the right hand, **d** NHPT scores for the left hand, **e** FFT für the right hand, and **f** FFT for the left hand. Data distribution is shown using violin plots. Each violin plot displays the median as a thick central line, the 25th and 75th percentiles as thin lines within the violin body, while the symmetrical "violin wings" depict the probability density of the data. BBT, Box and Block Test; higher scores

indicate better gross manual dexterity; BL, baseline; CON-ULT, conventional upper-limb training; FTT, Finger Tupping Test; higher scores indicate better finger coordination and motor speed; iVR-ULT, immersive virtual reality upper-limb training; NHPT, Nine Hole Peg Test; lower scores indicate better fine manual dexterity; ns, not significant; PI, post-intervention; *p < 0.05; **p < 0.01; ***p < 0.001

Table S4 Changes in disease severity across groups

Parameter	**iVR-ULT (n=29)**	**CON-ULT (n=29)**	**Effect size r**	**Overall p-value**
Unified Parkinson's Disease Rating Scale, Parts II & III (UPDRS II and III; score)				
UPDRS II BL[a]	13.50 (11.00 to 20.50)	11.00 (7.00 to 19.00)		
UPDRS III BL[a]	28.00 (18.00 to 38.50)	19.00 (10.00 to 37.00)		
UPDRS II PI[a]	12.00 (8.00 to 16.00)	10.00 (6.00 to 14.00)		
UPDRS III PI[a]	26.50 (15.00 to 39.00)	24.00 (10.00 to 35.00)		
Δ UPDRS II[a]	0.00 (-2.00 to 1.00) r=0.301; p=0.111	-1.00 (-4.00 to 0.00) r=0.496; p=0.009	0.104	0.432
Δ UPDRS III[a]	-0.50 (-3.50 to 0.50) r=0.369; p=0.047	0.00 (-2.00 to 0.00) r=0.260; p=0.161	0.045	0.732
Hoehn & Yahr scale (H&Y; score)				
H&Y BL[a]	2.00 (2.00 to 3.00)	2,00 (1.00 to 2.00)		
H&Y stage 1 BL[b]	7 (24.10)	11 (37.90)		
H&Y stage 2 BL[b]	12 (41.40)	11 (37.90)		
H&Y stage 2.5 BL[b]	2 (6.90)	0 (0.00)		
H&Y stage 3 BL[b]	7 (24.10)	6 (20.70)		
H&Y stage 4 BL[b]	1 (3.40)	1 (3.40)		
H&Y PI[a]	2.00 (2.00 to 3.00)	2.00 (1.00 to 3.00)		
H&Y stage 1 PI[b]	7 (24.10)	11 (37.90)		
H&Y stage 2 PI[b]	11 (37.90)	10 (34.50)		
H&Y stage 2.5 PI[b]	2 (6.90)	0 (0.00)		
H&Y stage 3 PI[b]	7 (24.10)	7 (24.10)		
H&Y stage 4 PI[b]	1 (3.40)	1 (3.40)		
Δ H&Y[a]	0.00 (0.00 to 0.00) r=0.000; p=1.000	0.00 (0.00 to 0.00) r=0.186; p=0.317	0.130	0.326

[a]Median (25^{th}-75^{th} percentiles). [b]Absolute and relative frequencies. BL, baseline; CON-ULT, conventional upper limb training; Δ, difference from baseline to post-intervention; H&Y, Hoehn & Yahr scale; lower scores indicate lower disease severity; iVR-ULT, immersive virtual reality upper-limb training; PI, post-intervention; UPDRS, Unified Parkinson's Disease Rating Scale; lower scores indicate lower disease severity

Table S5 Effect of the intervention on intrinsic motivation

Parameter	iVR-ULT (n=29)	CON-ULT (n=29)	Effect size r	p-value
Short Scale of Intrinsic Motivation Scale (KIM; score)				
Interest & enjoyment	5.00 (4.33-5.00)	4.33 (4.00-5.00)	0.244	0.066
Perceived choice	4.33 (3.67-4.67)	4.00 (3.33-4.33)	0.118	0.374
Perceived competence	4.00 (3.17-4.33)	3.33 (2.67-4.33)	0.084	0.525
Pressure / tension	4.17 (3.17-4.67)	4.00 (3.33-4.33)	0.077	0.562
Total	4.17 (3.79-4.46)	3.92 (3.50-4.33)	0.172	0.195

Data are presented as median (25th-75th percentiles). CON-ULT, conventional upper-limb training; iVR-ULT, immersive virtual reality upper-limb training; KIM, Short Scale of Intrinsic Motivation [Kurzskala Intrinsische Motivation]; higher scores indicate greater intrinsic motivation

Table S6 Changes in health-related quality of life across groups

Parameter	iVR-ULT (n=29)	CON-ULT (n=29)	Effect size r	Overall p-value
Parkinson's Disease Questionnaire-39 (PDQ39) Mobility				
Mobility BL	15.00 (5.00 to 25.00)	15.00 (7.50 to 35.00)		
Mobility PI	14.00 (5.00 to 24.50)	16.00 (7.00 to 35.00)		
Δ Mobility	0.00 (-0.50 to 0.00) r=0.260; p=0.169	0.00 (-0.50 to 0.00) r=0.235; p=0.269	0.121	0.322
PDQ39 Activities of daily living (ADL)				
ADL BL	20.83 (8.33 to 33.33)	20.83 (16.67 to 33.33)		
ADL PI	12.50 (4.17 to 29.17)	20.83 (8.33 to 29.17)		
Δ ADL	-6.25 (-16.67 to 0.00) r=0.494; p=0.009	-4.17 (-12.50 to 4.17) r=0.261; p=0.160	0.133	0.315
PDQ39 Emotional wellbeing (EWB)				
EWB BL	16.67 (8.33 to 37.50)	20.83 (12.50 to 45.83)		
EWB PI	14.58 (4.17 to 18.75)	20.83 (8.33 to 33.33)		
Δ EWB	-4.17 (-18.75 to 2.08) r=0.447; p=0.018	-4.17 (-12.5 to 0.00) r=0.406; p=0.029	0.054	0.682
PDQ39 Stigma				
Stigma BL	6.25 (0.00 to 12.50)	18.75 (0.00 to 31.25)		
Stigma PI	6.25 (0.00 to 12.50)	6.25 (0.00 to 18.75)		
Δ Stigma	0.00 (0.00 to 6.25) r=0.009; p=0.961	0.00 (-18.75 to 6.25) r=0.386; p=0.038	0.220	0.097
PDQ39 Social support				
Social support BL	0.00 (0.00 to 8.33)	0.00 (0.00 to 16.67)		
Social support PI	0.00 (0.00 to 0.00)	0.00 (0.00 to 16.67)		
Δ Social support	0.00 (0.00 to 0.00) r=0.304; p=0.107	0.00 (-8.33 to 8.33) r=0.025; p=0.895	0.110	0.408
PDQ39 Cognition				
Cognition BL	25.00 (12.50 to 37.50)	18.75 (12.50 to 37.50)		
Cognition PI	12.50 (3.13 to 28.13)	18.75 (12.50 to 31.25)		
Δ Cognition	-3.13 (-18.75 to 0.00)	-6.25 (-12.50 to 0.00)	0.074	0.692

	r=0.391; p=0.039	r=0.212; p=0.254		
PDQ39 Communication				
Communication BL	16.67 (0.00 to 25.00)	16.67 (0.00 to 25.00)		
Communication PI	8.33 (0.00 to 25.00)	8.33 (0.00 to 16.67)		
Δ Communication	0.00 (-8.33 to 0.00) r=0.356; p=0.060	0.00 (-8.33 to 0.00) r=0.131; p=0.480	0.087	0.512
PDQ39 Bodily discomfort				
Bodily discomfort BL	25.00 (16.67 to 50.00)	25.00 (16.67 to 50.00)		
Bodily discomfort PI	20.83 (8.33 to 37.50)	25.00 (8.33 to 33.33)		
Δ Bodily discomfort	-4.17 (-16.67 to 0.00) r=0.310; p=0.100	-8.33 (-16.67 to 8.33) r=0.313; p=0.092	0.001	0.994

Data are presented as median (25th-75th percentiles). BL, baseline; CON-ULT, conventional upper-limb training; Δ, difference from baseline to post-intervention; iVR-ULT, immersive virtual reality upper-limb training; PDQ-39, Parkinson's Disease Questionnaire-39; lower scores indicate better health-related quality of life; PI, post-intervention

Table S7 Intervention acceptability

Parameter	**iVR-ULT (n=29)**	**CON-ULT (n=29)**	**p-value**
Smiley Likert Scale (score)			
How did you like the training	1 (1-2)	1 (1-2)	0.479
How comfortable during the training	1 (1-2)	1 (1-2)	0.397
How satisfied with your performance	2 (1-2)	2 (1-2)	0.166
Like to continue the training at home	2 (1-2)	1 (1-2)	0.243

Data are presented as median (25th-75th percentiles). CON-ULT, conventional upper-limb training; iVR-ULT, immersive virtual reality upper-limb training; lower scores indicate higher acceptability

Table S8 Spearman's rank correlations between participants' baseline characteristics and changes in primary and relevant secondary outcomes

	Age	Disease duration	LEDD	UPDRS Part II	UPDRS Part III	H&Y stage
ARAT Grasp right	0,283	-0,119	0,022	0,318	0,284	0,090
ARAT Grasp left	0,130	0,006	0,055	0,263	0,152	0,055
ARAT Grip right	0,165	-0,065	0,145	0,353	0,322	0,211
ARAT Grip left	0,258	-0,038	0,026	0,513	0,454	0,178
ARAT Pinch right	0,032	0,169	0,167	0,154	0,129	0,012
ARAT Pinch left	0,065	0,002	0,187	0,376	0,266	0,002
ARAT Gross movement right	0,068	0,009	0,180	0,017	0,070	0,230
ARAT Gross movement left	0,083	0,005	0,197	-0,015	0,039	0,195
ARAT Total right	0,238	0,130	0,184	0,270	0,219	0,153

ARAT Total left	0,099	-0,032	0,097	0,380	0,261	-0,031
JTHFT right	0,022	-0,082	-0,095	-0,132	-0,172	0,056
JTHFT left	0,166	-0,047	-0,223	-0,095	-0,053	0,140
BBT right	-0,097	-0,071	0,048	-0,194	-0,178	-0,181
BBT left	-0,263	-0,137	0,113	-0,003	-0,002	-0,098
NHPT right	-0,217	-0,103	-0,309	-0,305	-0,271	-0,223
NHPT left	-0,136	0,036	-0,057	-0,139	-0,133	-0,032
FTT right	-0,091	-0,033	-0,034	0,070	0,163	-0,093
FTT left	-0,110	-0,135	-0,024	0,214	0,233	-0,022

Age and disease duration are reported in years; ARAT, Action Research Arm Test; BBT, Box and Block Test; FTT, Finger Tapping Test; H&Y, Hoehn and Yahr scale; LEDD, levodopa equivalent daily dose; NHPT, Nine Hole Peg Test UPDRS, Unified Parkinson's Disease Rating Scale, higher scores indicate greater impairment

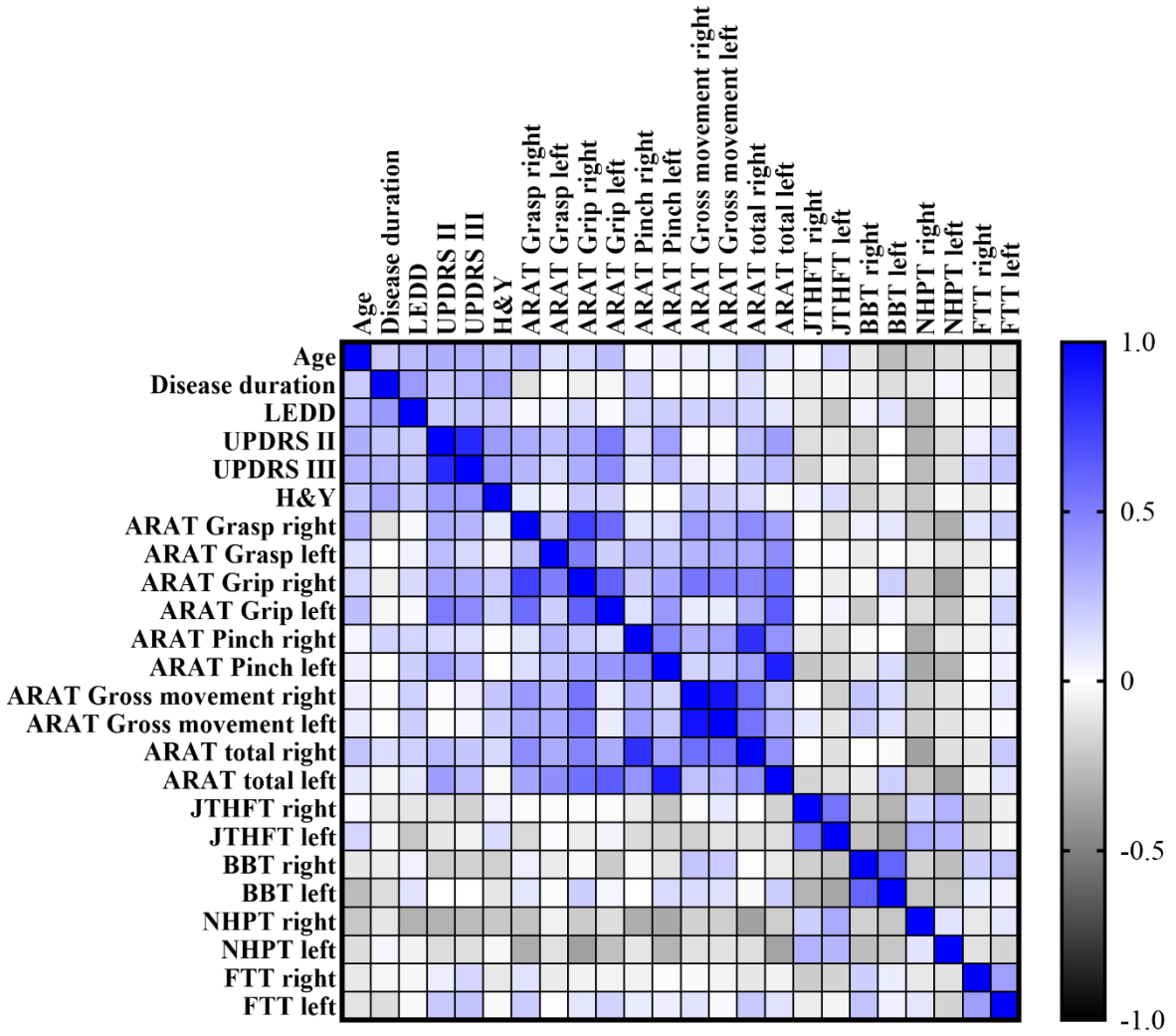

Fig. S3 Heatmap of Spearman's rank correlations between participants' baseline characteristics and changes in primary and relevant secondary outcomes