Int. J. Human–Computer Studies 184 (2024) 103206

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Adaptative computerized cognitive training decreases mental workload during working memory precision task - A preliminary fNIRS study

Aleksandra Landowska [a,*], Max L. Wilson [a], Michael P. Craven [b,c,d], Kyle Harrington [c,d]

[a] School of Computer Science, University of Nottingham, United Kingdom
[b] NIHR Nottingham Biomedical Research Centre, University of Nottingham, United Kingdom
[c] NIHR MindTech MedTech Co-operative, Institute of Mental Health, University of Nottingham, United Kingdom
[d] Human Factors Research Group, Faculty of Engineering, University of Nottingham, United Kingdom

ARTICLE INFO

Keywords:
Mental workload
fNIRS
Computerised cognitive training
HCI
Augmented reality
Prefrontal cortex
Working memory

ABSTRACT

With the growing concern for the health of ageing populations, much research continues to look at the impact of cognitive training, particularly in relation to cognitive decline. We sought to use novel techniques, including augmented reality and portable neurotechnology, to evaluate the impact of a dynamically adjusting cognitive training programme, in comparison to a statically challenging alternative. Before and after an 8-week training period, and at a 5-week follow-up, we used portable functional Near Infrared Spectroscopy to examine mental workload in a mixed battery of cognitive and transfer tasks. A recently developed tablet-based task was used to identify changes in cognitive misbinding. Augmented Reality was used to create a supermarket shopping experience, as a more ecologically valid and realistic transfer task relating to an everyday task relating to independence that quickly becomes difficult with cognitive decline. The analyses showed a decreased mental workload within the dorsolateral prefrontal cortex and that participants considerably increased their performance in the trained task. Some results were maintained at the 5-week follow-up assessment. In terms of transfer, we observed reliable group differences immediately after training completion, which were mainly driven by distinct conditions. Some behavioural memory gains were maintained during the follow-up. The use of novel technologies brought new insights into the effects produced by the dynamic computerised cognitive training programme, which has potential future applications in cognitive decline screening and prevention.

1. Introduction

In the past decade cognitive training (CT) and assessment became very popular. It is estimated that the CT revenue market will be worth US$ 2.04 billion in 2022 and is forecasted to be worth US$ 9.4 Billion by 2032[1]. CT is a programme of structured, regular, repeated, guided, cognitive activities designed to maintain or improve cognitive functioning, and many companies claim that their CT intervention might improve cognitive functions and therefore the overall quality of life. They are designed to train either a single cognitive domain, for example, working memory or attention (Flak et al., 2019; Gagnon and Belleville, 2012), or as combined programmes to train multiple domains (Mahncke et al., 2019; Rizkalla, 2018; Sukontapol et al., 2018). However, despite its growing popularity, evidence regarding the effectiveness of cognitive training remains conflicting. In particular, there is a lack of objective evidence of CT's success, and empirical methods that are able to evaluate its impact, particularly in settings that resemble real-world scenarios. Brain imaging technologies, however, offer a promising method to fill this knowledge gap, particularly more recent and portable neuroimaging techniquesuch as functional Near Infrared Spectroscopy (fNIRS), that can be applied in more natural, ecologically valid conditions. The goal of this work was to assess CT using fNIRS as a more objective measure and use augmented reality to assess task success in safe but more ecologically valid conditions. We state these goals in more formal terms after the related work below. Our work makes three contributions:

1. we demonstrate, using a recently validated novel memory precision task, that Computerised CT can have a sustained impact working memory after 8 weeks of training.

* Corresponding author.
E-mail address: aleksandra.landowska@nottingham.ac.uk (A. Landowska).

[1] https://www.prnewswire.com/news-releases/global-market-for-cognitive-assessment-training-to-reach-us-9-4-billion-by-2032-factmr-reports-301488500.html

https://doi.org/10.1016/j.ijhcs.2023.103206
Received 22 November 2022; Received in revised form 1 December 2023; Accepted 21 December 2023
Available online 28 December 2023

2. we demonstrate that fNIRS can detect significant differences in prefrontal cortex activity, including at the follow-up, at some medium task-difficulty levels.
3. we demonstrate the use of augmented reality to examine whether these effects transfer to a more ecologically valid task condition including follow-up.

2. Related work

Computerised cognitive training (CCT) is a more recent extension of the traditional pen and paper CT, using digital technologies to deliver cognitive interventions (Kueider et al., 2012). CCT often takes an approach to deliver treatment in the form of gamification. The advantage of gamified brain training is that it implements more entertaining tasks that might enhance motivation and engagement. Difficulty and content can be scaled through algorithms that allow games to be challenging and interesting (Khaleghi et al., 2021). Moreover, it allows for more objective real-time data collection using electronic devices (Basak et al., 2008).

CCT might potentially improve cognitive performance or delay the onset of cognitive decline, with companies and research typically focusing on cognitive reserve and neuroplasticity (Soldan et al., 2017). Cognitive reserve is defined as the brain's resilience to neural damage and ability to compensate for the loss of neural function due to ageing or injury (Soldan et al., 2017; Stern, 2002, 2012). Neuroplasticity is closely related to cognitive reserve, and it refers to the ability of neurons and neural networks within the brain to grow, modify, and adapt in response to experience, training, or new information (Fuchs and Flügge, 2014). Therefore, the efficacy of cognitive training should be reflected in changes in brain activity (ten Brinke et al., 2017). Employing neuroimaging techniques could provide a better insight into how CCT affects brain function and therefore deliver scientific validation of its potential efficacy. Previous studies have used fMRI to assess the effects of CCT (Kim et al., 2017; Lampit et al., 2015; Li et al., 2019; Rosen et al., 2011), as well as EEG (Fabio et al., 2016; Gandelman-Marton et al., 2017; Olfers and Band, 2018). Comparatively, Functional Near-Infrared Spectroscopy (fNIRS) was previously used only once to assess the efficacy of cognitive training in major depressive disorder (Payzieva and Maxmudova, 2014). fNIRS is a neuroimaging method which uses near-infrared light to measure changes in brain activity (Boas et al., 2014) and the advantage of fNIRS over other neuroimaging abilities is that it allows more natural movement while tolerating more movement artefacts that would create signal noise (Nieuwhof et al., 2012; Piper et al., 2014). fNIRS is widely used in cognitive neuroscience to understand neural substrates of human behaviour (Pinti et al., 2020) and in a clinical setting to measure brain activity in a range of mental disorders (Rahman et al., 2020). Recently fNIRS has become more popular also in neuroergonomics (Curtin and Ayaz, 2018) and human-computer interaction (HCI) to acquire better insight into users' cognitive state during natural interactions with technology (Maior et al., 2015; Solovey et al., 2009). Recent HCI studies employed fNIRS in research on usability testing for interfaces (Hill and Bohil, 2016; Hirshfield et al., 2009; L.M. 2011), web layouts (Bhatt et al., 2018; Lukanov et al., 2016), to measure suspicion towards online misinformation (Hirshfield et al., 2019), while dealing with mobile advertising (Mancini et al., 2022), or in gaming performance (Aksoy et al., 2019; Andreu-Perez et al., 2021; Kanatschnig et al., 2021). Because fNIRS probes are designed to be placed directly upon a user's scalp, often targeting the prefrontal cortex (PFC), it can be used to measure mental workload in more natural settings during interaction with technology (Causse et al., 2017). fNIRS has been previously used to measure mental workload in HCI studies in office workers in real working conditions (Midha et al., 2021), in digital manufacturing environments (Argyle et al., 2021) and in VR serious gaming for learning (Aksoy et al., 2019).

Mental workload is described as the level of mental resources required to meet a task's demands (van Acker et al., 2018; Wickens, 2008). Mental workload has an impact on performance and therefore deteriorated performance can be caused by a suboptimal workload – either overload or underload (Young et al., 2015). Overload occurs when there are insufficient resources to perform a task and therefore leading to a decrease in performance, attentional lapses, and errors. On the other hand, underload might lead to disengagement and boredom, and therefore impact on performance as well (Young et al., 2015; Young and Stanton, 2002). Previous literature demonstrated that mental workload and performance are correlated with brain activity. This is because task-related brain activity requires a certain amount of allocated mental resources which are limited and proportional to task difficulty (Wickens, 2008). In particular, activity within the PFC has been shown to be a function of mental workload (Fishburn and Norr, 2014) using fMRI (Causse et al., 2022; Lim et al., 2010 and EEG (Berka et al., 2007; Kutafina et al., 2021; Qu et al., 2020; So et al., 2017). Ayaz et al. (2012) demonstrated that fNIRS can measure mental workload changes in the dorsolateral prefrontal cortex (DLPFC) of operators in naturalistic settings. The study employed the n-back task as a baseline condition to simulated activities of air traffic controllers. The result demonstrated that activity within the DLPFC increased as the difficulty of the task increased, correlated with behavioural performance and NASA TLX. Subsequently, the study denoted that with expertise, oxygenation within the PFC decreases. Decreased neural activity is linked to increased neural efficiency correlated with improvement in task performance (Kelly et al., 2006).

The efficacy of CCT depends on improved neural efficiency, leading to either increased task performance, decreased workload or both. Therefore, a reduction in workload should entail more availability of cognitive resources, and subsequently, the ability to perform more complex tasks without error. The transfer effect occurs if skills and knowledge acquired during the training phase can be applied in different situations to different goals in real-life scenarios (Kelly et al., 2014). There are inconsistent results regarding CCT efficacy and real-life transfer. While the majority of studies showed improved performance on the trained task (near-transfer)(Luis-Ruiz et al., 2020; Tetlow and Edwards, 2017; Zhang et al., 2019), meta-analyses on the efficacy of CCT demonstrated the absence of consistent evidence for general cognitive improvement and transfer to the real-life situation (Hu et al., 2021; Kelly et al., 2014; Sala et al., 2019; Sala and Gobet, 2019; Vermeir et al., 2020). However, these results might be caused by the fact that many current studies often do not employ ecologically valid real-world transfer assessment methods. Therefore, developing better and more innovative approaches could improve measuring training-related transfer effects. Gonneaud et al. (2014) proposed that CT should also include 'training for transfer', delivered using immersive technologies that allow for practising trained skills in ecologically valid simulations of real-life situations. Immersive technologies offer to bridge a gap between experimental control and naturalness of the response providing an opportunity to assess the effects of CCT in ecologically valid situations.

Employing brain imaging methods could provide a valuable more objective tool for assessing the efficacy and progress of CT. The advantage of using portable neuroimaging methodsuch as fNIRS in CT studies is its ability to measure the cortical hemodynamics associated with brain activity non-invasively. fNIRS could provide insights into the functioning of the brain before and after CT interventions, allowing researchers to observe changes in brain activity patterns, which can be indicative of the effectiveness of the training. This could potentially benefit the field of CT, as it adds additional information beyond subjective reports or behavioural assessments. Previous studies employed fNIRS to measure the effects of CT in healthy individuals (Acevedo et al., 2022; Ge et al., 2021) or individuals with MCI (Vermeij et al., 2017). However, so far there are no studies that employed fNIRS to assess the efficacy of neurogames longitudinally using working memory training programmes and ecologically valid transfer tasks.

The aim of the project was two-fold. First, we aimed to evaluate the relative impact of a personalised CCT training programme, provided by our industry partner Brain+[2], in comparison to a standardised set of brain training activities. Second, we aimed to use new technologies to provide deeper insights into the effects of brain training. In particular, using wireless fNIRS, this study examined the efficacy of the training programme by measuring mental workload in the prefrontal cortex (PFC). The portability of fNIRS, in comparison to other neurotechnology, is that brain measurements can be taken both during seated test batteries, and more ecologically valid and immersive experiences. To achieve the goal of examining the effect of the CCT training in more ecologically valid everyday tasks, we created a memory-based shopping task in augmented reality, where the mental workload used during this task was also evaluated using fNIRS.

3. Methods

To meet our aims, we designed a longitudinal study, involving participants using the provided brain training app for 8 weeks. We took performance and brain-activity measurements from participants before training, after training, and 5 weeks after training was stopped. The study had the following hypotheses:

H1: There will be a significant change in brain activity within the PFC during the working memory task Starry Night after the ADDP training as measured by fNIRS
H2: There will be a significant change in brain activity within the PFC during ecologically valid shopping task after the ADDP training as measured by fNIRS
H3: Performance will increase for the cognitive working memory task after the ADDP training as measured via AD test battery Stary Night
H4: Performance will increase for the transfer task after the ADDP training as measured via an ecologically valid shopping task
H5: Participants will recognize the impact of ADDP training and have generally positive attitudes toward the technology after participating in the programme

Within the study, participants were divided into two groups: 1) active experimental group and 2) passive/control experimental group. Both groups received one of the Brain+ intervention programmes. The difficulty parameters in the active experimental version were dynamically adjusted to the individual's performance over the 8 weeks of using CCT. The difficulty parameters in the passive version were set at one constant level for the same period. This group served as a placebo condition to contrast the effect of the ADDP intervention. They received exactly the same set of games, however, the difficulty level was capped at the low level. All participants were assessed before the intervention using The Montreal Cognitive Assessment (MoCa) (Nasreddine et al., 2005) to determine baseline cognitive abilities. Repeated measures were conducted using a specially adapted version (for neuroscience block design studies) of the AD detection test battery Starry Night (Pertzov et al., 2012, 2013) and an ecologically valid augmented reality task in a 3D environment (using Magic Leap), each described further below. Ethical approval CS–2019-R32 was granted by the University of Nottingham School of Computer Science Ethics Committee. Written informed consent was obtained from all participants included in this study.

3.1. Participants

This study recruited older adults (aged 45+) who were concerned that their memory might be declining. For ethical reasons, we did not involve participants with a clinical diagnosis of dementia or Alzheimer's disease, as the research team did not include medical or care professionals, the study required an untested level of technical and practical competence, and the study context could provide false hope on a technology that was not yet proven. Further, recruitment for our participants was clear that we did not have the expertise to diagnose such conditions; during the study, if any participant described memory concerns to the researchers, we reiterated that we did not have the clinical expertise to diagnose participants and recommended that they speak to their doctor if they were concerned. Moreover, the inclusions of dementia patients might not be appropriate because the study was looking at preventative measures and due to the heterogeneous nature of dementia-related cognitive decline it would be difficult to compare groups. We discuss this further in future work.

One hundred and five ($N = 105$) participants (53 females and 52 males (zero declared an alternative gender or declined to say), mean age $M = 54.54$, $SD = 8.13$) were recruited via Join Dementia Research, a community of members of the public interested in supporting research relating to Dementia, and via a recruitment firm Patricia Turner. Nineteen participants dropped out from the study due to low training adherence or personal reasons, one was excluded because of poor data quality, and eighty-five participants completed the study.

3.2. Experimental design

The experiment employed a between-subject design. Participants were assigned to two groups – experimental or control. The assignment was not random and it wasingle-blinded only due to COVID-19 restrictions. The experimental group was recruited and took part first, and then when finished, recruitment began for the control group. Both groups received the ADDP app during their training phase, however, the apps differed in training parameters:

1. Active experimental group ($N = 42$) – received the adaptative version of the training program
2. Active control group ($N = 43$) – received the non-adaptive version of the training program with constant difficulty

Both groups participated in 3 data collections with the researcher, with CCT taking place between the pre- and post-training sessions:

1. Pre-training baseline assessments using several instruments and measures:
   - The Montreal Cognitive Assessment (MoCa)
   - Cognitive assessment: Stroop Test, digit span
   - the fNIRS-adapted AD detection test battery Starry Night
   - an ecologically valid augmented reality task in 3D environment (using Magic Leap)
   - Augmented Reality Immersion Questionnaire (ARI)
   - Cybersickness Questionnaire

   Training phase:

   - Both groups received one of the Brain+ training programmes to play at home on their own mobile device. Participants were asked to play 4 neurogames (5 min each), 20 min a day, 5 days a week for 8 weeks.
1. Post-training assessment session:
   - Cognitive assessment: Stroop Test, digit span
   - the fNIRS-adapted AD detection test battery Starry Night
   - an ecologically valid augmented reality task in a 3D environment (using Magic Leap)
   - Short feedback on the user's perception of the ADDP solution
2. 1-month follow-up
   - Cognitive assessment: Stroop Test, digit span
   - the fNIRS-adapted AD detection test battery Starry Night

[2] https://www.brain-plus.com/

- an ecologically valid augmented reality task in 3D environment (using Magic Leap)

Participants' other normal daily activities were not controlled during the study, where e.g. we did not ask people to start or stop playing Sudoku. Participants were asked, however, to use the CCT training consistently during the training phase, and then stop using the CCT training app between post-training and follow-up sessions. Correct participation and minimum required training levels were checked using server logs created by participants using the app. Participants were remunerated with £30 for participating in the first session, £20 for the post-training session, and £50 if reaching and participating in the final follow-up session. Participants were given free access to the commercial version of the CCT app if they wished to restart using it.

3.3. CCT training app (used between pre- and post-training sessions)

The training phase used the ADDP multi-modal cognitive training application developed by Brain+ to exercise concentration, memory, attention and planning. The training consisted of 4 different games:

1. Memory lane (Fig. 1a) to train attention and working memory. Participants moved a ball from side to side, whilst it rolled at speed down a tunnel, using their finger on a mobile device. The user was instructed to memorise sequences of between 1 – 4 symbols that appeared on the screen. The task involved moving the ball to collect the correct symbols, one by one, as they appeared in the tunnel, while also avoiding gaps in the tunnel. The interaction involved maintaining continuous finger contact with the screen, and moving the finger left and right to move the ball left and right as desired. This involved developing reaction times to avoid obstacles (aspeed increased), whilst remembering an increasing number of symbols. For the control group, the number of objects to remember and then collect was capped at a maximum of 2 objects.
2. Ocean of attention (Fig. 1b) to train attention and working memory. Participants were presented with a number of different objects (flowers, shells, insects) at different locations. The task involved memorising and recalling the objects after a short delay. The user then had to guide the character to the target location, either collecting or avoiding those objects. The interaction involved tapping on which square the user wanted the character to go to next, from their current square. The control group received a fixed high-exposure duration of items to memorise at 1.5 s, whereas a staircase algorithm adapted to the individual level for the experimental group.
3. Cocktail party (Fig. 1c) to train working memory. Participants were presented with information (name, age, occupation, country of birth) about characters on a virtual party. The task involved memorising and recalling the information about the characters when prompted with options. The control group had a cap on the number of associations to a person that should be remembered, and the time it was required before recall.
4. Path finder (Fig. 1d) to train planning skills. Participants were presented with the target location on the board, and a set of obstacles that would affect their position, such as turn them around or shift them to a different square. The task involved planning a sequence of moves to reach the target destination while on the board. Participants had to plan their moves using a limited number of moves (move forward, move backward, turn left, turn right), and predict where they would end up given the impact of the obstacles. In the control version of the game, participants only had to plan their moves one step at a time, whereas, in the experimental group, the complexity and size of the path increased according to a staircase routine.

In summary, the control group had what would be colloquially described as "easy" versions of the tasks, whilst the experimental group had versions that dynamically adjusted to challenge the capability of the user. The custom builds of this training app (adaptive and non-adaptive) were kept consistent from the beginning of the data collection so that the commercially available app's normal updates and other features did not change during data collection.

3.4. Assessment stimuli (using during sessions)

3.4.1. Starry night

The Starry Night AD detection test battery is a working memory game developed by Brain+, based upon the continuous working memory test- "What was where", which was proposed by Pertzov et al. (2012, 2013). The test builds on an extensive body of literature on continuous-report measures of working memory performance (Bays et al., 2009; Bays and Husain, 2008; Fougnie et al., 2010; Ma et al., 2014; Rose et al., 2016; Wolff et al., 2017; Zokaei and Husain, 2019). Continuous measures of working memory, unlike binary measures, examine the resolution in which items are retained in memory. Additionally, unlike many traditional memory tasks commonly used, these

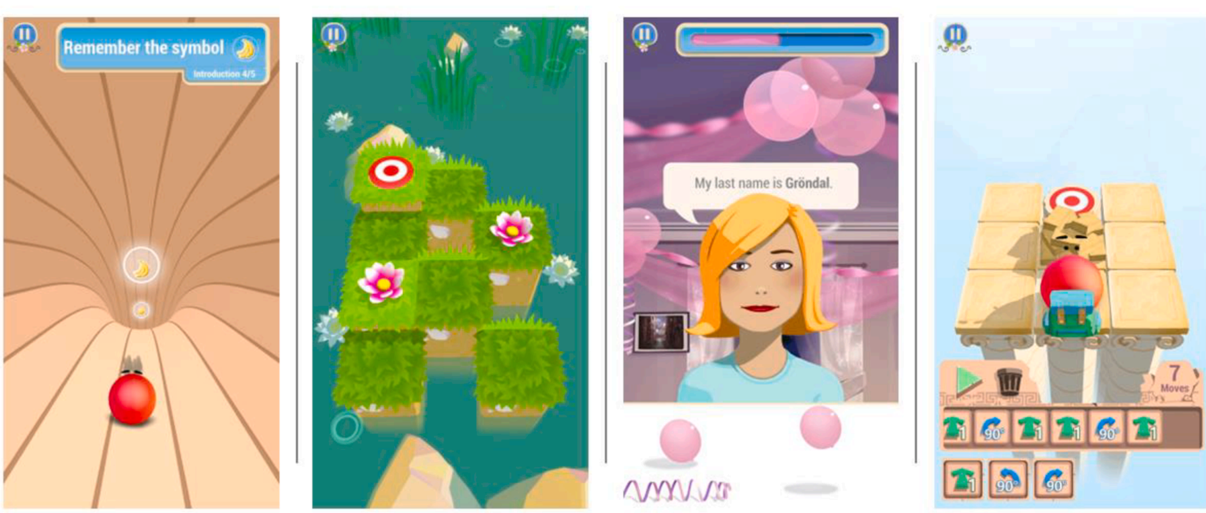

Fig. 1. Brain+ cognitive training app used during the training phase. The training consisted of 4 different games to train attention, memory and planning. Participants were asked to play 20 min a day, 5 days a week for 8 weeks.

tasks provide a means to dissect the different sources of error that can contribute to deficits in performance. Error can be separately attributed to the degradation of memory precision, failures in encoding or retaining an item (guesses), and incorrect binding of different features memoranda (e.g. swapping the identity and location of items). Therefore, the output of the task is far richer than simply indexing whether a participant remembered an item or not. Critically, precision measures of the recall have been shown to be more sensitive than binary measures of performance (Zokaei et al., 2015) and have proven successful in quantifying the reliability and quality of working memory in healthy ageing, neurodegenerative disorders and even in at-risk populations (Grogan et al., 2016; Liang et al., 2016; Pavisic et al., 2021; Peich et al., 2013; Rolinski et al., 2015; Zokaei et al., 2014, 2015; N. 2021; Zokaei and Husain, 2019). The paradigm aims to assess the ability to bind item features (for example the identity of the object and its location) in working memory, where prior work has demonstrated that miss-binding these pieces of information is an early and reliable indicator of cognitive decline.

In the Starry Night game (Fig. 2), participants are presented with either 1 or 3 different abstract star constellations at different random locations. The task is to maintain the constellations and their locations in working memory. Then, a blank screen is displayed for 1- or 4-second duration, and then the test array appears, containing 1 correct constellation from the original set and 1 non-target item (foil). The task is based on the previous research (Pertzov et al., 2012; Tabi et al., 2019, 2020; Tabi et al., 2021; Zokaei et al., 2014, 2015; 2019). The interference task was added as a novel approach to manipulate both difficulties and examine the impact of attentional load on memory performance. This was done specifically to make sure the challenge levels were suitable for healthy participants and to avoid ceiling effects, specifically following cognitive training. Participants were asked to drag the target constellation into its original location. Two of the conditions involved an additional distraction task (interference task), which requires participants to track 1 to 3 stellar objects on the screen by pointing a 'telescope' at them with their finger. The difficulty of the task is determined by the number of items held in working memory, delay duration, and interference task. The custom version of this game, designed for block-design fNIRS user studies, consisted of 6 blocks (conditions) with 40 -seconds "on" and 10-second "off" periods alternating in pseudo-random order:

- One Target Short Delay (1TSD): Involving 1 target constellation with a 1-second exposure followed by a 1-second delay.
- Three Targets Short Delay (3TSD): Involving 3 target constellations with a 3-second exposure followed by a 1-second delay.
- One Target Long Delay (1TLD): Involving 1 target constellation with a 1-second exposure followed by a 4-second delay.
- One Target with Interference (1TI): Involving 1 target constellation with a 1-second exposure during a 4-second interference task.
- Three Targets Long Delay (3TLD): Involving 3 target constellations with a 3-second exposure followed by a 4-second delay.
- Three Targets with Interference (3TI): Involving 3 target constellations with a 3-second exposure during a 4-second interference task.

The working memory precision was measured as: a) working memory accuracy in identifying the proportion of correct targets, b) working memory localisation performance in remembering the distance of where the user drags it to from its actual original location (localization errors in pixels) measured only on trails when the target was correctly identified, and c) misbinding (the distance to the nearest wrong target in pixels) (Tabi et al., 2020; Tabi et al., 2021, 2022). The test was optimized for and delivered through the Samsung Galaxy S3 Tablet, rather than using different sized devices, in order to maintain consistency of formal measurements for scientific rigour. This task took approximately 15 min to complete (including practice and tutorial).

3.4.2. Augmented reality shopping task

The shopping task (Fig. 3), developed using Unity 2020.2.2 game engine, consisted of a three-aisled virtual store with various items placed on shelves. The shop was located within a 5Mx3M space in the Mixed Reality Lab at the University of Nottingham. There were three different sections in the shop – food, fruits and vegetables, and household. Participants were presented with different shopping lists for 10 s to

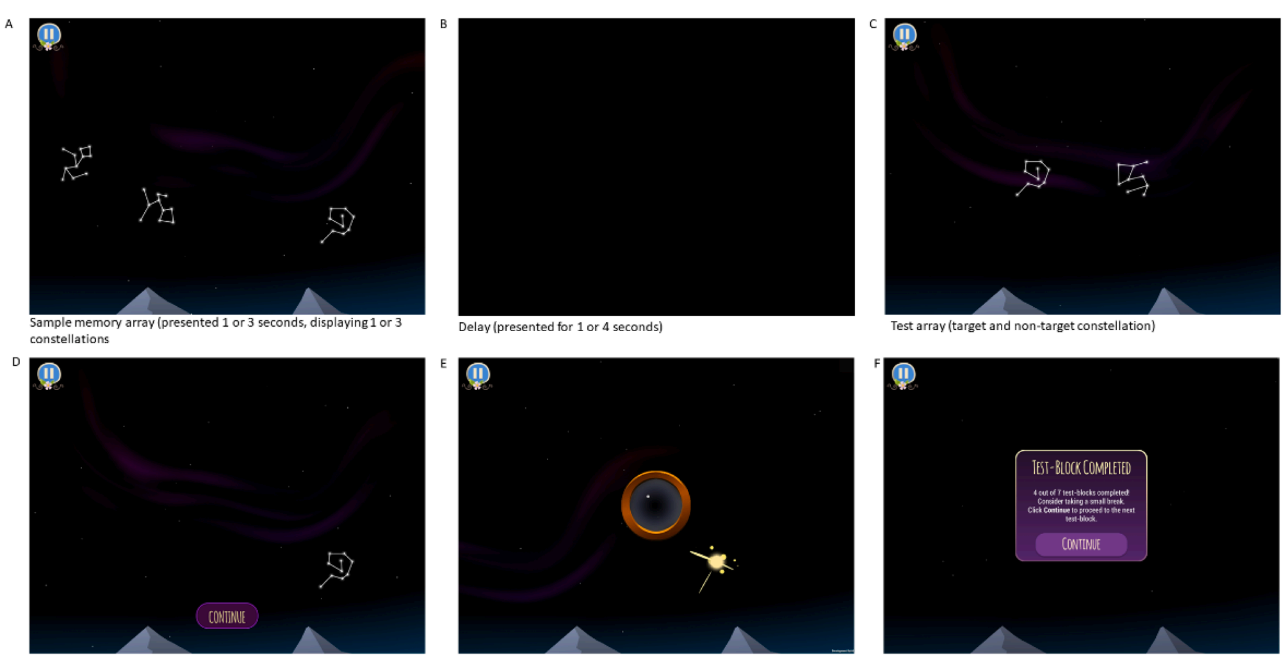

Fig. 2. Starry Night game. In sample memory array (A) one or three constellations are presented to memorise, following a 1- or 4-second delay (B). Next in a test array two constellations are presented (target and foil) (D). In two conditions participants also received a distraction task (a telescope to follow celestial objects) (E). The task was to drag the target constellation to the remembered original location (D). There were 6 40-second blocks and a 10-second rest between each block (F).

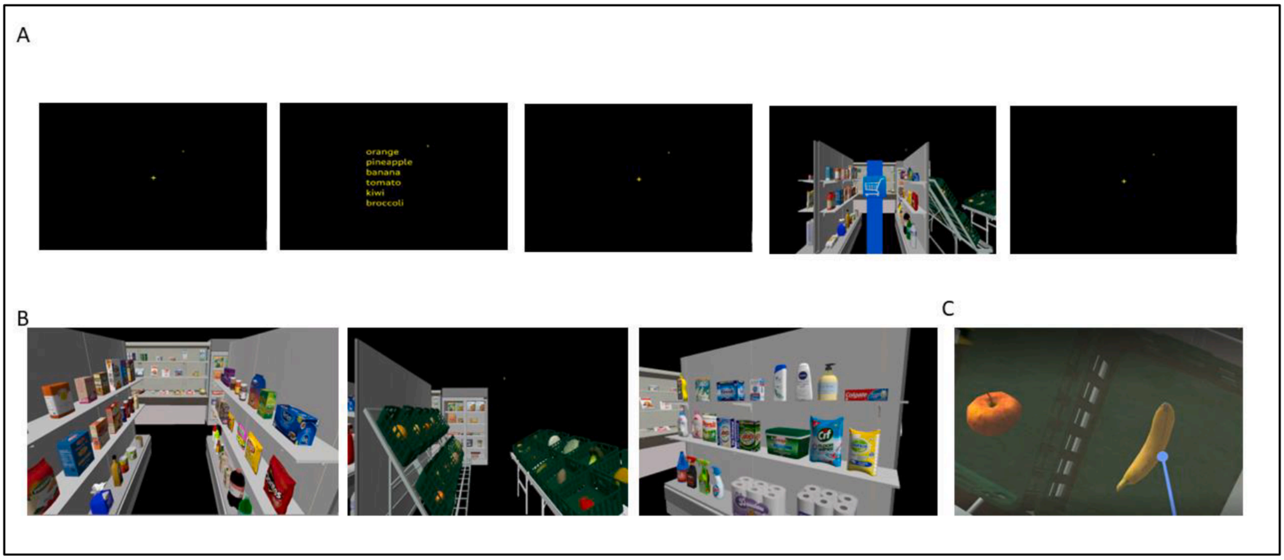

Fig. 3. Shopping task. A) Participants were presented with 10- a second fixation point following a 10-second shopping list (5 levels of difficulty) to memorise. Then after 10 s of rest, they were asked to recall items from the list and find them in the shop. B) view in different shop sections: food, fruits and vegetables and household. C) participantselected products using a controller by pointing at them.

memorise and then had to retrieve as many of these items as possible by physically exploring the shop and selecting the items using the trigger button of the hand-held controller. Participants were informed that they had as much time as needed to complete the task; however, on average, participants completed the task within 30 to 70 s. The difficulty of the task was determined by the number of items on the list:

- 3 items shopping list
- 4 items shopping list
- 5 items shopping list
- 6 items shopping list
- 7 items shopping list

The selection of item quantities on the shopping lists was informed by established cognitive psychology literature on working memory capacity. Miller's working memory model suggested a capacity limit of 7 ± 2 items (Miller, 1956). However, other subsequent studies suggested a limit of 4 ± 1 chunks of information (Cowan, 2001; Oberauer et al., 2018) or 3 to 4 chunks of information (Mathy and Feldman, 2012). Consistent with these findings, our preliminary pilot study conducted using a 2D version of the task revealed that participants were able to recall an average of 5 items for both pre- and post-training sessions (Fig. 13 in supplementary materials). Therefore, the number of items chosen for the task in this study reflects both the theoretical underpinnings of working memory capacity and empirical evidence from initial investigations, ensuring that the task parameters are appropriately scaled to capacities of working memory.

The shopping performance was measured as a proportion of correctly recalled items from the list (measured as%) and a proportion of incorrectly recalled items (measured as%). The task was delivered through the AR headset Magic Leap. Magic Leap One (Fig. 4) is an augmented/mixed reality system that blends the virtual and real-world to produce content where digital and real-world objects co-exist. The system consists of three pieces:

1. Lightwear – a see-through head-mounted display, with tracking cameras to map the environment and track the user's motion, and eye-tracking cameras
2. Lightpack – a small, wearable, wireless computer with Nvidia Tegra X2 chipset, 8GB RAM, 128GB storage, and a battery

Fig. 4. User wearing Artinins Brite23 fNIRS device together with Magic Leap AR Glasses.

3. Controller – an input device, tracked within a simulation to interact with a digital content

3.5. Data collection instruments

Further to the performance data gathered from the Starry Night and shopping tasks, activity in the PFC was measured by fNIRS, and three questionnaires were used to assess working memory, immersion in augmented reality, and their experience of motion sickness.

3.5.1. fNIRS

Changes in brain activity were measured using the Brite23 wireless fNIRS system (Artinis Medical Systems, Elst). The probe covering the frontal cortex consists of eleven sources and seven detectors arranged in twenty-three channels, including two short-separation channels (See Fig 4). Data were recorded using the Oxysoft software (3.3.70 - x64

Artinis Medical Systems, Elst, The Netherlands).

3.5.2. Questionnaires

1. The Montreal Cognitive Assessment (MoCA) (alternative) – is a 30-question test to measure executive functions and multiple cognitive domains (takes approximately 10 min).
2. Augmented Reality Immersion (ARI) Questionnaire – is the 42-item scale used for measuring the subjective sense of presence and immersion in AR. Participants are asked to rate their presence on a 5-point Likert scale. The three subscales which assess different components of immersion: engagement, engrossment, and total immersion (Georgiou and Kyza, 2017).
3. Simulator Sickness Questionnaire (SSQ) – is the 16-item tool that assesses possible side effects of VR exposure on a 4-point Likert scale Simulator measureside effects of the VR simulation in the scale from 0 to 3, where 0 is coded as "none", 1 – "slight", 2 – "moderate" and 3 – "severe". A total score on the questionnaire is 48 which indicatesevere symptoms and 0 indicates no symptoms (Kennedy et al., 1993).
4. Stroop Test – is a psychology test designed to measure a processing speed (reaction time). There were two conditions in the Stroop test: congruent – when names of colours are printed in the same ink colour as the name of the colour; incongruent – when the name of the colour is printed in a different ink colour as the name of the colour. In total, there were 30 trials (15 congruent and 15 incongruent), and the experiment took approximately 5 min to complete including practice.
5. Digit Span – is a psychology test designed to measure working memory span. In the digit span, participants were presented with a sequence of digits to memorise. The digits were presented on the screen one digit at a time, starting from easy (one digit to memorise) and building up to difficult (10 digits to memorise). The test took approximately 5 min to complete.

The MoCa, Stroop test and digit span were used to assess the cognitive abilities of participants at the baseline. Thiscreening approach was to ensure that participants did not exhibit symptoms indicative of MCI, Alzheimer's disease or dementia, inclusion criteria essential for the demographic of this study. Therefore, we did not administer the MoCA, Stroop or digit span in subsequent sessions, including the post-training and follow-up phases. This approach allowed us to concentrate on the primary objectives of our study to test the ADDP solution while ensuring the cognitive suitability of our participants. The ARI and SSQ were used to assess the quality of the simulation and the participant's comfort during the shopping task.

3.6. Procedure and data acquisition

Each participant received the Participant Information Sheet at least 24 h before their first pre-training session. When participants arrived at the laboratory, they were asked to read and sign the consent form. Then each participant was assessed with MoCa to determine baseline cognitive abilities. Afterwards, each participant was sat ~50 cm in front to the computer (Windows 10 machine, Intel Core i7-4790 CPU @ 3.60 GHz, 16.0 GB RAM) and they were asked to complete short versions of Stroop test and Digit Span available on https://pavlovia.org/. Afterwards, participants were fitted with fNIRS device and sat approximately 50 cm in front of the tablet touch screen. The tablet was placed on the plastic stand such that the screen was facing directly towards the participant (rather than flat on a table), to prevent participants from leaning forward, which could cause noise contamination in the fNIRS data. Participants were then introduced to Starry Night. The test was delivered through the Samsung Galaxy S3 Tablet and took approximately 15 min to complete (including practice and tutorial). Afterwards, Magic Leap was placed over the fNIRS device and participants were introduced to the shopping task. For the majority of participants, it was their first experience of head-mounted augmented reality, and so participants were allowed to explore and familiarize themselves with the technology before beginning the shopping task. As the study employed wireless AR and fNIRS devices, participants were allowed to move and navigate the lab space freely and untethered. The shopping task took approximately 20 min to complete (including 5 min of familiarization and 5 min of practice). After completion, the researcher removed the Magic Leap and fNIRS devices from participants and they were asked to fill in two questionnaires – ARI and SSQ to evaluate participant's general experience and feasibility of combining AR and fNIRS.

The first session took approximately 2 h in total. Participants were invited to the post-training assessment only if they had finished at least 90 % of their training program. The procedure during the post-training assessment was the same, however, MoCa, ARI and SSQ were not administered. The second session lasted 90 min. The follow-up session was the same as the post-training session, however, there was a short debrief with the researcher at the end and participants were asked to provide very short (1 or 2 sentences) feedback on their experience and perception of the CCT app.

4. Data analysis

Throughout the analysis, we compare across three variables:

- Group – between participants analyses depending on which version of the CCT they used.
- Session – within participants analyses comparing pre-training, post-training, and follow-up
- Condition – typically repeated measures within-participants condition, depending on the task (e.g., difficulty of shopping task, level in Starry Night).

4.1. fNIRS data analysis

The fNIRS data analysis was conducted using NIRS Toolbox (Santosa et al., 2018). The data was down-sampled to 4 Hz. Raw signals were converted to optical density changes and then to oxyhaemoglobin (HbO) and deoxyhaemoglobin (HbR) estimates using Beer-Lambert law with a partial path length correction of 0.1 for both wavelengths (Strangman et al., 2003). Motion artifacts were corrected using the Temporal Derivative Distribution Repair (TDDR) (Fishburn et al., 2019). On the first level analysis, beta coefficients for task activations were estimated using the autoregressive iteratively-reweighted least squares approach (Barker et al., 2013). To correct for physiological noise, we included a short-separation channel as a nuisance regressor in GLM (Gagnon et al., 2011). The BoxCar function was used to model hemodynamic response for the shopping task. This approach was taken because the tasks varied in duration (the 3-items list condition lasted approximately 30 s and 7-items condition would last approximately 60–70 s) and we trimmed longer tasks to 30 s, therefore resultant hemodynamic responses could deviate from the typical assumed shape and timing captured by a canonical HRF. This approach was more suitable for our situation focusing on the presence of the stimulus rather than its temporal dynamics. The canonical HRF was used to model hemodynamic response for the working memory task Starry Night. For group analysis, a mixed-effects model was used to determine the effects of each group, session and each condition as fixed effects and subject as a random effect. The false discovery rate (FDR) correction was used with the significance level set at 0.05 ($q \leq 0.05$) (Benjamini and Hochberg, 1995) to control for multiple comparisons. Contrast analyses were used to assess differences between groups, sessions, and conditions on the PFC, and conditions on the PFC.

4.2. Behavioural and questionnaire data analysis

All analyses on behavioural and questionnaire data were performed with R statistical computing environment, version 4.1.3 (RCoreTeam, 2021). The difference between groups reflected in MoCa scores at the baseline were assessed using the Wilcoxon signed-rank test. A generalised Linear Mixed Model (GLMM) was fitted to investigate the effects of training using packages 'lme4' for continuous variables (Bates et al., 2012) and 'glmmTMB' for skewed and zero-inflated variables (Brooks et al., 2017). Post-hock contrasts were calculated using the 'emmeans' package (Lenth, 2022). Plots were produced using 'ggplot2' package (Wickham et al., 2016).

Group waspecified as a grouping factor and condition and session were specified and fixed factors modelled with random intercepts and where possible also random slopes were set. Additionally, models with participant ID as random factors were tested when no group effects were found. We tested models with MoCA, time played, and immersion - set as both random and fixed effects models. Appropriate families were selected based on the distribution of the residuals. Numerous families and link functions were evaluated where necessary. When models failed to converge, they were gradually simplified by removing random slopes and then fixed factors and interactions. Then, the best-fit models that converged, were selected based on the Akaike information criterion (AIC). Parameter estimates (β), standard errors, t- or z-values and p-values are reported for each best-fit model that converged.

5. Results

The summary (mean, SD and median) of MoCa, Stroop test, digit span, ARI, and SQQ scores for both groups are presented in Table 1. The median MoCa score between groups was not statistically significant ($U = 1009.0$, $p = 0.34$) which indicates that both groups demonstrated similar levels of cognitive abilities at the baseline. There was no effect of group (*beta* = 0.01, 95 % *CI* [−0.48, 0.51], $p = 0.958$; *Std. beta* = −0.02, 95 % *CI* [−0.12, 0.07]) and no effect of session (*beta* = 0.09, 95 % *CI* [−0.27, 0.46], $p = 0.61$; *Std. beta* = 0.04, 95 % *CI* [−0.05, 0.14]) for the accuracy in the Stroop task. Also, there was no effect of group (*beta* = 0.03, 95 % *CI* [−0.23, 0.30], $t(434) = 0.24$, $p = 0.81$; *Std. beta* = 0.18, 95 % *CI* [0.02, 0.34]), and no effect of the session (*beta* = 4.58e-03, 95 % *CI* [−0.18, 0.19], $t(434) = 0.05$, $p = 0.96$; *Std. beta* = −0.09, 95 % *CI* [−0.15, −0.03]) for the reaction time during the Stroop test. However, there was the effect of condition (*beta* = 0.17, 95 % *CI* [0.14, 0.20], $t(440) = 10.85$, $p < 0$ 0.001; *Std. beta* = 0.61, 95 % *CI* [0.50, 0.72]) which demonstrated that, as would be expected, participants performed significantly faster during the congruent Stroop condition.

Table 1
The summary (mean, SD and median) of MoCa, Stroop test, digit span, ARI, and SQQ scores for both groups.

	Experimental			Control		
	Mean	SD	Median	Mean	SD	Median
MoCa	28.34	2.16	29.00	27.65	2.68	28.00
Stroop						
Accuracy Congruent	0.99	0.01	1.00	0.99	0.02	1.00
Accuracy Incongruent	0.96	0.10	1.00	0.95	0.07	1.00
RT Congruent	0.82	0.21	0.96	1.04	0.22	1.01
RT Incongruent	1.13	0.30	1.08	1.22	0.24	1.16
Digit Span	2.92	1.60	4.00	2.85	1.52	4.00
Time Played	802.23	73.34	786.5	895.93	142.40	859
Cybersickness (SSQ)	0.23	0.19	0.22	0.40	0.35	0.36
Presence (ARI)						
Engagement	4.21	0.41	4.22	4.12	0.39	4.44
Engrossment	4.10	0.43	4.00	3.99	0.47	4.16
Immersion	3.55	0.59	3.58	3.79	0.70	3.66

Additionally, there was no effect of group (*beta* = 0.02, 95 % *CI* [−0.25, 0.29], $p = 0.86$; *Std. beta* = 0.02, 95 % *CI* [−0.25, 0.29]) and no effect of session (*beta* = 0.14, 95 % CI [−0.12, 0.40], $p = 0.288$; *Std. beta* = 0.14, 95 % *CI* [−0.12, 0.40]) in the digit span test. Although the Stroop test and Digit Span did not demonstrate an improvement after the training, they also served as a method to characterise the cognitive abilities of participants and determine a difference between both groups at the baseline.

There was a significant difference between groups in how much time they played neurogames at home ($p = 0$ 0.001). The median time played for the experimental group was 786 min and the median time for the control group was 859 min.

Below we present the results of the experiment for fNIRS and behavioural data. In line with the hypotheses, we first present results for fNIRS data (changes in mental workload) and then behavioural data (accuracy, localisation performance and misbinding) results from the Starry Night experiment. Then we present results for fNIRS (changes in mental workload) and behavioural (accuracy and errors) results for the shopping task. We present results only of the best-fit models that converged.

5.1. Starry night

5.1.1. fNIRS results

All significantly activated channels with beta values, SE, t-stat values, p-values, q-values and relative power are presented in Table 2 and Fig. 5. Only channels with a negatively correlated haemoglobin species (an increased HbO and decreased HbR at $q < 0.05$) were considered asignificant (Cui et al., 2010). A mixed effect model was fit to predict beta (brain activity) with group, session and condition as fixed effects and id as random effect:

$$Beta \sim Group : Session : Condition + (1 \mid id)$$

The result revealed a significant effect of group, session, and condition. A contrast analysis demonstrated lower PFC activation (indicative of lower mental workload) in the experimental group relative to the control group after the training indicated by significantly lower HbO and higher HbR during condition 3TI in channels S4-D2, S5-D3, S5-D3 and S6-D4. At the follow-up, the experimental group demonstrated a significantly lower activity in the channel S6-D4 relative to the control group.

5.1.2. Behavioural data results

For the accuracy analysis, we fitted a generalised mixed model (Beta family with a logit link) to predict Accuracy with Condition and Session with ID as random effect:

$$Accuracy \sim Session * Conition + (1 \mid id)$$

The model's explanatory power related to the fixed effects alone (marginal R2) is −0.56. The analysis revealed that there was no significant effect of group and session for accuracy. However, we found a significant effect of the condition. Post-hoc analysis revealed that the accuracy was higher in condition 1TSD than 3TSD ($\beta = 0.02$, $SE = 0.01$, $t\text{-}ratio = 4.07$, $p = 0.001$), 1TLD than 3TLD ($\beta = 0.026$, $SE = 0.01$, t-ratio =4.07, p = 0.001), and 1TI than 3TI (β = 0.09, $SE = 0.01$, $t\text{-}ratio = 8.72$, $p = 0.001$) (Fig. 6). This finding essentially confirms that easier levels led to higher accuracy. All values for (beta estimates, SE, CI, Statistic and p-values are presented in the supplementary materials.

We fitted a general linear mixed model (Gamma family with an identity link) to predict Localisation performance with Condition and Session with ID as a random effect:

$$Localisation\_Performance \sim Condition * Session + (1 \mid id)$$

The model's total explanatory power is substantial (conditional R2 = 1.00), and the part related to the fixed effects alone (marginal R2) is of

Table 2
Results of contrast analysis for all significantly activated channels (source-detector pairs) during Starry Night game.

Source	Detector	Chromophore	Condition	Beta	SE	tstat	p	q	Power
4	2	HbO	Experimental-Control (Post-Pre)	−4.63	1.76	−2.62	0.01	0.04	0.34
4	2	HbR	Experimental-Control (Post-Pre)	2.87	0.90	3.18	0.00	0.01	0.67
5	2	HbR	Experimental-Control (Post-Pre)	2.57	0.92	2.81	0.01	0.02	0.66
5	3	HbO	Experimental-Control (Post-Pre)	−6.46	1.90	−3.40	0.00	0.00	0.32
6	4	HbO	Experimental-Control (Post-Pre)	−7.77	1.77	−4.38	0.00	0.00	0.34
6	4	HbR	Experimental-Control (Post-Pre)	2.89	0.77	3.76	0.00	0.00	0.79
6	4	HbO	Experimental-Control (Follow-up-Pre)	−7.23	1.71	−4.23	0.00	0.00	0.38
6	4	HbR	Experimental-Control (Follow-up-Pre)	2.22	0.81	2.74	0.01	0.03	0.80

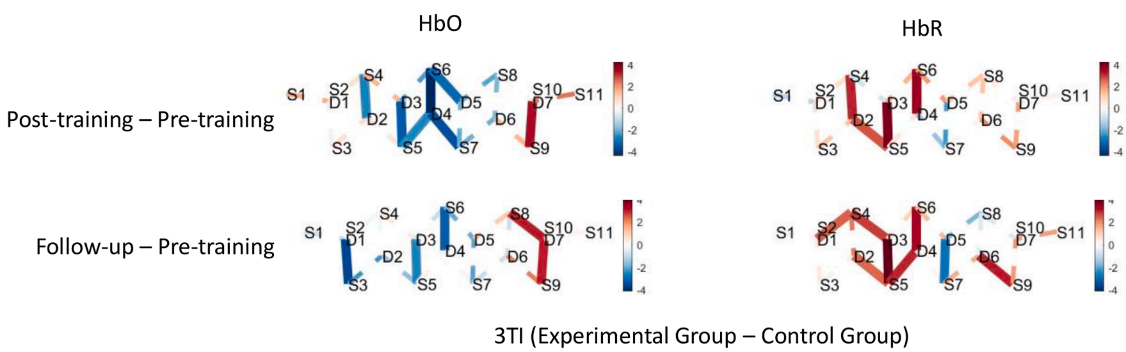

Fig. 5. - Results of contrast analysis from fNIRS data, showing significant channels (source–detector pairs) where differences in oxyhaemoglobin (HbO) and deoxyhaemoglobin (HbR) levels were observed between the experimental and control groups during a 3-target with interference condition (3TI). The differences are shown at two time points: after training (top) and at a follow-up session (bottom). The significance of the contrasts is indicated by a colour scale representing t-statistics, where the intensity of red and blue colours corresponds to the magnitude of HbO and HbR differences, respectively. All reported findings are statistically significant with $q < 0.05$, adjusted for multiple comparisons using FDR correction.

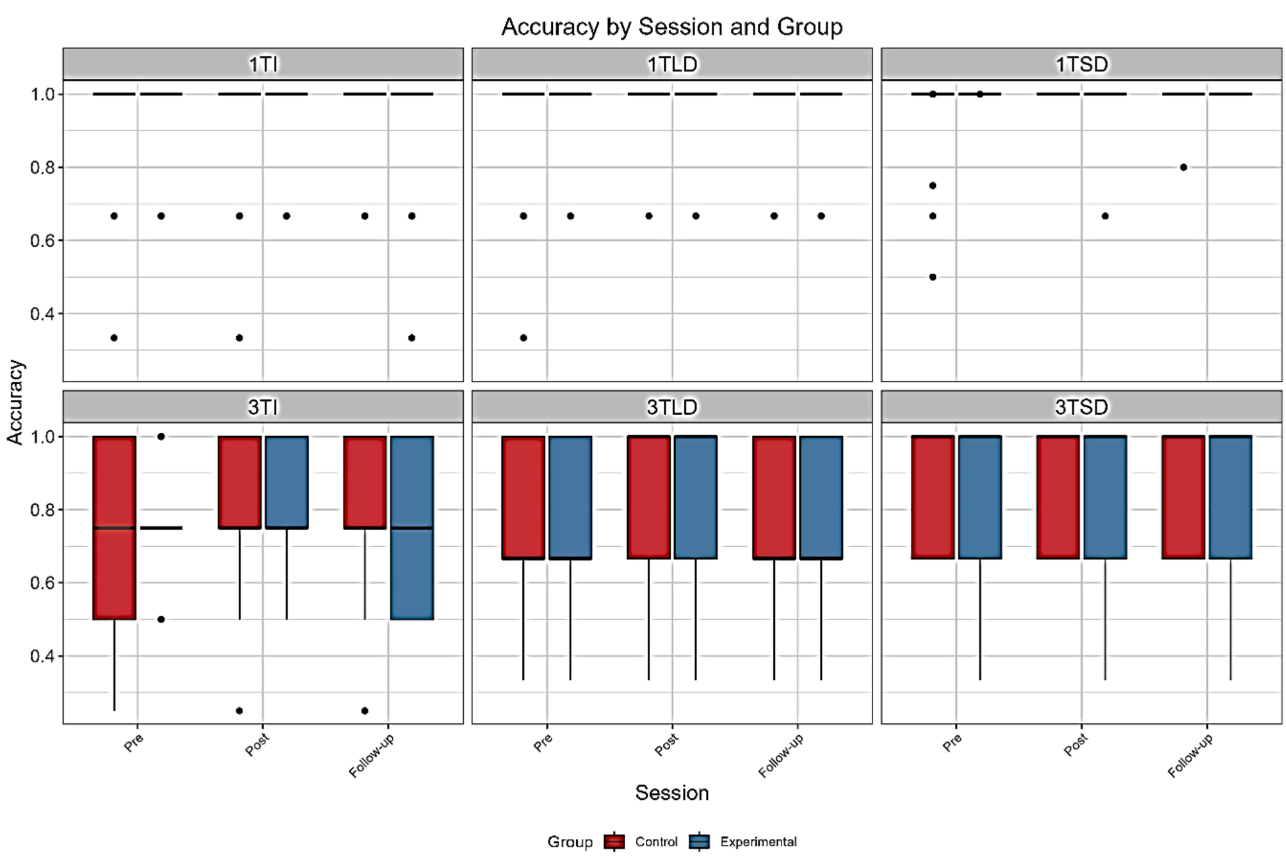

Fig. 6. Differences in accuracy from Starry Night test battery for the control and experimental group. Differences were found between conditions, but not groups and sessions.

0.98. The model demonstrated a significant effect of session and a significant effect of condition. Post-hoc analysis revealed that the overall localisation performance has increased from pre-training to follow-up in both groups ($\beta = -21.50$, $SE = 8.52$, $t\text{-}ratio = -2.52.$, p = 0.03). Moreover, the localisation performance was better in condition 1TSD than 3TSD ($\beta = -249.47$, $SE = 12.66$, $t\text{-}ratio = -19.71$, p = 0.001), in condition 1TLD than 3TLD ($\beta = -179.02$, $SE = 10.23$, $t\text{-}ratio = -17.49$, $p = 0.001$), and in condition 1TI than 3TI (β = −200.59, $SE = 12.65$, $t\text{-}ratio = -15.85.$, p = 0.001). Additionally, participants performed better in condition 1TSD than 1TI ($\beta = -40.66$, $SE = 4.99$, $t\text{-}ratio = 4.26$, p = 0.001), 3TSD than 3TLD (1TI β= 67.11, $SE = 15.73$, $t\text{-}ratio = -8.14$, $p =$ 0.001) and condition 3TLD than 3TI ($\beta = -58.89$, $SE = 15.39$, $t\text{-}ratio =$ -3.82, p = 0.001) (Fig. 7). These results mean that participantsignificantly improved their localisation performance in easier conditions. Moreover, the result lasted at the follow-up. All values for the model (beta estimates, SE, CI, Statistic and p-values are presented in the supplementary materials.

For the misbinding analysis, we fitted a zero-inflated generalised linear mixed model (Gamma family with an identity link) to predict Misbinding with Condition and Session and ID as a random effect:

$$Misbinding \sim Condition * Session + (1 \mid id)$$

The model included participant ID as a random effect. The model's total explanatory power is weak (conditional R2 = 0.04) and the part related to the fixed effects alone (marginal R2) is of 0.02. The result revealed that there was no significant effect of group and session. However, there was an effect of the condition. Post-hoc analysishowed that the chance of misbinding was higher in condition 3TI than in condition 3TSD ($\beta = 0.07$, $SE = 0.01$, $t\text{-}ratio = 3.90$, p = 0.001), and higher than in 3TLD ($\beta = 0.07$, $SE = 0.02$, $t\text{-}ratio = 3.53$, p = 0.001). (Figure 17 in the supplementary materials). All values for the model (beta estimates, SE, CI, statistics and p-values are presented in the supplementary materials.

5.2. Shopping task

Regarding the quality of the simulation and comfort of the equipment, both groups reported a high level of engagement, engrossment, and a moderate effect of immersion. Also, the overall cybersicknesscore for both groups was low, which indicates that combining fNIRS with augmented reality did not cause significant discomfort of adverse effects (Table 1).

5.2.1. fNIRS data results

All significantly activated channels with beta values, SE, t-stat values, p-values, q-values and relative power are presented in Table 3 and Fig. 8. Only channels with a negatively correlated haemoglobin species (an increased HbO and decreased HbR at $q < 0.5$) were considered asignificant (Cui et al., 2010) . We fitted a mixed effect model to predict beta (brain response) with group, session, condition as fixed effects and ID as random effects:

$$Beta \sim Group : Session : Condition + (1 \mid id)$$

We found no significant effects of groups (See Fig. 15 in the supplementary materials), therefore, we then tested less complex models with fewer variables. Next, we fit a mixed effect model with Session and Condition as fixed effects and subject ID as a random effect:

$$Beta \sim Session : Condition\ (1 \mid id)$$

The analysis revealed a significant effect session and condition. A contrast analysis demonstrated a decrease within the PFC (indicative of lower mental workload) after the training indicated by significantly

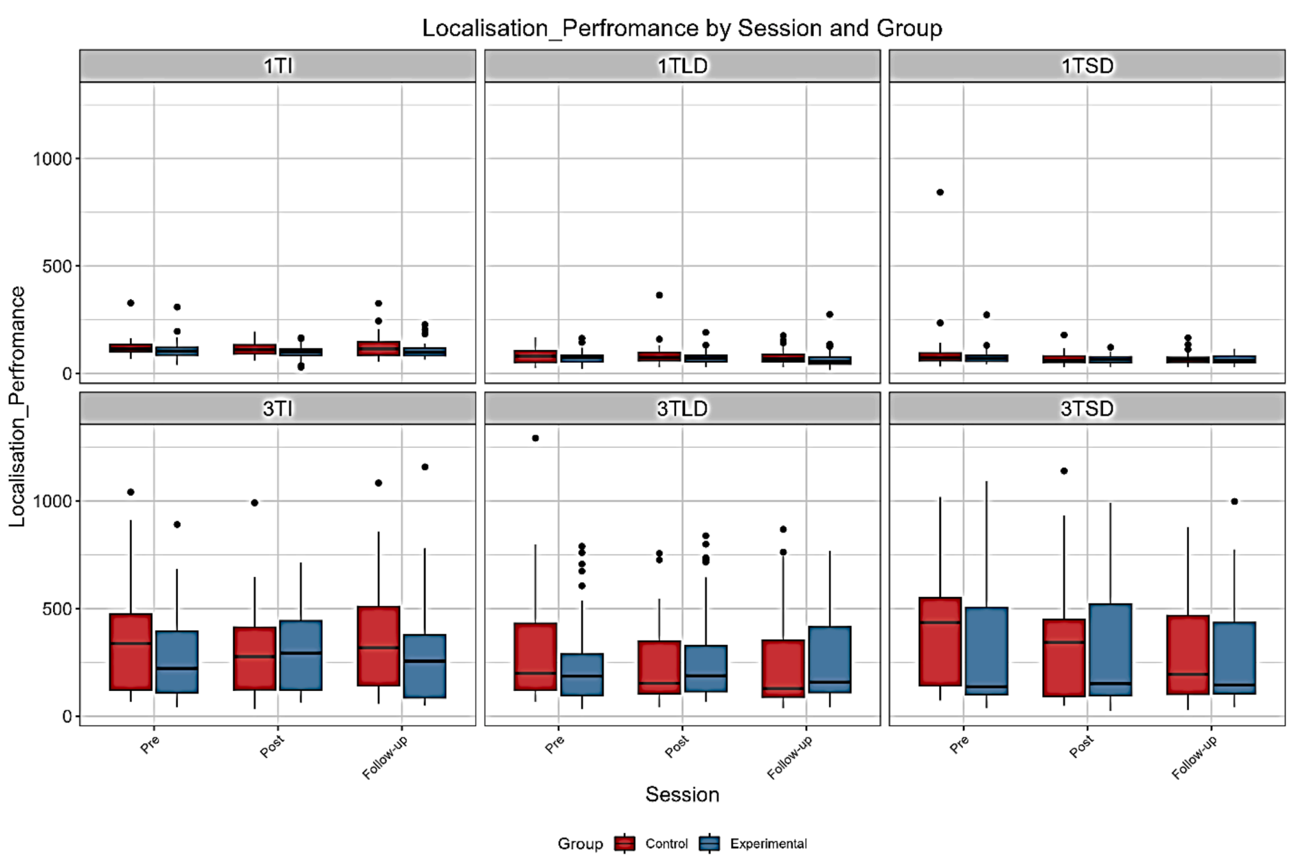

Fig. 7. Differences in localization performance from Starry Night test battery for the control and experimental group. Differences were found between conditions and sessions, but not between groups.

Table 3
Results of contrast analysis (Pre-training – Post-training) for all significantly activated channels (source-detector pairs) during the shopping task in augmented reality.

Source	Detector	Type	Condition	Beta	SE	tstat	p	q	Power
8	6	HbO	5-items-Pre - Post	−4.95	1.44	−3.42	0.001	0.005	0.34
8	6	HbR	5-items-Pre - Post	2.23	0.61	3.68	0.001	0.005	0.82

Fig. 8. Results of contrast analysis and the fNIRS probe with significant channelshowing oxyhaemoglobin (HbO) and deoxyhaemoglobin (HbR) difference between the pre- and post-training for 5-item shopping list condition) (source-detector pairs) ($q < 05$; FDR-corrected). The colour bar shows t-statistics for the differences between the pre- and post-training.

lower HbO and higher HbR for 5-item shopping list condition in channels S8-D6. These results show that participants performed better after training, but the effect was not maintained at the follow-up (See Fig. 16 in the supplementary materials).

5.2.2. Behavioural data results

We fitted a general linear mixed model (Beta family with a logit link) to predict Accuracy with Session, Condition and MOCA_standarised with ID as random effect:

$$Accuracy \sim Session * Condition + MOCA\_standarised(1\backslash id)$$

The model's total explanatory power is substantial (conditional R2 = 1.26) and the part related to the fixed effects alone (marginal R2) is of 1.15. The result revealed that there was no significant effect of the group. However, there was an effect of session, condition, MOCA, and the interaction between session and condition was also significant. The

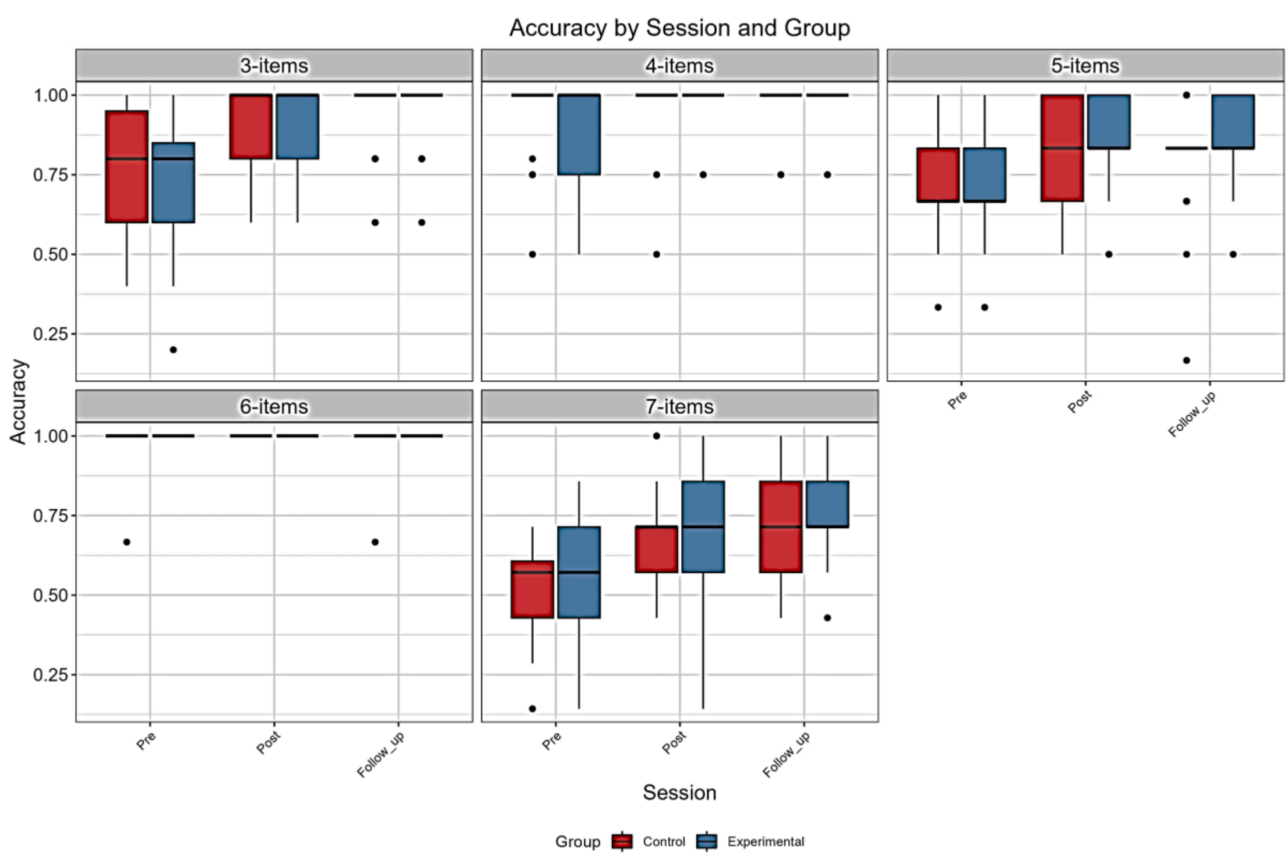

Fig. 9. The differences in task accuracy between control and experimental groups across various conditions (number of items on the shopping lists) and sessions (Pre, Post, Follow-up) of the shopping task. Statistical analysis revealed significant differences between conditions and sessions, suggesting a variation in performance over time and with task complexity. However, no significant group differences were found indicating that the experimental intervention did not significantly alter accuracy compared to the control.

accuracy increased significantly after the training (versus pre-training) in condition 3-items shopping list ($\beta = 0.08$, $SE = 0.01$, t-ratio = 5.63, $p = 0.001$), 5-items shopping list ($\beta = 0.13$, $SE = 0.02$, t-ratio = 6.36, $p = 0.001$), and 7-items shopping list ($\beta = 0.15$, $SE = 0.03$, *t-ratio* = 4.22, $p < 0.001$). The effects were also maintained at the follow-up for 3-items shopping list ($\beta = 0.09$, SE = 0.01, t-ratio = 6.39, p = .01), 5-items shopping list (β = 0.12, SE = 0.02, t-ratio = 5.80, $p = 0.001$), and 7-items shopping list ($\beta = 0.23$, $SE = 0.03$, *t-ratio* = 6.67, $p = 0.001$). This means that the behavioural improvements, where found, were maintained after the follow-up. The effect of MOCA_standarised wastatistically significant ($\beta = 0.11$, $SE = 0.05$, *t-ratio* = 2.46, $p = 0.014$), implying that participants with better MoCa scores generally performed better across conditions, as perhaps would be expected (Fig. 9). All values for (beta estimates, SE, CI, Statistic and p-values are presented in the supplementary materials.

For error analysis, we fitted a zero-inflated generalised linear mixed model (Gamma family with an inverse link) to predict Errors with Condition and Session as fixed effects and ID as random effect:

$$Errors \sim Condition * Session + (1 \mid id)$$

The model's total explanatory power is substantial (conditional R2 = 0.98), and the part related to the fixed effects alone (marginal R2) is of 0.57. Results revealed no significant effect of group and session. However, there was a significant effect of the condition. Post-hoc analysis revealed that there was a significant difference between the condition 4-items shopping list and the 5-items shopping list ($\beta = 0.06$, $SE = 0.01$, *t-ratio* =3.40, p = 0001), condition 4-items shopping list and 7-items shopping list ($\beta = 0.08$, $SE = 0.01$, *t-ratio* = 4.80, $p < 0.001$), condition 5-items shopping list and 7-items shopping list ($\beta = 0.02$, $SE = 0.00$, t-ratio = 3.65, p = 0001) and 6-items shopping list and 3-items shopping list ($\beta = -0.04$, $SE = 0.00$, *t-ratio* = −4.25. $p = 0.001$) (Fig. 10). This highlights that participants made fewer errors in the easier tasks, as would be expected. All values for (beta estimates, SE, CI, Statistic and p-values are presented in the supplementary materials.

5.3. Feedback analysis

5.3.1. Experimental group

Feedback was analysed using natural language processing (NLP) based on sentiment analysis. The overall sentiment score of the experimental version of the CCT was positive. Reviews were mostly positive indicating that the majority of people enjoyed the app. Positive emotions (such as joy, trust, and anticipation) were higher in frequency than negative. The most frequent words used by participants describing the app were "easy", "enjoyed" and "good". The results are presented in Fig. 11.

5.3.2. Control group

The overall sentiment score of the non-adaptive version of the CCT was positive. Reviews were mostly positive indicating that the majority of people enjoyed the app. Positive emotions (such as joy, trust and anticipation) were higher in frequency than negative. The most frequent words used by participants describing the app were "easy" and "good", however, there were also some negative emotionsuch as "repetitive" and "boring". The results are presented in Fig. 12.

6. Discussion

The main aim of this study was to evaluate the efficacy of a dynamically adapting CCT intervention programme by measuring

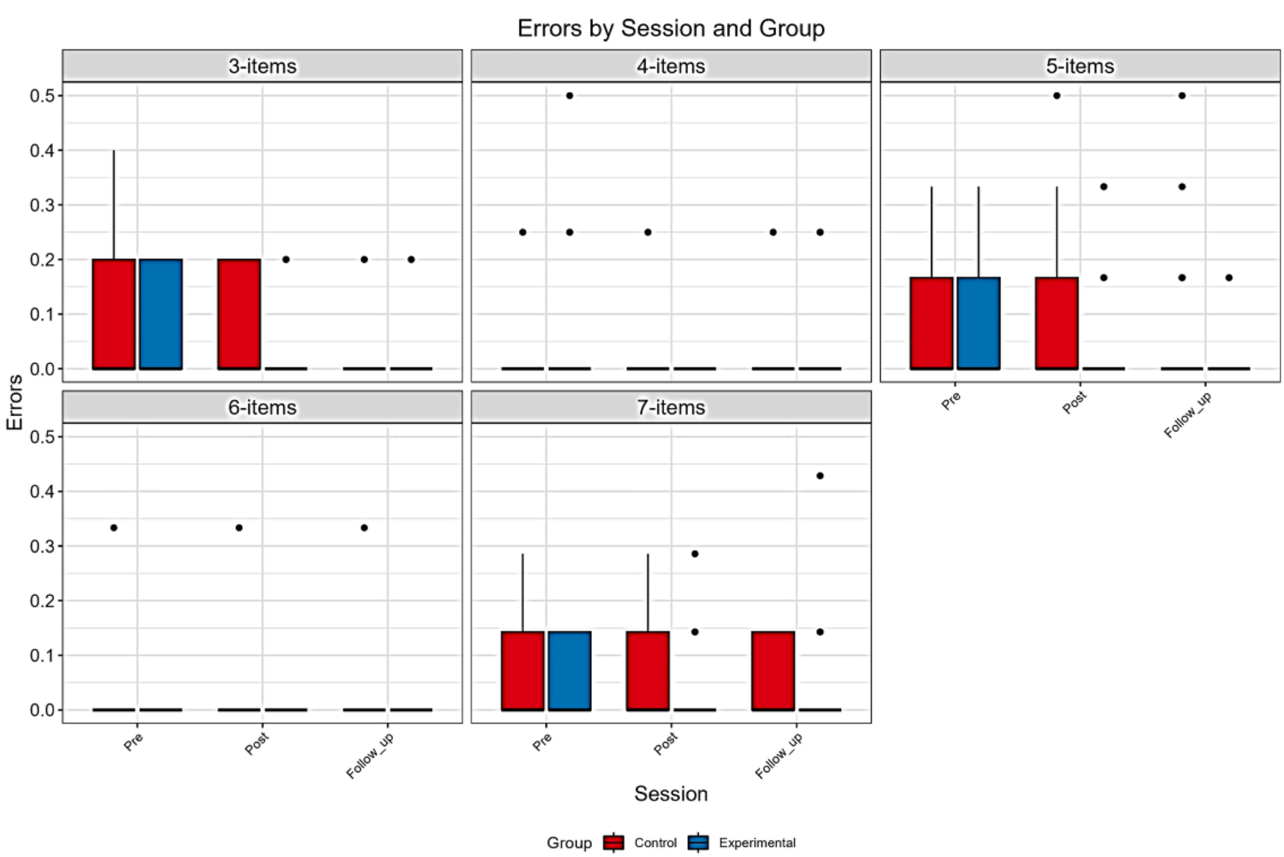

Fig. 10. The differences in task errors between control and experimental groups across various conditions (number of items on the shopping lists) and sessions (Pre, Post, Follow-up) of the shopping task. Statistical analysis revealed significant differences between conditions and sessions, suggesting a variation in performance over time and with task complexity. However, no significant group differences were found indicating that the experimental intervention did not significantly alter errors compared to the control.

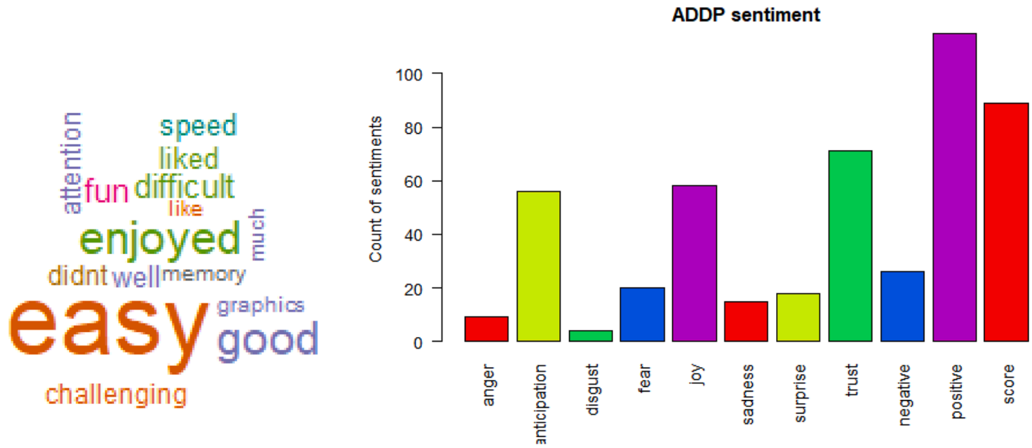

Fig. 11. Results of sentiment analysis for the experimental version of AD intervention. Most frequent words (left) and participant'sentiment towards the app (right).

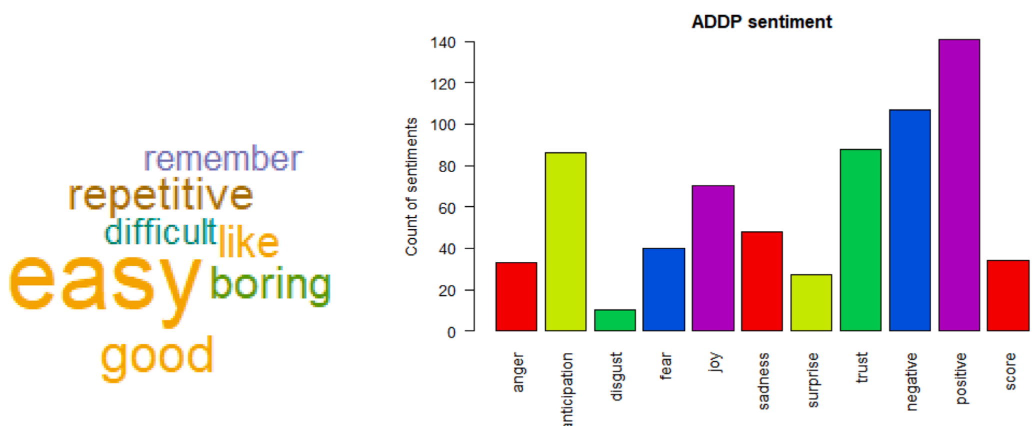

Fig. 12. Results of sentiment analysis for the placebo version of AD intervention. Most frequent words (left) and participant'sentiment towards the app (right).

changes in hemodynamic indicatives of mental workload at three different time points: prior to the CCT, following an 8-week intervention, and at a 1-month follow-up. The objective was to measure changes in brain activity using fNIRS and cognitive abilities following the ADDP programme using a detection test battery (Starry Night) and an ecologically valid transfer task (UoN shopping game in AR). Participants were divided into two groups: 1) the active experimental group (the difficulty in the active experimental version was adjusted to the individual performance) 2) the control active placebo group (the difficulty in the experimental passive version was set at a constant level). Below we discuss the evidence relating to each of the hypotheses.

H1: There will be a significant change in brain activity within the PFC during the working memory task Starry Night after the ADDP training as measured by fNIRS

The results of the study demonstrated decreased prefrontal activity as measured by fNIRS during the Starry Night test after the AD intervention in the experimental group that used the dynamic training. In particular, we found decreased activity in the bilateral DLPFC (channels S4-D2, S5-D3, S5-D3 and S6-D4) during the condition when participants were presented with 3 targets with the interference task. The result was also maintained at the follow-up (channel S6-D4). These results are in line with previous fMRI studies which demonstrated decreased activity within the PFC regions following cognitive training (Brehmer et al., 2011; Chang et al., 2017; Clark et al., 2017; Heinzel et al., 2016; Miró-Padilla et al., 2019). This result might represent reduced demand for metabolic and mental resources and therefore increased neural efficiency to perform a task following cognitive training. This increased neural efficiency could also translate to enhanced working memory performance, which was detected using a working memory precision task – Starry Night. Moreover, the effect was maintained at the follow-up, which potentially demonstrates the efficacy of the adaptive version of the ADDP training over the static version.

We did not find any significant results for the other high workload conditions (3 targets with 4-second delay) and the low workload conditions (1 target with 1-second delay). This might be for three reasons. First, it is possible that fNIRS is not sensitive enough to detect such subtle variations in task difficulty. Previous studies showed that fNIRS can differentiate between large task difficulties (easy vs difficult), but smaller differences could not be distinguished (Ayaz et al., 2012). The second reason may be related to the signal variability which might result from individual differences in resource allocation. Perhaps, some individuals require fewer mental resources than others to complete a task, depending on their expertise or cognitive abilities and therefore show different levels of activity (di Domenico et al., 2015; Grabner et al., 2006). Thirdly, there are some study design limitations created by the longitudinal study design, which are not ideal for fNIRS data collection. Because fNIRS signals are often contaminated by physiological noise or motion artefacts (despite being more tolerant of movement than e.g. EEG), ideal data collection for fNIRS data should mean that each condition is repeated multiple times to resolve task-evoked response (Kirilina et al., 2012; Yücel et al., 2021) typically called block design. In our longitudinal study design, we needed to minimise participants' discomfort related to repeated sessions and wearing fNIRS for too long across both tasks. This meant that the duration of each Stary Night condition was 40 s and there was only one repetition of each unique configuration. Future studies could build in a block design in order to focus more specifically on this aspect of the results.

H2: There will be a significant change in brain activity within the PFC during ecologically valid shopping task after the ADDP training as measured by fNIRS

There was no difference between the groups in brain activity during the shopping task. However, the analysis demonstrated a decreased mental workload in the bilateral dorsolateral prefrontal cortex after the training phase during the 5-items condition regardless of the group. This result could potentially indicate that both versions of ADDP intervention were effective in enhancing cognitive efficiency in improving everyday functioning, which wasimulated in the ecologically valid shopping task. Importantly, the specific observation of changes during the 5-items condition aligns with existing models of working memory. According to these models, an average of 4–5 items represents the typical capacity that individuals can maintain in their working memory (Cowan, 2001; Oberauer et al., 2018). This capacity is crucial as it underpins a range of cognitive tasks in everyday life. The improvement in the 5-items condition suggests that cognitive training may be particularly effective in optimizing brain function up to this average capacity.

The absence of similar improvements in other conditions might be related to two potential reasons. Firstly, participants may already be operating at peak efficiency during tasks that exceed the average working memory capacity, leaving limited room for further improvement through cognitive training. Secondly, tasks involving more than 5 items might constitute an overload for the working memory capacity and therefore could lead to a plateau in performance improvement. Moreover, the lack of similar significant results for the other conditions might again result from the fact that fNIRS is not the optimal method to distinguish between subtle differences in the workload as discussed above. On the other hand, McKendrick et al. (2017) demonstrated that walking itself, and an increase in environmental complexity, might lead to the reduction of total haemoglobin in the PFC as a result of the distribution of mental resources to other brain areas. The results were not maintained at the follow-up.

Previous fMRI literature shows mixed results regarding long-lasting neuronal changes following cognitive training. Some studies demonstrated that cerebral changes were maintained 5 weeks (Miró-Padilla et al., 2019) or 6 months (Subramaniam et al., 2014) after the training, while other studies found no effect of the training at follow-up after 10-weeks (Kable et al., 2017) or 12 months (Li et al., 2019). As for now, there are no studies that have employed fNIRS to assess the long-term impact of cognitive training on brain function, therefore more studies are needed.

H3: Performance will increase for the cognitive working memory task after the ADDP training as measured via AD test battery Stary Night

We did not find an improved accuracy for working memory task after the AD intervention as measured via Starry Night. There was no effect of the group and no effect of the session. However, we did find a significant difference between conditions regardless of the session and group, confirming that the conditions did vary in difficulty. The participants were more accurate in conditions when there was one target to memorise versus three targets. However, the duration of the delay (1 vs 4 s) and interference task did not impact significantly on the accuracy. These results imply that memorising fewer items requires less mental workload, therefore the accuracy is higher. When mental workload increases, then the accuracy decreases (Pagnotta et al., 2021). The results are in line with the previous “What was where” study which demonstrated a significant effect of a number of items, but no effect of delay on the identification performance (Pertzov et al., 2012).

The results revealed overall improved localisation performance after the training in both groups from pre-training to follow-up regardless of the condition. There was also a significant effect of the condition. The post-hoc analysis also revealed that the localisation performance was determined by the number of targets to memorise (one versus three targets), the duration of the delay (1 versus 4 s) and also by the interference task. Again, this result is consistent with the study conducted by Pertzov et al. (2012, 2013) which demonstrated that localisation performance is correlated with a number of objects to be remembered and the longer retention interval. This means that having more items in working memory for longer delays leads decrease in localisation performance.

Localization tasks might respond better to cognitive training because they use different brain networks than just remembering or recognizing things (Manohar et al., 2019). These tasks involve spatial memory (remembering where things are, understanding space and planning) and are linked to brain areasuch as the hippocampus or the parietal lobe. Training these brain networks might be more susceptible to neuroplasticity and cognitive improvement but more detailed research is needed.

For misbinding, we found no significant effects of the group and session. However, the chances of misbinding were higher when participants were distracted by the interference task versushort or long delay. In general, the results of this study might indicate that localisation performance could be more sensitive measure of performance than accuracy and misbinding and could be more susceptible to CT.

H4: Performance will improve for the transfer task after the ADDP training as measured via an ecologically valid shopping task

The shopping test accuracy demonstrated no significant effect of the group, however, there was a significant effect of the session and condition. The accuracy increased after the AD intervention in 3-, 5- and 7-items shopping list conditions. The effect was maintained at the follow-up which means that the intervention leads to the lasting behavioural improvements in the transfer task. Participants however did not improve on 4- and 6-item shopping list conditions. There could be many reasons for that outcome. Firstly, for the easy and difficult conditions, we found a ceiling effect therefore participants have no more room for improvement. The ceiling effect could be due to the food categories, asome participants have reported that items from the category “fruits & vegetables” were easier to recognise and remember. Therefore, future studies could focus on using only “fruits and vegetables” category as its identification can be more culture-independent. Secondly, it is also possible that our shopping task design did not have enough sensitivity to detect a discrete difference between task difficulties. Future studies could therefore investigate the validation sensitivity of the task in detecting lessubtle differences – easy, medium, and hard.

We also found that the overall accuracy could be predicted by the MoCa, demonstrating that those participants that had higher initial MoCa scores were more accurate during the shopping game, regardless of the session and condition. Just like for Starry Night, there was no difference between groups, which could mean that engaging in any cognitive training improves accuracy. The analysis revealed no difference in errors between groups and conditions, however, there was a difference in errors between conditions. Participants committed significantly fewer errors when the workload was low (4-items vs 5-items, 5-items vs 7-items), confirming the difficulty of the task varied.

The lack of significant results in the shopping task might indicate high between-subject variation, therefore individualized approach to cognitive training could provide better results. Previous studies indicated that individual differences in cognitive skill measures taken at the baseline may predict variations in training outcomes although the direction of that relationship remains inconclusive. The results of this study demonstrated that a high MoCa score measured at the baseline can predict only a higher accuracy score in the transfer task but no other measures. This result might be due to the ceiling effect as there was not much variation in the scores given that this study recruited only healthy participants. The reliability and sensitivity of MoCa applies rather to a clinical sample, where MoCa can predict post-training far transfer of the skill rather than the near transfer (Weng et al., 2019). The follow-up assessment one month after the training revealed that effects were maintained for the shopping task. However other training gains did not persist. This might suggest that in order to maintain training benefits, it is important to engage in cognitive training continuously.

Our shopping task contributes to the understanding of application of immersive technologies in CCT transfer assessment. Although our shopping task does not capture transfer to the non-trained domains, it measures how the effects of CCT generalize to real-life scenarios. Employing immersive technologies allows for the creation of more ecologically valid scenarios that bridge the gap between the naturalness of the response and the controllability of the measure. Previous CCT studies yielded improvement in the trained tasks, however showed limited evidence of transfer to other untrained domains or outside of the training paradigm (Ball et al., 2002; Boot et al., 2010; Green and Bavelier, 2008; Lee et al., 2012; Luis-Ruiz et al., 2020; Owen et al., 2010; Willis et al., 2006). Therefore, although behavioural and neural changes can be observed from training, these changes have not been shown to consistently translate to meaningful improvements outside of the training paradigm. However previous research has been criticised for not implementing ecologically valid outcome measures and transfer tasks (Zhang et al., 2019). Although there are some studies that employed immersive technologies to deliver cognitive training (Liao et al., 2019; Mrakic-Sposta et al., 2018; Optale et al., 2010; Park et al., 2019; Schreiber, 1999), they did not use it to measure real-life transfer effects. The advantage of using AR for this task, over VR, is that it can be potentially more beneficial in clinical settings too. Thanks to AR'see-through nature, it allowsafe navigation without the risk of bumping into walls or objects. AR allows better interaction with other users (Xiong et al., 2021) and observers, therefore a patient can maintain a visual contact with a therapist what providesafety cues during a session (Roberts et al., 2016). AR also offers a potential to minimise the risk of cybersickness during simulation (Hughes et al., 2020). The Magic Leap headset was also easily combined with fNIRS thanks to its portable and small and lightweight design, which is an important practical concern when considering the involvement of mixed reality and on-head physiological measurements. In particular, the lightweight design minimises the risk or the probe displacement and therefore noise in the recorded brain data.

H5: Participants will recognize the impact of ADDP training and have generally positive attitudes toward the technology after participating in the programme

Participants were mostly positive for both groups indicating that the majority of people enjoyed both apps. Participants expressed positive emotions (joy, trust and anticipation), which were higher in frequency than negative emotions. The most frequent words used by both groups describing the app were "easy" and "good", however, while the experimental group reported that they mostly enjoyed the app, the control group described the placebo app as boring and repetitive. In our related work using public and patient involvement methods, we report further on the perspectives that members of our studied demographic had about cognitive training technologies for managing cognitive decline (Harrington et al., 2022).

7. Limitations and future research

We observe four main unavoidable limitations, from our study design, that could be investigated by future work. Firstly the recruitment was affected by COVID-19, and so the sample was not random and only single-blinded. Future work could employ a double-blinded randomized controlled trial design to minimise the risk of bias. Our data showed that participants who were assigned to the control group spent significantly more time on their daily training than the experimental group, even though they reported it was more repetitive and boring. This could be because both groups were recruited separately and during different times of the year. While the experimental group was recruited during summer (June-September), the control group was recruited in winter (October-February). The seasons, including the amount of daylight, could have played a role in this factor, as could have the changing restrictions in place due to COVID in those different time periods. Further, it's possible that the recruitment team emphasised the importance of sustained engagement after observing the engagement levels in the first phase (experimental group). However, even though the control groups played more than the experimental group, in general, the experimental group performed better at their post-training assessment, which might indicate that the time played in the control group may have yielded less benefit for participants.

Secondly, the chosen longitudinal design of our study are not optimal to gather a some kinds of data (e.g. fNIRS). In particular, the number of repetitions and duration of trails may not have been sufficient to detect effect sizes. This approach was taken to reduce the length of the session and therefore discomfort for participants. Future work could introduce additional sessions dedicated to technology training and familiarisation to shorten the session duration and minimise participant fatigue. Alternatively, this could be done by reducing the number of outcome measures and focusing only on those which showed promise of demonstrating statistically significant difference. Further, due to the ethics-related decisions in our non-clinical department, the study involved only healthy participants that are concerned about their memory decline, however do not have an official diagnosis of Alzheimer's disease or a cognitive impairment. This could explain why we did not find significant effects of some conditions, and the ceiling effects observed. Future work should ideally also involve a clinical population to test the efficacy of the cognitive training.

Thirdly, the design of our experiment was not optimal in relation to fNIRS analysis, particularly during the shopping task. In our protocol, a BoxCar function was employed to model the hemodynamic response. This approach was motivated by the variable duration of tasks within our experimental conditions which could last between 60 and 70 s. To maintain consistency, we trimmed the longer tasks to a 30-second window. Consequently, the hemodynamic responses observed could deviate from the canonical HRF often assumed in fNIRS research, both in shape and timing. Our approach focused on the detection of hemodynamic changes associated with the presence of the stimulus, rather than the detailed temporal dynamics of the task, which may not be sensitive enough to detect small effect sizes. This methodological choice introduces the possibility that our analysis might not have been sensitive enough to detect significant effects and therefore we saw some discrepancies between neurological and behavioural results. Future investigation should aim to improve the experimental design to better control for task duration.

Lastly, it is worth noting that our transfer task was measuring memory performance, therefore capturing only an applied version of near-transfer, where far transfer is typically considered to test alternative cognitive function to the ones being trained. Although one of our aims in this approach was to test the feasibility and usability of the method of combining fNIRS and AR technologies, future studies could use this technique to create further far-transfer tasks to capture performance within an untrained cognitive domain.

8. Conclusion

The key challenges for evaluating the success of Computerised Cognitive Training are 1) the lack of objective measures of it sustained success, and 2) the lack of empirical evaluation methods that can be applied in more ecologically valid conditions. We sought to overcome these key challenges using portable neuroimaging techniques, and in more ecologically valid conditions created using virtual reality. The aim of this work was to conduct a substantial study that collected pre-training, post-training, and follow-on data, from participants concerned about their cognitive decline having used CCT for a given period. The results were intended to provide evidence that might enable a clinical trial, with an increased number participants (including those living with cognitive decline), for an extended period of time.

Overall, the novel methods used in this study (portable fNIRS to observe mental workload, and augmented reality to study example transfer tasks) did allow us to identify some changes in task performance and working memory activity in participants with subjective memory decline, validating the structure of the study. These results were especially evident for those using dynamically adjusted computerized cognitive training, in comparison to a baseline control condition that did not adapt to challenge the user's actual cognitive capabilities. We found that dynamic-CCT group had decreased mental workload within the prefrontal cortex during the working memory precision task. Notably, this effect was also maintained at the follow-up, where lack of sustained impact is a common criticism of CCTs. Behavioural results demonstrated that localisation performance (remembering where an object was previously) increased after the training and was also maintained at the follow up for both groups that received CCT. These results indicate that CCTs can have a sustained impact on cognitive activity, as observed by fNIRS in our case, especially if they continue to adapt to challenge the user as they get more experienced with a game.

A common concern for evaluating CCTs is whether the impact, and sustained impact, transfer to more real-world tasks, for which we simulated a shopping task in augmented reality but in safe lab conditions. For the shopping task we found decreased mental workload after the training within the PFC, but the result was not maintained at the follow-up. However, the behavioural results again showed that effects were maintained at the follow-up. For both Starry Night and the shopping task, however, we found that participants improved significantly only in some distinct conditions. For shopping, for example, brand familiarity anecdotally played a role, which may relate to longer-term memory or further factors that need considering. Overall, our results are promising both for this combination of technologies to evaluate CCTs and their lasting impact. We hope that future work will be able to consider these approaches in clinical-scale studies with more participants and for more sustained periods of training.

CRediT authorship contribution statement

Aleksandra Landowska: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Max L. Wilson:** . **Michael P. Craven:** . **Kyle Harrington:** Conceptualization, Investigation, Methodology, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Aleksandra Landowska reports financial support was provided by European Union. Our industry partners Brain+ provided the CCT and led the European innovation project (ADDP) that funded the research programme. The work was performed as a preliminary study to provide a grounding and recommendations for ADDP to consider a future clinical trial, by testing the feasibility of the protocol, validity, and usability of employed instruments: 1) AD test battery Stary Night (an implementation of the cognitive misbinding task by Pertzov et al. (2012, 2013), who were also members of the ADDP research grant helping Brain+ to create Starry Night, 2) the mobile augmented reality the common shopping memory task (developed by the University of Nottingham), and 3) the ADDP intervention CCT app for cognitive screening. Further, for the project partners, one aim was to gather feedback from participants.

Data availability

The raw data, except of videos, that support the findings of this study are available on request from the cosrsponding author.

Acknowledgements

This research was funded by the European Union's Horizon 2020 research and innovation programme – Fast Track to Innovation (FTI), grant agreement 820636. We thank the recruitment company, Patricia Turner, for their efforts and guidance, as well as the members of Join Dementia Research. We thank the participants for their active engagement throughout the process. More than normal appreciation is due to the first author of the manuscript, from the research team. Dr. Landowska managed this research through challenging COVID restrictions, developed and implemented complex emergency health and safety procedures to work with participants when able, and managed a sustained effort to complete the research study within the extended funding period. We thank also our industry Brain+ who provided free licenses to participants as additional post-study remuneration for participating in the research. This research was also supported by the National Institute for Health and Care Research (NIHR) Nottingham Biomedical Research Centre and the NIHR MindTech MedTech Co-operative. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health and Social Care in England.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijhcs.2023.103206.

References

Acevedo, B.P., Dattatri, N., Le, J., Lappinga, C., Collins, N.L., 2022. Cognitive training with neurofeedback using fNIRS improves cognitive function in older adults. Int J Environ Res 19 (9). https://doi.org/10.3390/ijerph19095531.

Aksoy, E., Izzetoglu, K., Baysoy, E., Agrali, A., Kitapcioglu, D., Onaral, B., 2019a. Performance monitoring via functional near infrared spectroscopy for virtual reality based basic life support training. Front. Neurosci. 13 https://doi.org/10.3389/fnins.2019.01336.

Aksoy, E., Izzetoglu, K., Onaral, B., Kitapcioglu, D., Sayali, M.E., Guven, F., 2019b. Assessing Correlation Between Virtual Reality Based Serious Gaming Performance and Cognitive Workload Changes via Functional Near Infrared Spectroscopy. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11580 LNAI. https://doi.org/10.1007/978-3-030-22419-6_26.

Andreu-Perez, A.R., Kiani, M., Andreu-Perez, J., Reddy, P., Andreu-Abela, J., Pinto, M., Izzetoglu, K., 2021. Single-trial recognition of video gamer's expertise from brain haemodynamic and facial emotion responses. Brain Sci 11 (1). https://doi.org/10.3390/brainsci11010106.

Argyle, E.M., Marinescu, A., Wilson, M.L., Lawson, G., Sharples, S., 2021. Physiological indicators of task demand, fatigue, and cognition in future digital manufacturing environments. Int. J. Man-Mach. Stud. 145 https://doi.org/10.1016/j.ijhcs.2020.102522.

Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B., 2012. Optical brain monitoring for operator training and mental workload assessment. Neuroimage 59 (1). https://doi.org/10.1016/j.neuroimage.2011.06.023.

Ball, K., Berch, D.B., Helmers, K.F., Jobe, J.B., Leveck, M.D., Marsiske, M., Morris, J.N., Rebok, G.W., Smith, D.M., Tennstedt, S.L., Unverzagt, F.W., 2002. Effects of cognitive training interventions with older adults: a randomized controlled trial. Jama 288 (18), 2271–2281.

Barker, J.W., Aarabi, A., Huppert, T.J., 2013. Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. Biomed Opt. https://doi.org/10.1364/boe.4.001366.

Basak, C., Boot, W.R., Voss, M.W., Kramer, A.F., 2008. Can Training in a Real-Time Strategy Video Game Attenuate Cognitive Decline in Older Adults? Psychol Aging 23 (4), 765–777. https://doi.org/10.1037/a0013494.

Bates, D., Mächler, M., Zurich, E., Bolker, B.M., Walker, S.C., 2012. Fitting linear mixed-effects models using lme4. J Stat Softw 67, 1–48.

Bays, P.M., Husain, M., 2008. Dynamic shifts of limited working memory resources in human vision. Science 321 (5890). https://doi.org/10.1126/science.1158023.

Bays, P.M., Catalao, R.F.G., Husain, M., 2009. The precision of visual working memory iset by allocation of a shared resource. J Vis 9 (10). https://doi.org/10.1167/9.10.7.
Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Series B Stat. Methodol. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L., 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat. Space Environ. Med. 78 (5 II).
Bhatt, S., Agrali, A., McCarthy, K., Suri, R., Ayaz, H., 2018. Web usability testing with concurrent fnirs and eye tracking. In Neuroergonomics: The Brain at Work and in Everyday Life. https://doi.org/10.1016/B978-0-12-811926-6.00030-0.
Boas, D.A., Elwell, C.E., Ferrari, M., Taga, G., 2014. Twenty years of functional near-infrared spectroscopy: introduction for the special issue. Neuroimage 85 Pt 1, 1–5. https://doi.org/10.1016/j.neuroimage.2013.11.033.
Boot, W.R., Basak, C., Erickson, K.I., Neider, M., Simons, D.J., Fabiani, M., Gratton, G., Voss, M.W., Prakash, R., Lee, H., Low, K.A., 2010. Transfer of skill engendered by complex task training under conditions of variable priority. Acta Psychol. 135 (3), 349–357.
Brehmer, Y., Rieckmann, A., Bellander, M., Westerberg, H., Fischer, H., Bäckman, L., 2011. Neural correlates of training-related working-memory gains in old age. Neuroimage 58 (4). https://doi.org/10.1016/j.neuroimage.2011.06.079.
Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M., Bolker, B.M., 2017. glmmTMB balancespeed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J 9 (2), 378–400. https://doi.org/10.32614/rj-2017-066.
Causse, M., Chua, Z., Peysakhovich, V., del Campo, N., Matton, N., 2017. Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. Sci Rep 7 (1). https://doi.org/10.1038/s41598-017-05378-x.
Causse, M., Lepron, E., Mandrick, K., Peysakhovich, V., Berry, I., Callan, D., Rémy, F., 2022. Facing successfully high mental workload and stressors: an fMRI study. Hum Brain Mapp 43 (3). https://doi.org/10.1002/hbm.25703.
Chang, L., Løhaugen, G.C., Andres, T., Jiang, C.S., Douet, V., Tanizaki, N., Walker, C., Castillo, D., Lim, A., Skranes, J., Otoshi, C., Miller, E.N., Ernst, T.M., 2017. Adaptive working memory training improved brain function in human immunodeficiency virus–seropositive patients. Ann. Neurol. 81 (1) https://doi.org/10.1002/ana.24805.
Clark, C.M., Lawlor-Savage, L., Goghari, V.M., 2017. Functional brain activation associated with working memory training and transfer. Brain Res 334. https://doi.org/10.1016/j.bbr.2017.07.030.
Cowan, N., 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav. Brain Sci. 24 (1) https://doi.org/10.1017/S0140525x01003922.
Cui, X., Bray, S., Reiss, A.L., 2010. Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. Neuroimage 49 (4), 3039–3046. https://doi.org/10.1016/j.neuroimage.2009.11.050.
Curtin, A., Ayaz, H., 2018. The Age of Neuroergonomics: towards Ubiquitous and Continuous Measurement of Brain Function with fNIRS. Japanese Psychological Research 60 (4). https://doi.org/10.1111/jpr.12227.
di Domenico, S.I., Rodrigo, A.H., Ayaz, H., Fournier, M.A., Ruocco, A.C., 2015. Decision-making conflict and the neural efficiency hypothesis of intelligence: a functional near-infrared spectroscopy investigation. Neuroimage 109. https://doi.org/10.1016/j.neuroimage.2015.01.039.
Fabio, R.A., Billeci, L., Crifaci, G., Troise, E., Tortorella, G., Pioggia, G., 2016. Cognitive training modifies frequency EEG bands and neuropsychological measures in Rett syndrome. Res Dev Disabil 53–54. https://doi.org/10.1016/j.ridd.2016.01.009.
Fishburn, F., Norr, M., 2014. Sensitivity of fNIRS to cognitive state and load. Frontiers in Human … 8, 76 https://doi.org/10.3389/fnhum.2014.00076 (February).
Fishburn, F.A., Ludlum, R.S., Vaidya, C.J., Medvedev, A.v., 2019. Temporal Derivative Distribution Repair (TDDR): a motion correction method for fNIRS. Neuroimage. https://doi.org/10.1016/j.neuroimage.2018.09.025.
Flak, M.M., Hol, H.R., Hernes, S.S., Chang, L., Engvig, A., Bjuland, K.J., Pripp, A., Madsen, B.O., Knapskog, A.B., Ulstein, I., Lona, T., Skranes, J., Løhaugen, G.C.C., 2019. Adaptive computerized working memory training in patients with mild cognitive impairment. A randomized double-blind active controlled trial. Front Psychol 10. https://doi.org/10.3389/fpsyg.2019.00807 (APR).
Fougnie, D., Asplund, C.L., Marois, R., 2010. What are the units of storage in visual working memory? J Vis 10 (12). https://doi.org/10.1167/10.12.27.
Fuchs, E., Flügge, G., 2014. Adult neuroplasticity: more than 40 years of research. Neural Plast. (2014) https://doi.org/10.1155/2014/541870.
Gagnon, L.G., Belleville, S., 2012. Training of attentional control in mild cognitive impairment with executive deficits: results from a double-blind randomised controlled study. Neuropsychol Rehabil 22 (6). https://doi.org/10.1080/09602011.2012.691044.
Gagnon, L., Perdue, K., Greve, D.N., Goldenholz, D., Kaskhedikar, G., Boas, D.A., 2011. Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state-space modeling. Neuroimage 56 (3), 1362–1371.
Gandelman-Marton, R., Aichenbaum, S., Dobronevsky, E., Khaigrekht, M., Rabey, J.M., 2017. Quantitative EEG after brain stimulation and cognitive training in Alzheimer disease. J. Clin. Neurophysiol. 34 (1) https://doi.org/10.1097/WNP.0000000000000301.
Ge, R., Wang, Z., Yuan, X., Li, Q., Gao, Y., Liu, H., Fan, Z., Bu, L., 2021. The effects of two game interaction modes on cortical activation in subjects of different ages: a functional near-infrared spectroscopy study. IEEE Access 9. https://doi.org/10.1109/ACCESS.2021.3050210.
Georgiou, Y., Kyza, E.A., 2017. The development and validation of the ARI questionnaire: an instrument for measuring immersion in location-based augmented reality settings. Int. J. Hum. Comput. Stud. 98, 24–37.
Gonneaud, J., Piolino, P., Lecouvey, G., Madeleine, S., Orriols, E., Fleury, P., Eustache, F., Desgranges, B., 2014. Assessing prospective memory in young healthy adults using virtual reality. Virtual Reality: People with Special Needs.
Grabner, R.H., Neubauer, A.C., Stern, E., 2006. Superior performance and neural efficiency: the impact of intelligence and expertise. Prog Brain Res 69 (4). https://doi.org/10.1016/j.brainresbull.2006.02.009.
Green, C.S., Bavelier, D., 2008. Exercising your brain: a review of human brain plasticity and training-induced learning. Psychol. Aging 23 (4), 692.
Grogan, J.P., Husain, M., Manohar, S.G., Fallon, S.J., Zokaei, N., Husain, M., Zokaei, N., Manohar, S.G., Coulthard, E.J., Coulthard, E.J., 2016. A new toolbox to distinguish the sources of spatial memory error. J Vis 20 (13). https://doi.org/10.1167/jov.20.13.6.
Harrington, K., Craven, M.P., Wilson, M.L., Landowska, A., 2022. Perceptions of cognitive training games and assessment technologies for dementia: acceptability study with patient and public involvement workshops. JMIR Serious Games 10 (2), e32489.
Heinzel, S., Lorenz, R.C., Pelz, P., Heinz, A., Walter, H., Kathmann, N., Rapp, M.A., Stelzel, C., 2016. Neural correlates of training and transfer effects in working memory in older adults. Neuroimage 134. https://doi.org/10.1016/j.neuroimage.2016.03.068.
Hill, A.P., Bohil, C.J., 2016. Applications of Optical Neuroimaging in Usability Research. Ergonomics in Design 24 (2). https://doi.org/10.1177/1064804616629309.
Hirshfield, L.M., Solovey, E.T., Girouard, A., Kebinger, J., Jacob, R.J.K., Sassaroli, A., Fantini, S., 2009. Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy. In: Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/1518701.1519035.
Hirshfield, L.M., Gulotta, R., Hirshfield, S., Sam, H., Russell, M., Ward, R., Williams, T., Jacob, R., 2011. This is your brain on interfaces: enhancing usability testing with functional near-infrared spectroscopy. In: Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/1978942.1978996.
Hirshfield, L., Bobko, P., Barelka, A., Sommer, N., Velipasalar, S., 2019. Toward Interfaces that Help Users Identify Misinformation Online: using fNIRS to Measure Suspicion. Augmented Human Research 4 (1). https://doi.org/10.1007/s41133-019-0011-8.
Hu, M., Wu, X., Shu, X., Hu, H., Chen, Q., Peng, L., Feng, H., 2021. Effects of computerised cognitive training on cognitive impairment: a meta-analysis. J. Neurol. 268 (5) https://doi.org/10.1007/s00415-019-09522-7.
Hughes, C.L., Fidopiastis, C., Stanney, K.M., Bailey, P.S., Ruiz, E., 2020. The psychometrics of cybersickness in augmented reality. Front. Virtual Real. 1, 602954.
Kable, J.W., Caulfield, M.K., Falcone, M., McConnell, M., Bernardo, L., Parthasarathi, T., Cooper, N., Ashare, R., Audrain-McGovern, J., Hornik, R., Diefenbach, P., Lee, F.J., Lerman, C., 2017. No effect of commercial cognitive training on brain activity, choice behavior, or cognitive performance. J. Neurosci. 37 (31) https://doi.org/10.1523/JNEUROSCI.2832-16.2017.
Kanatschnig, T., Wood, G., & Kober, S.E. (2021). The Potential of Functional Near-Infrared Spectroscopy (fNIRS) for Motion-Intensive Game Paradigms. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13134 LNCS. https://doi.org/10.1007/978-3-030-92182-8_9.
Kelly, C., Foxe, J.J., Garavan, H., 2006. Patterns of normal human brain plasticity after practice and their implications for neurorehabilitation. Arch. Phys. Med. 87 (12 SUPPL) https://doi.org/10.1016/j.apmr.2006.08.333.
Kelly, M.E., Loughrey, D., Lawlor, B.A., Robertson, I.H., Walsh, C., Brennan, S., 2014. The impact of cognitive training and mental stimulation on cognitive and everyday functioning of healthy older adults: a systematic review and meta-analysis. Ageing Res. 15 (1) https://doi.org/10.1016/j.arr.2014.02.004.
Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G., 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. Int. J. Aerosp. Psychol. 3 (3), 203–220.
Khaleghi, A., Aghaei, Z., Mahdavi, M.A., 2021. A gamification framework for cognitive assessment and cognitive training: qualitative study. J. Med. Internet Res. 9 (2) https://doi.org/10.2196/21900.
Kim, H., Chey, J., Lee, S., 2017. Effects of multicomponent training of cognitive control on cognitive function and brain activation in older adults. Neurosci. Res. 124 https://doi.org/10.1016/j.neures.2017.05.004.
Kirilina, E., Jelzow, A., Heine, A., Niessing, M., Wabnitz, H., Brühl, R., Ittermann, B., Jacobs, A.M., Tachtsidis, I., 2012. The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. Neuroimage 61 (1). https://doi.org/10.1016/j.neuroimage.2012.02.074.
Kueider, A.M., Parisi, J.M., Gross, A.L., Rebok, G.W., 2012. Computerized cognitive training with older adults: a systematic review. PLoS ONE 7 (7). https://doi.org/10.1371/journal.pone.0040588.
Kutafina, E., Heiligers, A., Popovic, R., Brenner, A., Hankammer, B., Jonas, S.M., Mathiak, K., Zweerings, J., 2021. Tracking of mental workload with a mobile EEG sensor. Sensors 21 (15). https://doi.org/10.3390/s21155205.
Lampit, A., Hallock, H., Suo, C., Naismith, S.L., Valenzuela, M., 2015. Cognitive training-induced short-term functional and long-term structural plastic change is related to gains in global cognition in healthy older adults: a pilot study. Front. Aging 7. https://doi.org/10.3389/fnagi.2015.00014. MAR.
Lee, H., Boot, W.R., Basak, C., Voss, M.W., Prakash, R.S., Neider, M., Erickson, K.I., Simons, D.J., Fabiani, M., Gratton, G., Low, K.A., 2012. Performance gains from directed training do not transfer to untrained tasks. Acta Psychol. 139 (1), 146–158.

Lenth, R.v. (2022). *Estimated Marginal Means, aka Least-Squares Means*. https://cran.r-project.org/package=emmeans.
Li, B.Y., He, N.Y., Qiao, Y., Xu, H.M., Lu, Y.Z., Cui, P.J., Ling, H.W., Yan, F.H., Tang, H.D., Chen, S.di., 2019. Computerized cognitive training for Chinese mild cognitive impairment patients: a neuropsychological and fMRI study. Neuroimage Clin. 22 https://doi.org/10.1016/j.nicl.2019.101691.
Liang, Y., Pertzov, Y., Nicholas, J.M., Henley, S.M.D., Crutch, S., Woodward, F., Leung, K., Fox, N.C., Husain, M., 2016. Visual short-term memory binding deficit in familial Alzheimer's disease. Cortex. https://doi.org/10.1016/j.cortex.2016.01.015.
Liao, Y.Y., Chen, I.H., Lin, Y.J., Chen, Y., Hsu, W.C., 2019. Effects of virtual reality-based physical and cognitive training on executive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. Front. Aging Neurosci. 11, 162.
Lim, J., Wu, W.chau, Wang, J., Detre, J.A., Dinges, D.F., Rao, H, 2010. Imaging brain fatigue from sustained mental workload: an ASL perfusion study of the time-on-task effect. Neuroimage 49 (4). https://doi.org/10.1016/j.neuroimage.2009.11.020.
Luis-Ruiz, S., Caldú, X., Sánchez-Castañeda, C., Pueyo, R., Garolera, M., Jurado, M.Á., 2020. Is cognitive training an effective tool for improving cognitive function and real-life behaviour in healthy children and adolescents? A systematic review. Neurosci. Biobehav. Rev. (116) https://doi.org/10.1016/j.neubiorev.2020.06.019.
Lukanov, K., Maior, H.A., Wilson, M.L., 2016. Using fNIRS in usability testing: understanding the effect of web form layout on mental workload. In: Conference on Human Factors in Computing Systems Proceedings. https://doi.org/10.1145/2858036.2858236.
Ma, W.J., Husain, M., Bays, P.M., 2014. Changing concepts of working memory. Nat. Neurosci. 17 (3) https://doi.org/10.1038/nn.3655.
Mahncke, H.W., Kim, S.J., Rose, A., Stasio, C., Buckley, P., Caroff, S., Duncan, E., Yasmin, S., Jarskog, L.F., Lamberti, J.S., Nuechterlein, K., Strassnig, M., Velligan, D., Ventura, J., Walker, T., Stroup, T.S., Keefe, R.S.E., 2019. Evaluation of a plasticity-based cognitive training program in schizophrenia: results from the eCaesar trial. Schizophr. Res. 208 https://doi.org/10.1016/j.schres.2019.03.006.
Maior, H.A., Pike, M., Sharples, S., Wilson, M.L., 2015. Examining the reliability of using fNIRS in realistic HCI settings for spatial and verbal tasks. In: *Conference on Human Factors in Computing Systems Proceedings, 2015-April*. https://doi.org/10.1145/2702123.2702315.
Mancini, M., Cherubino, P., di Flumeri, G., Cartocci, G., Martinez, A., Sanchez, A., Santillo, C., Modica, E., Vozzi, A., Ronca, V., Trettel, A., Borghini, G., Babiloni, F., 2022. Neuroscientific methods for exploring user perceptions while dealing with mobile advertising: a novel and integrated approach. Front. Neuroergonomics 3. https://doi.org/10.3389/fnrgo.2022.835648.
Manohar, S.G., Zokaei, N., Fallon, S.J., Vogels, T.P., Husain, M., 2019. Neural mechanisms of attending to items in working memory. Neurosci. Biobehav. Rev. 101, 1–12.
Mathy, F., Feldman, J., 2012. Chunking and data compression in short-term memory. Cognition 122 (3). https://doi.org/10.1016/j.cognition.2011.11.003.
McKendrick, R., Mehta, R., Ayaz, H., Scheldrup, M., Parasuraman, R., 2017. Prefrontal hemodynamics of physical activity and environmental complexity during cognitive work. Hum. Factors 59 (1). https://doi.org/10.1177/0018720816675053.
Midha, S., Maior, H.A., Wilson, M.L., Sharples, S., 2021. Measuring mental workload variations in office work tasks using fNIRS. Int. J. Man-Mach. Stud. 147 https://doi.org/10.1016/j.ijhcs.2020.102580.
Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63 (2) https://doi.org/10.1037/h0043158.
Miró-Padilla, A., Bueichekú, E., Ventura-Campos, N., Flores-Compañ, M.J., Parcet, M.A., Ávila, C., 2019. Long-term brain effects of N-back training: an fMRI study. Brain Imaging Behav. 13 (4) https://doi.org/10.1007/s11682-018-9925-x.
Mrakic-Sposta, S., Di Santo, S.G., Franchini, F., Arlati, S., Zangiacomi, A., Greci, L., Moretti, S., Jesuthasan, N., Marzorati, M., Rizzo, G., Sacco, M., 2018. Effects of combined physical and cognitive virtual reality-based training on cognitive impairment and oxidative stress in MCI patients: a pilot study. Front. Aging Neurosci. 10, 282.
Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L, Chertkow, H., 2005. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J. Am. Geriatr. Soc. 53 (4), 695–699.
Nieuwhof, F., Reelick, M., Rikkert, M.O., Mirelman, A., Hausdorff, J., Claassen, J., 2012. Wireless fNIRS for neuroimaging during dual task walking and obstacle negotiation in the elderly: feasible, reliable and valid? Eur. Geriatr. Med. 3, S54 https://doi.org/10.1016/j.eurger.2012.07.071.
Oberauer, K., Lewandowsky, S., Awh, E., Brown, G.D.A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G.J., Hurlstone, M.J., Ma, W.J., Morey, C.C., Nee, D.E., Schweppe, J., Vergauwe, E., Ward, G., 2018. Benchmarks for models of short-term and working memory. Psychol. Bull. 144 (9) https://doi.org/10.1037/bul0000153.
Olfers, K.J.F., Band, G.P.H., 2018. Game-based training of flexibility and attention improves task-switch performance: near and far transfer of cognitive training in an EEG study. Psychol. Res. 82 (1) https://doi.org/10.1007/s00426-017-0933-z.
Optale, G., Urgesi, C., Busato, V., Marin, S., Piron, L., Priftis, K., Gamberini, L., Capodieci, S., Bordin, A., 2010. Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. Neurorehabilit. Neural Repair 24 (4), 348–357.
Owen, A.M., Hampshire, A., Grahn, J.A., Stenton, R., Dajani, S., Burns, A.S., Howard, R.J., Ballard, C.G., 2010. Putting brain training to the test. Nature 465 (7299), 775–778.
Pagnotta, M., Jacobs, D.M., de Frutos, P.L., Rodríguez, R., Ibáñez-Gijón, J., Travieso, D., 2021. Task difficulty and physiological measures of mental workload in air traffic control: a scoping review. Ergonomics. https://doi.org/10.1080/00140139.2021.2016998.
Park, E., Yun, B.J., Min, Y.S., Lee, Y.S., Moon, S.J., Huh, J.W., Cha, H., Chang, Y., Jung, T.D., 2019. Effects of a mixed reality-based cognitive training system compared to a conventional computer-assisted cognitive training system on mild cognitive impairment: a pilot study. Cogn. Behav. Neurol. 32 (3), 172–178.
Pavisic, I.M., Nicholas, J.M., Pertzov, Y., O'Connor, A., Liang, Y., Collins, J.D., Lu, K., Weston, P.S.J., Ryan, N.S., Husain, M., Fox, N.C., Crutch, S.J., 2021. Visual short-term memory impairments in presymptomatic familial Alzheimer's disease: a longitudinal observational study. Neuropsychologia 162. https://doi.org/10.1016/j.neuropsychologia.2021.108028.
Payzieva, S., Maxmudova, D., 2014. NIRS study of the effects of computerized brain training games for cognitive rehabilitation of major depressive disorder patients in remission: a pilot study. Annu. Rev. Cyber Therapy Telemed. 12.
Peich, M.C., Husain, M., Bays, P.M., 2013. Age-related decline of precision and binding in visual working memory. Psychol. Aging 28 (3). https://doi.org/10.1037/a0033236.
Pertzov, Y., Dong, M.Y., Peich, M.C., Husain, M., 2012. Forgetting what was where: the fragility of object-location bnding. PLoS ONE 7 (10). https://doi.org/10.1371/journal.pone.0048214.
Pertzov, Y., Miller, T.D., Gorgoraptis, N., Caine, D., Schott, J.M., Butler, C., Husain, M., 2013. Binding deficits in memory following medial temporal lobe damage in patients with voltage-gated potassium channel complex antibody-associated limbic encephalitis. Brain 136 (8). https://doi.org/10.1093/brain/awt129.
Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., Burgess, P.W., 2020. The present and future use of functional near-infrared spectroscopy (fnirs) for cognitive neuroscience. Ann. N. Y. Acad. Sci. 1464 (1) https://doi.org/10.1111/nyas.13948.
Piper, S.K., Krueger, A., Koch, S.P., Mehnert, J., Habermehl, C., Steinbrink, J., Obrig, H., Schmitz, C.H., 2014. A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. Neuroimage 85, 64–71. https://doi.org/10.1016/j.neuroimage.2013.06.062.
Qu, H., Shan, Y., Liu, Y., Pang, L., Fan, Z., Zhang, J., Wanyan, X., 2020. Mental workload classification method based on EEG independent component features. Appl. Sci. 10 (9) https://doi.org/10.3390/app10093036.
Rahman, M.A., Siddik, A.B., Ghosh, T.K., Khanam, F., Ahmad, M., 2020. A Narrative review on clinical applications of fNIRS. J. Digit. Imaging 33 (5). https://doi.org/10.1007/s10278-020-00387-1.
RCoreTeam. (2021). *A language and environment for statistical computing. R Foundation for Statistical Computing*. 2013. https://www.r-project.org/.
Rizkalla, M.N., 2018. Cognitive training in the elderly: a randomized trial to evaluate the efficacy of a self-administered cognitive training program. Aging Ment. Health 22 (10). https://doi.org/10.1080/13607863.2015.1118679.
Roberts, D.J., Fairchild, A.J., Campion, S.P., Garcia Jimenez, Wolff, R., 2016. Bringing the client and therapist together in virtual reality telepresence exposure therapy.
Rolinski, M., Zokaei, N., Mackay, C.E., Husain, M., Hu, M.T.M., 2015. Pattern of working memory deficit in REM sleep behaviour disorder is the same as in Parkinson's disease. Mov. Disord. 30.
Rose, N.S., LaRocque, J.J., Riggall, A.C., Gosseries, O., Starrett, M.J., Meyering, E.E., Postle, B.R., 2016. Reactivation of latent working memories with transcranial magnetic stimulation. Science 354 (6316). https://doi.org/10.1126/science.aah7011.
Rosen, A.C., Sugiura, L., Kramer, J.H., Whitfield-Gabrieli, S., Gabrieli, J.D., 2011. Cognitive training changes hippocampal function in mild cognitive impairment: a pilot study. J. Alzheimers Dis. 26 (SUPPL. 3) https://doi.org/10.3233/JAD-2011-0009.
Sala, G., Gobet, F., 2019. Cognitive training does not enhance general cognition. Trends Cogn. Sci. (Regul. Ed.) 23 (1). https://doi.org/10.1016/j.tics.2018.10.004.
Sala, G., Deniz Aksayli, N., Semir Tatlidil, K., Tatsumi, T., Gondo, Y., Gobet, F., 2019. Near and far transfer in cognitive training: a second-order meta-analysis. Collabra Psychol. 5 (1) https://doi.org/10.1525/collabra.203.
Santosa, H., Zhai, X., Fishburn, F., Huppert, T., 2018. The NIRS brain analyzIR toolbox. Algorithms. https://doi.org/10.3390/A11050073.
Schreiber, M., 1999. Potential of an interactive computer-based training in the rehabilitation of dementia: An initial study. Neuropsychol. Rehabil. 9 (2), 155–167.
So, W.K.Y., Wong, S.W.H., Mak, J.N., Chan, R.H.M, 2017. An evaluation of mental workload with frontal EEG. PLoS ONE 12 (4). https://doi.org/10.1371/journal.pone.0174949.
Soldan, A., Pettigrew, C., Cai, Q., Wang, J., Wang, M.C., Moghekar, A., Miller, M.I., Albert, M., 2017. Cognitive reserve and long-term change in cognition in aging and preclinical Alzheimer's disease. Neurobiol. Aging 60, 164–172. https://doi.org/10.1016/j.neurobiolaging.2017.09.002.
Solovey, E.T., Girouard, A., Chauncey, K., Hirshfield, L.M., Sassaroli, A., Zheng, F., Fantini, S., Jacob, R.J.K., 2009. Using fNIRS brain sensing in realistic HCI settings. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology - UIST '09, 157. https://doi.org/10.1145/1622176.1622207.
Stern, Y., 2002. What is cognitive reserve? theory and research application of the reserve concept. J. Int. Neuropsychol. Soc. 8 (3) https://doi.org/10.1017/S1355617702813248.
Stern, Y., 2012. Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol. https://doi.org/10.1016/S1474-4422(12)70191-6.
Strangman, G., Franceschini, M.A., Boas, D.A., 2003. Factors affecting the accuracy of near-infrared spectroscopy concentration calculations for focal changes in oxygenation parameters. Neuroimage. https://doi.org/10.1016/S1053-8119(03)00021-1.

Subramaniam, K., Luks, T.L., Garrett, C., Chung, C., Fisher, M., Nagarajan, S., Vinogradov, S., 2014. Intensive cognitive training in schizophrenia enhances working memory and associated prefrontal cortical efficiency in a manner that drives long-term functional gains. Neuroimage 99. https://doi.org/10.1016/j.neuroimage.2014.05.057.

Sukontapol, C., Kemsen, S., Chansirikarn, S., Nakawiro, D., Kuha, O., Taemeeyapradit, U., 2018. The effectiveness of a cognitive training program in people with mild cognitive impairment: a study in urban community. Asian J. Psychiatr. 35 https://doi.org/10.1016/j.ajp.2018.04.028.

Tabi, Y.A., Eriksen, U.D., Sørensen, N., Semság, P., Manohar, S., Husain, M., 2019. A portable tablet task for assessment of short-term memory. IBRO Rep. https://doi.org/10.1016/j.ibror.2019.07.776.

Tabi, Y.A., Manohar, S., Husain, M., 2020. Assessment of short-term memory using a gamified tablet-based task (91). Neurology 94 (15 Supplement), 91. http://n.neurology.org/content/94/15_Supplement/91.abstract.

Tabi, Y.A., Maio, M.R., Fallon, S.J., Udale, R., Dickson, S., Idris, M.I., Nobis, L., Manohar, S.G., Husain, M., 2021. Impact of processing demands at encoding, maintenance and retrieval in visual working memory. Cognition 214. https://doi.org/10.1016/j.cognition.2021.104758.

Tabi, Y.A., Maio, M.R., Attaallah, B., Dickson, S., Drew, D., Idris, M.I., Kienast, A., Klar, V., Nobis, L., Plant, O., Saleh, Y., Sandhu, T.R., Slavkova, E., Toniolo, S., Zokaei, N., Manohar, S.G., Husain, M., 2022. Vividness of visual imagery questionnaire scores and their relationship to visual short-term memory performance. Cortex 146. https://doi.org/10.1016/j.cortex.2021.10.011.

ten Brinke, L.F., Davis, J.C., Barha, C.K., Liu-Ambrose, T., 2017. Effects of computerized cognitive training on neuroimaging outcomes in older adults: a systematic review. BMC Geriatr. 17 (1) https://doi.org/10.1186/s12877-017-0529-x.

Tetlow, A.M., Edwards, J.D., 2017. Systematic literature review and meta-analysis of commercially available computerized cognitive training among older adults. J. Cogn. Enhanc. 1 (4) https://doi.org/10.1007/s41465-017-0051-2.

van Acker, B.B., Parmentier, D.D., Vlerick, P., Saldien, J., 2018. Understanding mental workload: from a clarifying concept analysis toward an implementable framework. Cogn. Technol. Work 20 (3). https://doi.org/10.1007/s10111-018-0481-3.

Vermeij, A., Kessels, R.P.C., Heskamp, L., Simons, E.M.F., Dautzenberg, P.L.J., Claassen, J.A.H.R., 2017. Prefrontal activation may predict working-memory training gain in normal aging and mild cognitive impairment. Brain Imaging Behav. 11 (1) https://doi.org/10.1007/s11682-016-9508-7.

Vermeir, J.F., White, M.J., Johnson, D., Crombez, G., van Ryckeghem, D.M.L., 2020. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. J. Med. Internet Res. 8 (3) https://doi.org/10.2196/18644.

Weng, W., Liang, J., Xue, J., Zhu, T., Jiang, Y., Wang, J., Chen, S., 2019. The transfer effects of cognitive training on working memory among Chinese older adults with mild cognitive impairment: a randomized controlled trial. Front. Aging 10 (JUL). https://doi.org/10.3389/fnagi.2019.00212.

Wickens, C.D., 2008. Multiple resources and mental workload. Hum. Factors 50 (3). https://doi.org/10.1518/001872008×288394.

Wickham, H., Chang, W., Wickham, M.H., 2016. Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics*. Version 2 (1), 1–189.

Willis, S.L., Tennstedt, S.L., Marsiske, M., Ball, K., Elias, J., Koepke, K.M., Morris, J.N., Rebok, G.W., Unverzagt, F.W., Stoddard, A.M., Wright, E., 2006. Long-term effects of cognitive training on everyday functional outcomes in older adults. Jama 296 (23), 2805–2814.

Wolff, M.J., Jochim, J., Akyürek, E.G., Stokes, M.G., 2017. Dynamic hidden states underlying working-memory-guided behavior. Nat. Neurosci. 20 (6) https://doi.org/10.1038/nn.4546.

Xiong, J., Hsiang, E.L., He, Z., Zhan, T., Wu, S.T., 2021. Augmented reality and virtual reality displays: emerging technologies and future perspectives. Light Sci. Appl. 10 (1), 216.

Yücel, M.A., Lühmann, A.v., Scholkmann, F., Gervain, J., Dan, I., Ayaz, H., Boas, D., Cooper, R.J., Culver, J., Elwell, C.E., Eggebrecht, A., Franceschini, M.A., Grova, C., Homae, F., Lesage, F., Obrig, H., Tachtsidis, I., Tak, S., Tong, Y., Wolf, M., 2021. Best practices for fNIRS publications. Neurophotonics 8 (01). https://doi.org/10.1117/1.nph.8.1.012101.

Young, M.S., Stanton, N.A., 2002. Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. Hum. Factors 44 (3). https://doi.org/10.1518/0018720024497709.

Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A., 2015. State of science: mental workload in ergonomics. Ergonomics 58 (1). https://doi.org/10.1080/00140139.2014.956151.

Zhang, H., Huntley, J., Bhome, R., Holmes, B., Cahill, J., Gould, R.L., Wang, H., Yu, X., Howard, R., 2019. Effect of computerised cognitive training on cognitive outcomes in mild cognitive impairment: a systematic review and meta-analysis. BMJ 9 (8). https://doi.org/10.1136/bmjopen-2018-027062.

Zokaei, N., Husain, M., 2019. Working memory in alzheimer's disease and Parkinson's disease. Curr. Top. Behav. Neurosci. 41 https://doi.org/10.1007/7854_2019_103.

Zokaei, N., Ning, S., Manohar, S., Feredoes, E., Husain, M., 2014. Flexibility of representational states in working memory. Front. Hum. Neurosci. 8 https://doi.org/10.3389/fnhum.2014.00853. November.

Zokaei, N., Burnett Heyes, S., Gorgoraptis, N., Budhdeo, S., Husain, M., 2015. Working memory recall precision is a more sensitive index than span. J. Neuropsychol. 9 (2) https://doi.org/10.1111/jnp.12052.

Zokaei, N., Gillebert, C.R., Chauvin, J.J., Gresch, D., Board, A.G., Rolinski, M., Hu, M.T., Nobre, A.C., 2021. Temporal orienting in Parkinson's disease. Eur. J. Neurosci. 53 (8) https://doi.org/10.1111/ejn.15114.