Can Relational Training Improve Cognitive Functioning and Subjective Complaints in Older Adults? A Randomised Controlled Trial Using SMART

Abstract

Slowing normal cognitive aging with cognitive training interventions has become a significant societal challenge. Recently, SMART (Strengthening Mental Abilities with Relational Training), a Relational Frame Theory-based intervention, has emerged as a promising intervention in the cognitive training literature. However, to our knowledge, the programme has not been applied in older adults, and there is need for validation using rigorous methods. This study compares the effects of SMART with Dual N-Back training in 105 older adults aged 62-93 with subjective cognitive complaints. Results showed that SMART successfully trained relational reasoning abilities, but there was no indication for transfer effects to neurological status, non-verbal IQ, depression, or quality of life. Both training groups demonstrated similar improvements in executive functioning. These findings suggest limited effectiveness of SMART in older adults, consistent with patterns in cognitive training literature where initial positive effects diminish after rigorous replication. This highlights the need for more rigorous studies on SMART, and on adaptations of the programme across diverse populations.

Keywords: SMART, Relational Frame Theory, cognitive training, normal cognitive aging

Introduction

In an increasingly aging population, many older adults face the prospect of a decline of cognitive functions with age, that is, normal cognitive aging, unrelated to disease or specific cognitive impairments (Harada et al., 2013). Cognitive abilities such as memory, processing speed, and executive functions tend to begin declining from as early as early adulthood, with an accelerated decline occurring from age 65 (Salthouse, 2019). There is, however, notable variability among individuals in both the onset and trajectory of age-related cognitive decline (Glisky, 2007; Lobo et al., 2021). Cognitive decline can have a significant impact on everyday life. This can range from simply remembering where you have left your keys, to medication use, telephone use, financial management (Bezdicek et al., 2023; Tucker-Drob, 2011), and driving performance (Clay et al., 2005; Derafshi et al., 2024). Technology- and internet use can also be impacted, which can lead to notable consequences like difficulties using e-health platforms (Czaja et al., 2013; Kamin & Lang, 2020). As such, the cognitive decline associated with natural aging is emerging as a significant societal challenge.

Given its widespread impact on daily functioning, there is a growing need for interventions that can delay or slow the progression of normal cognitive aging. Efforts to address cognitive aging extend across different domains, including pharmacological studies, physical exercise and nutritional interventions (Krivanek et al., 2021). Among these, a potential intervention to combat cognitive aging is the implementation of cognitive training (Giuli et al., 2016; Velloso et al., 2025). Cognitive training typically refers to structured, often computer-based exercises designed to improve performance on specific cognitive tasks. These programmes target specific abilities such as working memory, attention, or processing speed, with the goal of fostering improvements in those abilities (von Bastian et al., 2022).

A common example is the N-Back task, a working memory training where users have to observe and remember a stream of stimuli and respond when the current item matches one presented 'n' steps earlier (Jaeggi et al., 2008). Other examples include task-switching training or dual-tasks, where participants have to switch between tasks or execute two tasks simultaneously to improve multitasking and cognitive flexibility, or speed-of-processing training, aimed at improving perception and attentional control (von Bastian et al., 2022). These programs aim to elicit near transfer, where improvements extend to similar abilities, or far transfer, where gains influence more different, general cognitive skills (Gobet & Sala, 2023). However, many cognitive training programmes, including the ones described above, fail to reach the latter goal, as training effects often do not generalize to other abilities, particularly in older adults (von Bastian et al., 2023). Supporting this, a meta-analysis by Melby-Lervåg et al. (2016) indicates that working memory training can produce near transfer effects, but effects on far transfer measures such as intelligence or general executive functioning are minimal to non-existent.

In contrasts to traditional cognitive training programmes that target working memory, attention, or processing speed, SMART (Strengthening Mental Abilities with Relational Training; Cassidy et al., 2011) has emerged as a promising cognitive training. SMART is rooted in Relational Frame Theory (RFT; Hayes et al., 2001) which posits that relational reasoning skills lie at the centre of human cognition. This idea has emerged across many fields in psychology (McLoughlin et al., 2020), including cognitive psychology (e.g., Halford et al., 2010) and behavioural analysis (e.g., Hayes et al., 2001). Within the behavioural tradition, RFT put forward the idea that human cognition centres on one particular type of relational responding that is referred to as arbitrarily applicable relational responding or AARR. It can be defined as the ability to derive relationships between stimuli in an abstract manner where the relation is not based on physical features but on contextual cues that signal how the stimuli are related (Hayes & Hayes, 1992). Humans possess a more expansive repertoire of derived relational responding than non-human animals, which contributes to the highly generative nature of language and cognition in humans. For example, if a person is taught that “A is bigger than B” and “B is bigger than C”, they can derive the untrained relation that “A is bigger than C”.

By training relational responding skills, SMART aims to improve general cognitive abilities (Cassidy et al., 2011). The training programme entails different training phases with increasing difficulty. These phases involve trials with relational premises using arbitrary, nonsense stimuli (e.g., 'CUB is more than KES; KES is more than RUL') followed by a yes/no question about the relation (e.g., 'Is CUB more than RUL?'). While this example demonstrates a quantity relation (more/less), earlier SMART studies also trained same/opposite relations (Cassidy et al., 2011; Colbert et al., 2018; McLoughlin et al., 2021, 2022). In our studies, we extended training beyond this limited set of relations by also including relational frames such as difference (same/different), temporal (before/after), and containment (contains/is within). SMART training is quite intensive in that it involves multiple training sessions per week (typically three 30 minute sessions) over a period of multiple weeks (typically 8 weeks).

Previous studies on SMART suggest that it may offer more than the task-specific improvements often observed in traditional cognitive trainings. Unlike these traditional programmes which struggle to achieve far transfer effects (Gobet & Sala, 2023; von Bastian et al., 2023), SMART has been associated with improvements in IQ scores in children, adolescents and adults (Cassidy et al., 2016; Colbert et al., 2018; Thirus et al., 2016). This has also been confirmed in a recent meta-analysis, indicating a moderate effect size of SMART on nonverbal IQ (May et al., 2022). More recently, active-controlled studies by McLoughlin and colleagues have also reported improvements in standardized measures of nonverbal intelligence following SMART training in primary and secondary school children (McLoughlin et al., 2021, 2022). These studies showed far transfer effects despite variability in training progression, providing further support to the potential generalizability of relational training effects.

It has also been suggested that the programme should be tested in other populations which may benefit from it. One group which could especially benefit from relational training, is older adults, as argued by Kelly (2020). She states that cognitive training in older adults should target the core executive functions which support everyday life activities such as planning, flexible thinking and problem solving. Since arbitrarily applicable relational responding is considered analogous to these executive functions, training relational responding may offer a promising method to enhance cognitive flexibility and daily functioning in older adults (Kelly, 2020).

However, it should also be noted that studies on SMART are not without their limitations. Past studies on SMART have often faced challenges like small sample sizes, limited or absent control conditions, and a lack of preregistered methods and analyses (May et al., 2022). These limitations are not unique to SMART, but are prevailing challenges in the cognitive training literature, contributing to unreliable results on the effectiveness of cognitive training programmes. In fact, these limitations have often led to cognitive trainings appearing promising at first, only for the training effects to fail to replicated in later, more rigorous studies (Simons et al., 2016; von Bastian et al., 2023).

Indeed, Gobet and Sala (2023) argue that the overall effect of far transfer in the cognitive training literature is essentially zero, underscoring an unrealistic optimism in the field and the tendency of cognitive training researchers to ignore results of meta-analyses. Consistent with this, a second-order meta-analysis by Sala et al. (2019) showed that when factors like placebo effects and publication bias were controlled for, there was practically no evidence of far transfer, suggesting that gains from cognitive training do not generalize widely. Both studies further emphasize the importance of rigorous, transparent research practices, including preregistration, careful reporting, and replication, to better understand the true effects of cognitive training (Gobet & Sala, 2023; Sala et al., 2019).

The aim of the current study is to assess whether SMART can be effectively applied in older adults, and whether it can improve or stabilize cognitive complaints due to normal cognitive aging, while also overcoming the methodological issues present in previous literature and training a broader range of relational frames than typically used in earlier SMART studies. This was realized with a randomized control trial, in which one group of older adults followed 8 weeks of the SMART training, while another group of older adults received 8 weeks of Dual N-Back training as an active control condition. Pre- and post-measures of relational responding, neurological status and non-verbal IQ were compared between the conditions to evaluate the relative effectiveness of the training programmes. As secondary outcomes we also measured executive functioning, depression, and quality of life pre and post training, as Kelly (2020) suggested that relational training could also impact core executive processes that support daily functioning in older adults, which may in turn be associated with mental wellbeing and perceived quality of life (Saraçlı et al., 2015; Sun et al., 2024).

We predicted that the participant group who received SMART training would improve more in relational responding skills than the group which received the control training. We also expected that the SMART group would show a larger improvement in neurological status, non-verbal IQ, executive functioning, and quality of life, and a larger decrease in depressive symptoms compared to the control group.

Method

Transparency and Openness

Materials, including data and code used for analysis, have been made available on the Open Science Framework (OSF):

https://osf.io/sdxv5/overview?view_only=49940b32ba4643c6a29ee40f0706e5cf. We provide complete information on sample size decisions, exclusions, manipulations, and measures. The data were processed and analysed using R Statistical Software (v4.3.1; R Core Team, 2021). Originally, we had planned to preregister this study. However, due to an administrative error, the preregistration was never formally registered on the Open Science Framework. Therefore, the study does not meet the formal criteria to be considered preregistered. Nonetheless, the original study design, hypotheses, power estimations, and confirmatory analyses are publicly available on GitHub and timestamped via version control since 25/01/2023 (https://github.com/JamieCummins/subjective-cognitive-complaints/tree/09408caaa6e1db4dd713c76fdea9f8cf2d8ac270/preregistration). This document reflects our intended analyses and hypotheses prior to the completion of data collection. We have uploaded the same materials to OSF for transparency and archival purposes (https://osf.io/sdxv5/overview?view_only=49940b32ba4643c6a29ee40f0706e5cf). Any deviations from these plans are clearly described in the method section. This study was approved by the ethical committee of the Faculty of Psychology and Educational Sciences with reference number 2020/47.

Participants

A total of 173 older adults participated in the study, of which 105 (22 m, 83 f, ages 62-93, M_{age} = 70.26) completed the entire procedure and were included in the analyses. Despite intensive recruitment, we did not reach the planned sample size of 230 participants. However, we also revised our analysis because the originally planned analysis did not account for baseline scores. A post-hoc sensitivity analysis using the pwr package (Champely, 2020) indicated that with our final sample size and the revised analysis plan, we had 80% power to detect a small-to-moderate effect size of $\eta^2_p = 0.07$. This is similar to the initial preregistered goal of 80% power to detect a medium interaction effect of $d = .40$. The large difference in sample size to obtain this power can be explained by the lack of an interaction effect in the revised analysis, which was present in the original plan.

Participants were recruited through Facebook advertisement and through dispersing of flyers in local libraries and health centres. Exclusion criteria were handled with a pre-screening self-report questionnaire assessing age, the subjective cognitive complaints, neurological or psychiatric diagnoses and history with brain injury. Participants with a clinical diagnosis of dementia and/or disorders which may affect ability to complete the training (e.g., attentional dysfunction), or who were using medication which can influence cognition (either inhibitory or facilitatory) were excluded from the study. Participants who experienced a mild traumatic brain injury (TBI) in their younger years were included provided those participants did not experience any cognitive impairment resulting from that mild TBI and made a full recovery (e.g., a minor concussion as a teenager). Participants were randomly assigned to either the SMART group or the Dual N-Back (control) group. Demographics per group are summarized in Table 1. Participant flow through the study is illustrated in the CONSORT diagram (see Figure 1).

Table 1

Demographic Information of Participants Included in Analyses per Group

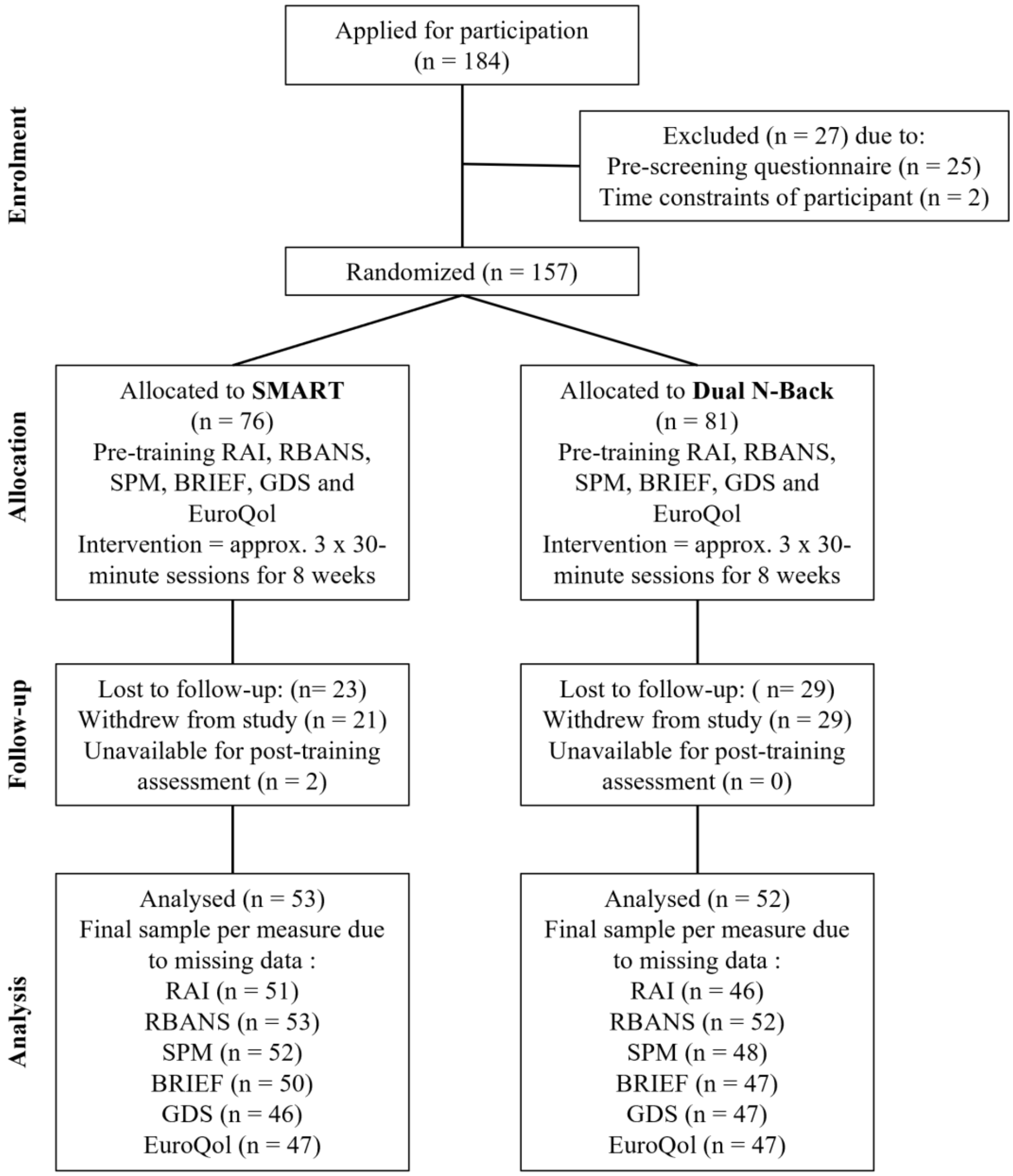
Sample Characteristics	SMART	N-Back
n	53	52
Mean Age (sd)	70.57 (6.07)	69.94 (4.39)
Sex	39 = f 14 = m	44 = f 8 = m

Note. n = number of participants, sd = standard deviation, f = female, m = male.

Alt text: The SD of age in the SMART group was 6.07, while the N-Back group had an SD of 4.39.

Figure 1

CONSORT Diagram of Study Design



Note. n = number or participants, SMART = Strengthening Mental Abilities with Relational Training, RAI = Relational Abilities Index, RBANS = Repeated Battery for the Assessment of Neurological Status, SPM = Standard Progressive Matrices, BRIEF = Behaviour Rating Inventory of Executive Function, GDS = Geriatric Depression Scale.

Alt text: Consort diagram showing participant flow throughout the study including enrolment, participant allocation, and drop-out rates during follow up and analysis.

Materials

Training

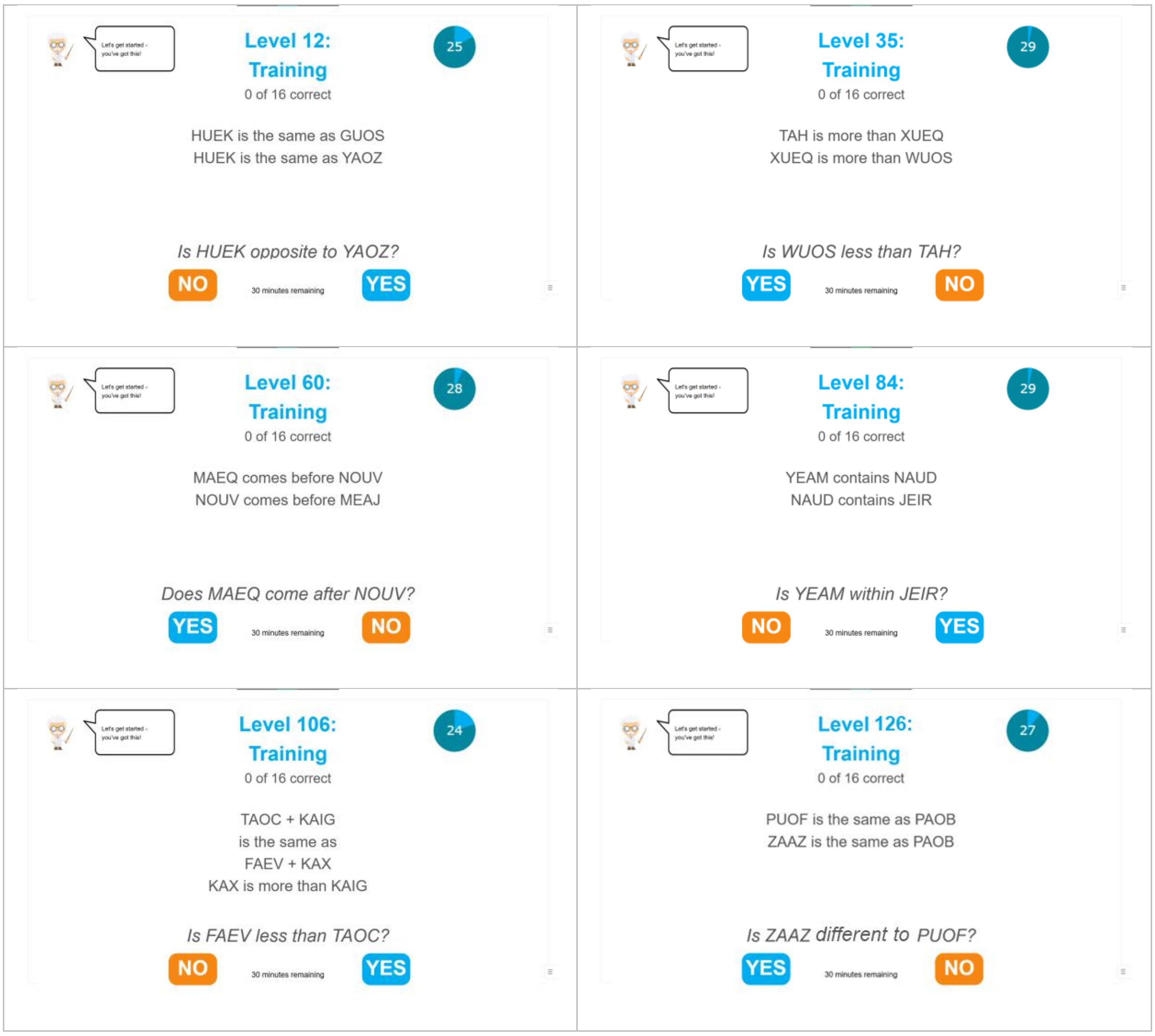
SMART. The SMART training followed the protocol of Cassidy et al. (2011) and was programmed in lab.js. The training consisted of 143 levels. Each level consists of a series of trials with between one to four premises regarding the relation between nonsense stimuli (e.g., CUG is the same as VEK; VEK is opposite to JOM*)* and a yes/no question regarding the relation between those stimuli (e.g., is CUG the same as JOM?) which must be answered within 30 seconds. The nonsense stimuli were randomly generated three or four syllable pseudowords, and were never repeated within the same level.

Difficulty of the levels was varied along multiple dimensions. Firstly, different relations are used across the levels: in levels 1-29, opposition relations (same/opposite) were used; in levels 30-55, quantity relations (more than/less than) were used; in levels 56-81 temporal relations (before/after) were used; in levels 82-103 containment relations (contains/is in) were used; and in levels 104-122 mathematical relations were used. Levels 123-143 were created for this study, and used difference relations (same/different). See Figure 2 for example trials per subscale. Secondly, the number of premises and the order in which they were presented (linear or randomly) varied within these different relations. Also the variation in the relations presented (e.g., all same or opposite relations vs. a mix of same and opposite relations), the presence of a relation in both the sentences and question, and the direction of the relations in the question (e.g., CUG is the same as JOM, is CUG the same as JOM? vs. CUG is the same as JOM, is JOM the same as CUG?) varied between levels.

Each level of SMART is divided in a training phase and a testing phase. In the training phase, participants receive feedback after each trial. They continue completing trials in this level until they achieve a criterion of 16 consecutive correct responses. Once participants have achieved this criterion, they continue to a testing phase, which consists of 16 trials wherein no feedback is presented to participants. If participants answer all of the 16 testing phase trials correctly, they progress to the next level of training. Otherwise, they return to the training phase of the current level.

Figure 2

Example Trials of each of the Subscales of SMART



Note. SMART = Strengthening Mental Abilities with Relational Training. Reading from left to right and from top to bottom: opposition, quantity, temporal, containment, mathematical, and difference relations.

Alt text: Screenshots of SMART trials. Every screenshot depicts a trial's premises in the middle of the screen (e.g., 'HUEK is the same as GUOS, HUEK is the same as YAOZ"), a question at the bottom of the screen (e.g., "Is HUEK opposite to YAOZ?") with a yes and a no button underneath. At the top of every screenshot there is a professor in the left corner with a speech bubble reading "Let's get started- you've got this!", text stating the current level and the number of correct answers given in the middle, and a timer circle counting down from 30 in the right corner.

Dual N-Back. The control group received the Dual N-Back training (Jaeggi et al., 2008), which is a training designed to improve working memory. Participants were instructed to follow a visual and an auditory stream of information simultaneously. The visual stream consisted of a grid of 3 x 3 cells, in which a square would appear in a seemingly random place in the grid. The auditory stream were voice recordings of letters. The goal is to track both streams and to indicate whether the square's position and the letter of the current trial, match those of n-steps back (for example those of 2 trials ago for a 2-back level). Participants had to respond by pressing the 's' button on their keyboard if the square was in the same place, and the 'l' button if the letter matched.

Each level included 20 target trials (i.e., trials where a response was required, such as 22 total trials for a 2-back level). They would start each training session with a 1-back level, and would advance to the next level (e.g., from a 1-back to a 2-back level) after making less than 3 mistakes in both the visual and the auditory task. When they made more than 5 errors in either task they would go down a level. If they made between 3 and 5 errors, they would stick to the current N-Back level.

Outcome Variables

Relational Abilities Index (RAI). The RAI (Colbert et al., 2017; Cummins, 2023; Cummins et al., 2023) was employed before and after the training to measure participants' relational responding skills, which the SMART programme specifically targets. It contains 128 trials, which follow the same structure as the SMART trials, with yes/no questions which must be answered based on 2 to 4 given premises about nonsense stimuli. Eight subscales were included in the RAI, differentiated by the relational frames in the trials: opposition, difference, quantity, temporal, containment, analogy (relations between relations; e.g., "SEY is more than LUD, VED is less than XUF. Is SUY to LUD the same as VED to XUF?") , deictic (perspective-taking relations involving place and time, and reversals; e.g., "GAJ is here now, WOM was there than. If here was there and there was here, and if GAJ is WOM and WOM is GAJ, is GAY there then?") and mathematical relations. These correspond to the subscales trained in the SMART training. Each subscale consisted of 16 trials, with increasing difficulty. Participants had 30 seconds per trial to answer. The RAI has been shown to correlate strongly with several subscales of the Wechsler Adult Intelligence Scale (WAIS), and has been suggested as a potential proxy for Full Scale IQ (Colbert et al., 2017).

Repeated Battery for the Assessment of Neurological Status. The RBANS is a multiple form test instrument used to repeatably measure cognitive decline (Randolph et al., 1998). The test takes approximately 30 minutes to administer, and measures performance on five subscales: immediate memory, delayed memory, attention, language, and visuospatial/constructional skills. For the primary analyses, we used the total test score. We used the A-form as pre-training and the B-form as post-training assessments. The RBANS total score has demonstrated good diagnostic accuracy for detecting cognitive decline in older adults (Duff et al., 2008) and its subscale scores have been shown to differentiate between different types of cognitive impairment, such as Alzheimer's and Huntington's disease

(Randolph et al., 1998). If the assessment took place online through videocall, participants held their response sheets up to the camera, allowing the researcher to capture screenshots of drawings and written responses, ensuring that responses could not be altered afterward.

Standard Progressive Matrices. Non-verbal IQ was measured using the SPM. The SPM is a non-verbal test which is widely used to assess general IQ (Raven, 1936). The test consists of 60 trials, each involving a matrix of visual stimuli with one element missing. The task is to select the correct missing element from 6 options. The trials were untimed, but the participants had 45 minutes to complete the whole task. The participants were allowed to skip trials, but were not able return to previous or skipped trials. In terms of validity, the SPM has been shown to correlate strongly with other measures of intelligence in older adults, including the Wechsler Adult Intelligence Scale (WAIS; Deary et al., 2004).

Behaviour Rating Inventory of Executive Function – Adult Version. The BRIEF-A self-report questionnaire (Gioia et al., 2000) was administered to assess executive functioning in everyday life. It exists of 75 items focussing on functions like inhibition, emotion regulation, planning and organizing etc. Participants are asked to indicate how frequently certain events (e.g., 'I say things without thinking') occur by selecting 'never,' 'sometimes,' or 'often'. The scores get summed up and transformed into T-scores. Lower scores indicate better executive functioning.

Geriatric Depression Scale - 15. Depression was measured with the GDS-15 (Yesavage & Sheikh, 1986), a self-report screening questionnaire for depression in older adults. The questionnaire entails 15 yes/no questions about how the older adults have felt the past 7 days, today included, at the time of the assessment. Higher scores indicate more severe depressive symptoms.

EuroQol – 5D – 5L. The EuroQol – 5D – 5L (Herdman et al., 2011) was used to measure quality of life of the older adults. It is a self-reported questionnaire in which 5 dimensions of health are considered: mobility, selfcare, usual activities, pain/discomfort, and anxiety/depression. Each dimension is rated on a 5-point scale (no problem to extreme problems). The scores are converted in a single index score per participant based on country specific value sets (in this case the Belgian value set). The scores generally range from -0.594 to 1 (where 1 = full health, 0 = death, and negative values indicate states worse than death).

Procedure

After signing the informed consent form, participants were asked to do a first test session. During this session, the RAI, RBANS, SPM, BRIEF, GDS-15, and EuroQol were administered. This session either took place at the university, or online through videocall. The RBANS was administered using a pen-and-paper format, the other measures were programmed as in-browser assessments using Lab.js (Henninger et al., 2022). After the first test session, participants would start a training period of 8 weeks with either SMART or the Dual N-Back training. The training was assigned randomly. During the 8 week period, the participants were asked to do 30 minute online training sessions, for 3 times a week. The participants were free to choose during which days they completed the training but were advised to distribute the sessions across three separate days. However, if necessary, they were permitted to complete multiple sessions on a single day to meet the required quota. Participants were also allowed to complete more than three sessions a week, either by choice or to compensate for missed sessions. However, they were required to engage in training across the full 8-week period and were not allowed to complete all sessions in a shorter time frame. The frequency of these sessions was monitored by the researchers, and followed up by weekly emails to the participants. Following the training period, a second test session would occur in which the same tests were administered as in the first test session. After complete participation, participants were given full access to both trainings.

Statistical Analyses

The data were processed using R Statistical Software (v4.3.1; R Core Team, 2021) and the *Tidyverse* package (Wickham et al., 2019). The confirmatory hypotheses were tested using an ANCOVA. The model predicted the post-training scores (T2) by group, controlling for the pre-training scores (T1). For this, the *car* package (Fox & Weisberg, 2019) was used. This method differs from the initial analysis plan available on GitHub (https://github.com/JamieCummins/subjective-cognitive-complaints/tree/09408caaa6e1db4dd713c76fdea9f8cf2d8ac270/preregistration), because the pre-data collection analysis plan did not account for baseline scores (see Roche et al., 2023 for the importance of controlling for baseline scores in RCT studies).

To explore the relationship between the dependent variables, Pearson correlations were computed between the baseline full-scale scores of all outcome measures. P-values were Holm-Bonferroni corrected to account for multiple testing. In addition, to explore how the different relational frames were associated with the other outcomes, Pearson correlations were computed between the RAI subscale baseline scores and the pre-training scores of the other outcome variables. P-values were Holm-Bonferroni corrected per RAI subscale.

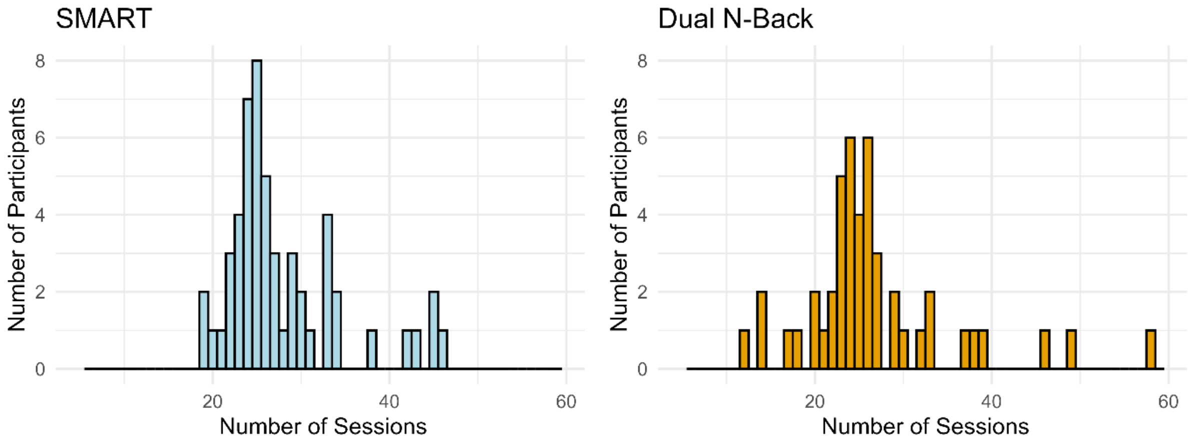
Results

Training Performance

Participants of the SMART group completed between 19 and 46 training sessions (M = 27.94, SD = 6.66). They progressed to levels ranging from 15 to 143 (M = 83.53, SD = 35.47). In the Dual N-Back group, participants completed between 5 and 58 training sessions (M = 26.26, SD = 9.12). They reached levels ranging from 2-back to 5-back (M = 3.66, SD = 0.60). A Welch two-sample t-test indicated that the number of completed training sessions did not differ significantly between groups, $t(83.3) = -1.05$, $p = .30$, 95% CI [-4.90, 1.52]. The distributions of number of completed training sessions are visualized in Figure 3 per training condition.

Figure 3

Distribution of Completed Training Sessions per Condition



Note. SMART = Strengthening Mental Abilities with Relational Training.

Alt text: bar graphs showing a similar distribution pattern between conditions.

Confirmatory Analyses

Primary Analyses

Mean scores of the primary outcome measure are summarized by group in Table 2 and Figure 4. For the RAI, post-training scores were significantly predicted by group, controlling for pre-training scores, $F(1, 94) = 9.48$, $p < .005$, $\eta^2_p = 0.09$, 95% CI [0.01, 0.22]. Post-hoc one-sided paired-samples t-tests revealed that post-training scores were higher than pre-training scores in both the SMART group, $t(50) = 7.39$, $p < .001$, *Cohen's d* = 0.74, 95% CI [0.52, 0.97], and the Dual N-back group, $t(45) = 4.40$, $p < .001$, *Cohen's d* = 0.42, 95% CI [0.22, 0.62].

For the other two primary outcomes was there no significant main effect of group on the post-training scores, controlling for pre-training scores: RBANS, $F(1, 85) = 0.83$, $p = .36$, $\eta^2_p = 0.01$, 95% CI [0, 0.07]; SPM, $F(1, 97) = 3.01$, $p = .09$, $\eta^2_p = 0.03$, 95% CI [0, 0.12].

Post-hoc one-sided paired-samples t-tests revealed that post-training scores were not significantly higher than pre-training scores over both groups for the RBANS, $t(104) = -0.07$, $p = .47$, *Cohen's d* = -0.01, 95% CI [-0.15, 0.14], and the SPM, $t(99) = 0.81$, $p = .21$, *Cohen's d* = 0.05, 95% CI [-0.07, 0.18].

Secondary Analyses

Secondary Outcome Measures. Mean scores of the secondary outcome measure are summarized by group in Table 2 and Figure 4. None of the secondary outcomes were significantly predicted by group, controlling for pre-training scores: BRIEF, $F(1, 94) = 0.08$, $p = .69$, $\eta^2_p = 0$, 95% CI [0, 0.04]; GDS, $F(1, 90) = 0.16$, $p = .78$, $\eta^2_p = 0.002$, 95% CI [0, 0.06]; EuroQol, $F(1, 91) = 0.19$, $p = .66$, $\eta^2_p = 0.002$, 95% CI [0, 0.06].

Post hoc one-sided paired-samples t-tests revealed that post-training scores were significantly lower than pre-training scores for the BRIEF overall, $t(96) = -3.08$, $p < .01$, *Cohen's d* = -0.17, 95% CI [-0.28, -0.06]. Post-training scores were not significantly lower over the groups for the GDS, $t(92) = -1.58$, $p = .06$, *Cohen's d* = -0.12, 95% CI [-0.28, 0.03], nor higher for the EuroQol, $t(93) = 0.72$, $p = .76$, *Cohen's d* = 0.05, 95% CI [-0.09, 0.20].

Table 2

Pre- and Post-Training Scores per Group

Measure	Mean (SD)			
	SMART		DUAL N-BACK	
	Pre-training	Post-training	Pre-training	Post-training
Primary outcomes				
RAI	79.80 (13.50)	91.12 (16.21)	78.07 (12.68)	83.57 (13.13)
RBANS	108.70 (12.46)	107.23 (11.89)	104.46 (11.25)	105.83 (11.25)
SPM	43.56 (8.37)	44.71 (8.70)	42.88 (7.13)	42.48 (7.07)
Secondary outcomes				

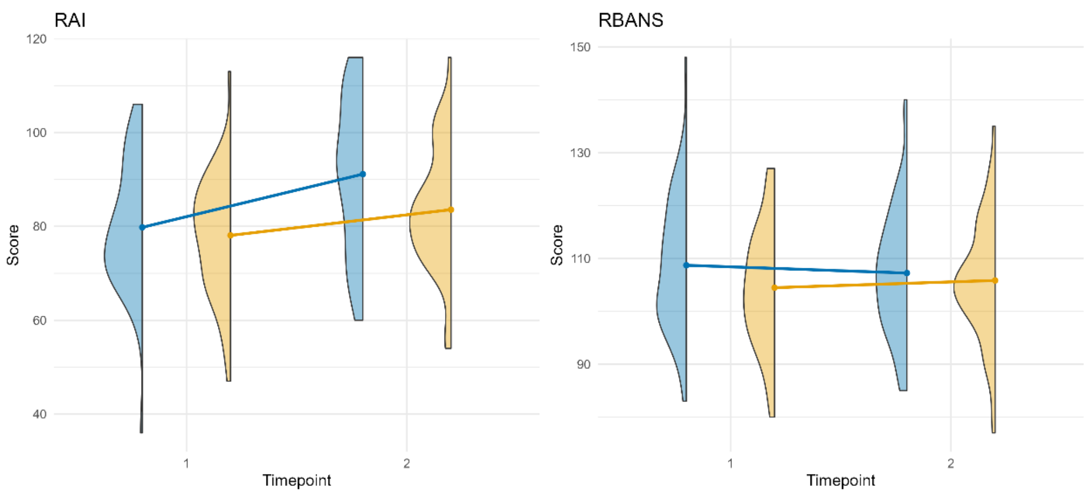
Measure	Mean (SD)			
	SMART		DUAL N-BACK	
	Pre-training	Post-training	Pre-training	Post-training
BRIEF - A	57.02 (7.88)	55.60 (7.94)	57.06 (8.12)	55.79 (7.32)
GDS - 15	3.04 (2.67)	2.61 (2.84)	2.60 (1.93)	2.40 (2.40)
EuroQol – 5D	0.86 (0.14)	0.86 (0.16)	0.86 (0.14)	0.87 (0.14)

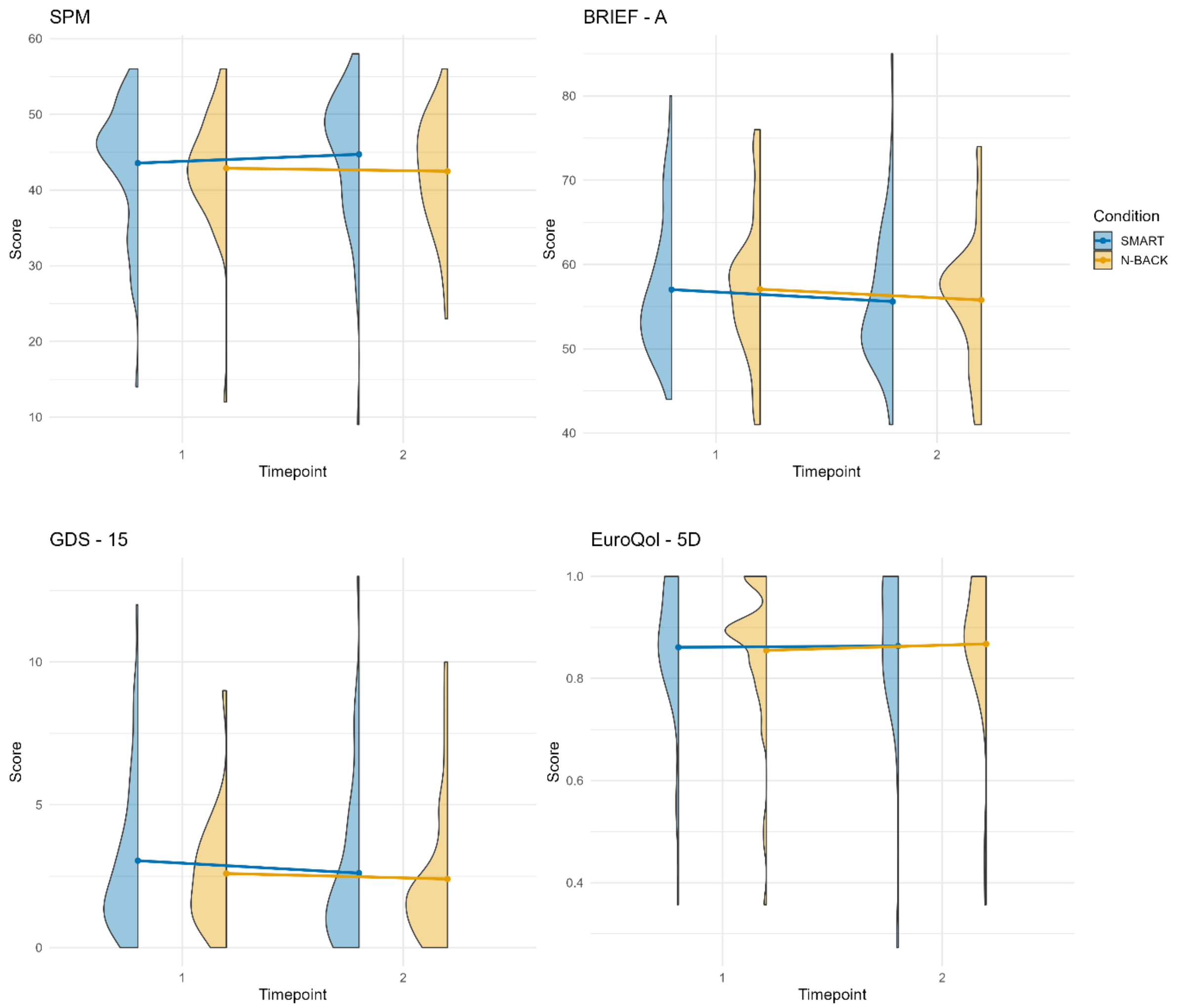
Note. SD = standard deviation, SMART = Strengthening Mental Abilities with Relational Training, RAI = Relational Abilities Index, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status, SPM = Standard Progressive Matrices, BRIEF-A = Behavior Rating Inventory of Executive Function–Adult Version, GDS-15 = 15-item Geriatric Depression Scale, EuroQol–5D = 5-Dimension EuroQol quality-of-life measure.

Alt text: Mean scores and SD's per primary and secondary measure per group pre- and post training. Participants in both groups showed improvements on the RAI, with larger gains in the SMART group. RBANS scores remained stable across groups, while SPM scores slightly increased in SMART but not in Dual N-Back. Both groups showed small reductions in BRIEF-A and GDS-15 scores, and EuroQol-5D scores remained stable.

Figure 4

Score Distributions and Mean Performance per Condition Across Timepoints





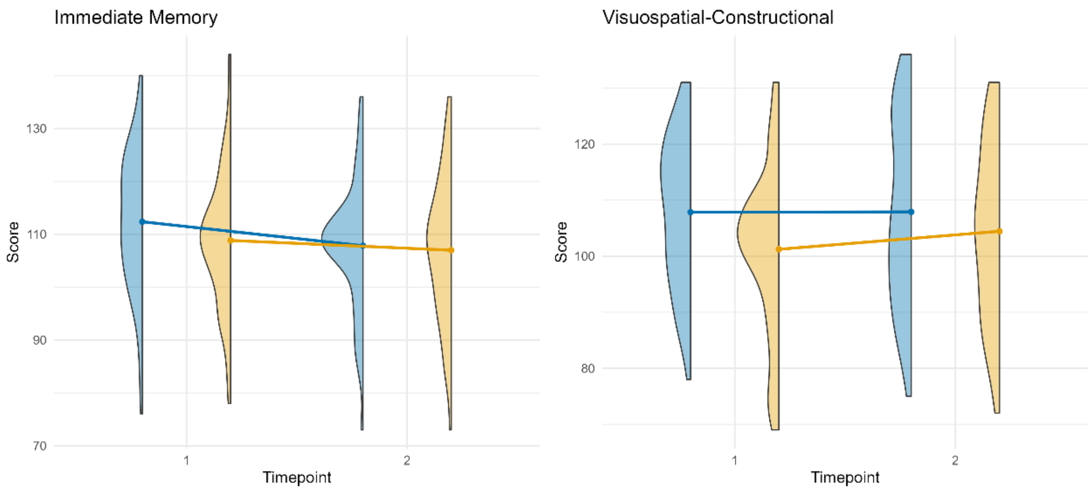
Note. SMART = Strengthening Mental Abilities with Relational Training, RAI = Relational Abilities Index, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status, SPM = Standard Progressive Matrices, BRIEF-A = Behavior Rating Inventory of Executive Function–Adult Version, GDS-15 = 15-item Geriatric Depression Scale, EuroQol–5D = 5-Dimension EuroQol quality-of-life measure. Violin plots represent the distribution of scores per condition at each timepoint. Points and lines indicate (the difference in) group means over time.

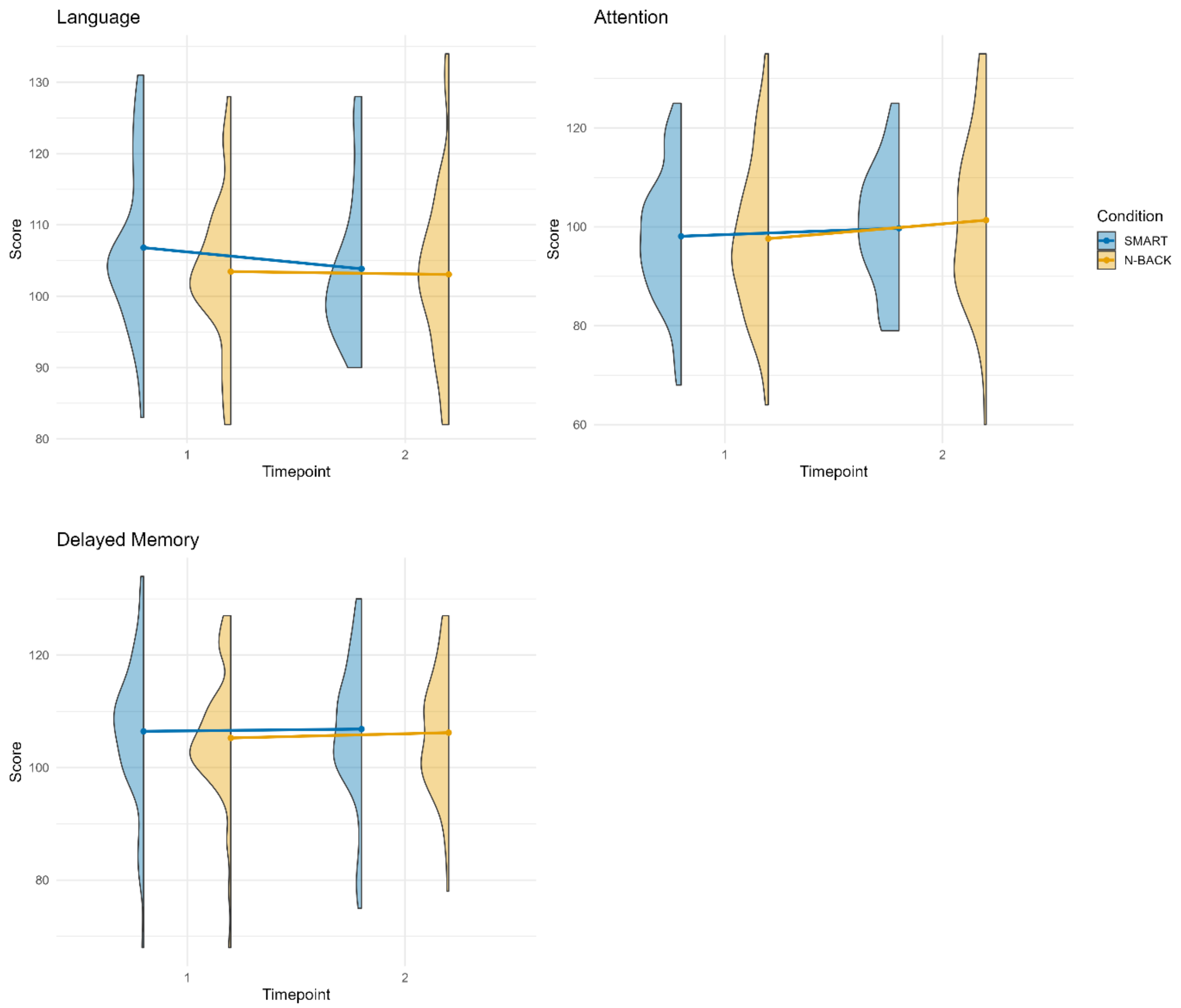
Alt text: half violin plots illustrating pre- to post-training changes in outcome measures, showing an increase in RAI scores in both groups, with a larger increase in the SMART group. RBANS scores remained stable across groups. SPM scores showed a slight increase in the SMART group but not in Dual N-Back. Both groups exhibited small reductions in BRIEF-A and GDS-15 scores, while EuroQol-5D scores remained unchanged.

Subscales RBANS. To test whether there were differential training effects on any of the RBANS subscales, we ran ANCOVAs for each subscale. There was no significant main effect of group on the post-training scores, controlling for pre-training scores for any of the subscales: immediate memory, $F(1, 102) = 0.33$, $p = .57$, $\eta^2_p = 0.003$, 95% CI [0, 0.06]; visuospatial-constructional, $F(1, 102) = 0.01$, $p = .94$, $\eta^2_p = 0$, 95% CI [0, 0.02]; language, $F(1, 102) = 0.25$, $p = .62$, $\eta^2_p = 0.002$, 95% CI [0, 0.05]; attention, $F(1, 102) = 0.78$, $p = .38$, $\eta^2_p = 0.008$, 95% CI [0, 0.07]; delayed memory, $F(1, 102) = 0.01$, $p = .94$, $\eta^2_p = 0$, 95% CI [0, 0]. Pre- and post-training scores per group are visualized in Figure 5.

Figure 5

Score Distributions and Mean Performance on RBANS Subscales Across Timepoints





Note. SMART = Strengthening Mental Abilities with Relational Training, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status. Violin plots represent the distribution of scores per condition at each timepoint. Points and lines indicate (the difference in) group means over time.

Alt text: half violin plots illustrating pre- to post-training changes in RBANS subscales, showing a small reduction in immediate memory in both groups, a slight increase in the visuospatial-constructional subscale in the N-Back group but not the SMART group, a small reduction in language in the SMART group but not the N-Back group, a small increase in attention for both groups, and no noticeable difference in delayed memory for both groups.

Correlations at Baseline. To explore the relationship between the dependent variables at baseline, we computed Pearson correlations between the pre-training full scale scores. Results can be found in Table 3.

Table 3

Correlation Matrix between Dependent Variables at Baseline

Measure	1	2	3	4	5
1. RAI	-				
2. RBANS	0.39 [0.21, 0.55] **	-			
3. SPM	0.51 [0.35, 0.64] ***	0.28 [0.08, 0.45]	-		
4. BRIEF - A	0.16 [-0.04, 0.35]	-0.01 [-0.21, 0.2]	0.08 [-0.12, 0.28]	-	
5. GDS-15	-0.11 [-0.31, 0.1]	-0.12 [-0.32, 0.09]	-0.06 [-0.26, 0.15]	0.58 [0.42, 0.7] ***	-
6. EuroQol-5D	0.16 [-0.04, 0.36]	0.04 [-0.17, 0.24]	0.17 [-0.03, 0.37]	-0.28 [-0.46, -0.08]	-0.38 [-0.54, -0.19] **

Note. RAI = Relational Abilities Index, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status, SPM = Standard Progressive Matrices, BRIEF-A = Behavior Rating Inventory of Executive Function–Adult Version, GDS-15 = 15-item Geriatric Depression Scale, EuroQol–5D = 5-Dimension EuroQol quality-of-life measure. Values between square brackets indicate 95% CI. * indicates that $p < .05$, ** indicates $p < .01$, and *** indicates that $p <.001$. P-values were Holm-Bonferroni corrected.

Alt text: RAI correlated moderately with RBANS (r = .39) and SPM (r = .51). RBANS was also positively correlated to SPM (r = .28). BRIEF showed no significant correlations with cognitive measures. GDS-15 correlated strongly with BRIEF (r = .58), and negatively with EuroQol-5D (r = –.38). EuroQol-5D was otherwise weakly related to other outcomes.

Correlations with RAI Subscales. To explore the relationship between the RAI subscales and the other dependent variables, we computed Pearson correlations between the pre-training scores. Results can be found in Table 4.

Table 4

Correlation Matrix between RAI Subscales and Other Outcomes at Baseline

Measure	opposition	difference	quantity	temporal	containment	analogy	deictic	mathematical
RBANS	0.25 [0.05, 0.43]	0.29 [0.1, 0.46] *	0.28 [0.08, 0.45]	0.32 [0.13, 0.49] *	0.2 [0, 0.39]	0.05 [-0.15, 0.24]	0.35 [0.16, 0.51] **	0.12 [-0.08, 0.31]
SPM	0.29 [0.1, 0.46] *	0.38 [0.19, 0.54] **	0.44 [0.26, 0.59] ***	0.35 [0.17, 0.52] **	0.24 [0.04, 0.42]	0.16 [-0.04, 0.35]	0.3 [0.11, 0.47] *	0.34 [0.14, 0.5] **
BRIEF - A	-0.04 [-0.24, 0.16]	0.15 [-0.05, 0.35]	0.15 [-0.05, 0.34]	0.14 [-0.07, 0.33]	0.18 [-0.02, 0.37]	0.15 [-0.05, 0.34]	0.05 [-0.16, 0.25]	0.01 [-0.2, 0.21]
GDS-15	-0.15 [-0.34, 0.06]	-0.02 [-0.22, 0.19]	0 [-0.21, 0.21]	-0.08 [-0.28, 0.13]	-0.04 [-0.24, 0.17]	0.06 [-0.15, 0.27]	-0.18 [-0.38, 0.02]	-0.13 [-0.33, 0.08]
EuroQol-5D	0.05 [-0.16, 0.25]	0.11 [-0.1, 0.31]	0.15 [-0.06, 0.34]	0.21 [0, 0.4]	0.07 [-0.14, 0.27]	-0.17 [-0.36, 0.04]	0.06 [-0.14, 0.27]	0.25 [0.05, 0.44]

Note. RAI = Relational Abilities Index, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status, SPM = Standard Progressive Matrices, BRIEF-A = Behavior Rating Inventory of Executive Function–Adult Version, GDS-15 = 15-item Geriatric Depression Scale, EuroQol–5D = 5-Dimension EuroQol quality-of-life measure. Values between square brackets indicate 95% CI. * indicates that $p < .05$, ** indicates $p < .01$, and *** indicates that $p < .001$. P-values were Holm-Bonferroni corrected per RAI subscale.

Alt text: RBANS showed small-to-moderate positive associations with most relational responding measures, particularly temporal (r = .32) and deictic (r = .35). SPM displayed consistent moderate correlations across domains, strongest with quantity (r = .44) and difference (r = .38). BRIEF and GDS-15 showed no significant correlations. EuroQol-5D was weakly related to temporal (r = .21) and mathematical (r = .25).

Exploratory Analyses

Two potential exploratory analyses were outlined in the publicly available document on GitHub (https://github.com/JamieCummins/subjective-cognitive-complaints/tree/09408caaa6e1db4dd713c76fdea9f8cf2d8ac270/preregistration). The first one was to explore the impact of whether assessments were completed in person or in an online setting would impact the outcomes. However, given the null results of this study, and because we had no a priori hypothesis regarding format of administration, we decided not to pursue this exploratory analysis further in this paper. This was decided post hoc and is reported here for transparency. The second planned exploratory analysis was to examine the rates of drop out between the training conditions. To test this, we conducted a Pearson's Chi-squared test with Yates' continuity correction. The test was not significant, $\chi^2(1) = 0.59$, $p = .44$, *Cramer's V* = 0.05, 95% CI [0, 0.19]. Drop out rates per condition can be found in Figure 1.

An additional exploratory analysis was carried out to examine whether a dosage effect was present for the primary outcome measures, given the variety in number of training sessions completed by the participants. This was tested by adding number of training sessions as an independent variable to the confirmatory ANCOVA model. Number of training sessions did not significantly predict training outcomes for the RAI, $F(1, 90) = 1.11$, $p = .30$, $\eta^2_p = 0.01$, 95% CI [0, 0.09]; for the RBANS, $F(1, 96) = 3.48$, $p = .07$, $\eta^2_p = 0.03$, 95% CI [0, 0.13]; nor for the SPM, $F(1, 93) = 1.49$, $p = .23$, $\eta^2_p = 0.02$, 95% CI [0, 0.10].

Discussion

Within the literature on cognitive training programmes, SMART emerged as a promising alternative (Cassidy et al., 2011; Kelly, 2020). Although some studies have reported positive findings, the programme's effectiveness has yet to be validated through methodological rigorous research (May et al., 2022). Given that cognitive decline in older adults is a pressing societal challenge (Clay et al., 2005; Tucker-Drob, 2011), our study aimed to evaluate the possible training effects of SMART in older adults by comparing it to Dual N-Back training. The SMART training was also extended to include not only opposition and quantity relations, but also temporal, containment, mathematical, and difference relations. A range of outcome measures were taken into account to assess relational reasoning, neurological status, non-verbal IQ, executive functioning, depression, and quality of life.

Our results indicated that participants who received SMART training demonstrated a greater increase in relational reasoning abilities compared to the participants who completed the Dual N-Back training. This implies that SMART had a specific training effect on relational reasoning, as was expected. However, the data showed no significant group effects on the other cognitive and behavioural measures, and neither of the groups significantly improved after training on neurological status, non-verbal IQ, depression, or quality of life, indicating a lack of transfer to these untrained skills. Also when looking at the different facets of neurological status (immediate and delayed memory, attention, language, and visuospatial-constructional skills), no apparent training effects were observed. Participants did show a significant improvement in executive functioning, regardless of the training condition. Considering that both SMART and Dual N-Back training elicited an improvement, this could be interpreted as a general benefit of engaging in cognitive training. However, taking into account that we used a self-report questionnaire, it remains unclear whether executive functioning itself has improved. It is equally possible that only self-perception of executive functioning improved rather than objective cognitive gains.

Furthermore, there were no exceptionally strong correlations between the RAI or the RAI subscales with any of the outcome measures. This suggests limited potential for training effects, as stronger correlations at baseline increase the chance at transfer from one skill to another (Hawes et al., 2022; Melby-Lervåg & Hulme, 2013). While such correlations can increase the possibility of transfer effects, it is important to note that transfer is a complex phenomenon, that also relies on other factors such as training conditions and motivation (Blume et al., 2010).

It is possible that SMART is effective only under certain training conditions which were not met in this study, explaining the weak training effects. One such factor is training dosage (Amd & Roche, 2018). In the current study, the mean reached stage in the SMART condition is equivalent or even higher compared to previous studies (Amd & Roche, 2018; Cassidy et al., 2011; Colbert et al., 2018; Roche et al., 2023). Therefore insufficient training dosage is unlikely to explain the observed weak training effects. This was also tested formally in an exploratory analysis which showed that the number of training sessions completed by the participants did not affect the results. Relatedly, the present study also implemented a more extensive SMART protocol than most earlier work (Cassidy et al., 2011; Colbert et al., 2018; McLoughlin et al., 2021, 2022) by also training temporal, containment, mathematical, and difference relations. This suggests that the observed limited transfer effects are unlikely to be explained by restricted training content.

Another training factor which should be considered is the target population. SMART has mostly been applied in children and not older adults, who may differ in cognitive flexibility and training gains. It has indeed been considered that especially younger kids are more responsive to cognitive training interventions (Birtwistle et al., 2025). In the context of SMART, studies from McLoughlin and colleagues (2021, 2022) in primary and secondary school children have demonstrated far transfer effects on nonverbal reasoning and educational outcomes using comparable active-control designs. These findings suggest that SMART can produce broader cognitive benefits in children. Thus, age-related differences could explain the limited training effects of SMART in the present sample of older adults.

As mentioned, transfer effects can also depend on motivation (Blume et al., 2010). Participants received no reimbursement for participation, suggesting that those who completed the study were driven by intrinsic motivation. As drop out rates were higher in the Dual N-Back condition, this could indicate that this programme was less engaging or motivating for participants. However, the drop out rates were not significantly related to training condition, so this should be interpreted with caution. Remarkably, some participants in the Dual N-Back condition reached levels as high as 5-back. Performance on N-Back tasks typically declines as load increases, and older adults generally show larger decreases in accuracy and slower responses at higher N-Back levels compared with younger adults (Bopp & Verhaeghen, 2020). Similarly, in the SMART condition, some participants completed the full programme, reaching the ultimate stage (i.e., stage 143). In comparison, previous SMART studies in children and adolescents typically report lower maximum stages reached, often due to variations in the number of levels implemented in the programme (Cassidy et al., 2011; Colbert et al., 2018; McLoughlin et al., 2021, 2022). Achieving high levels on the Dual N-Back and completing all 143 SMART stages highlights the considerable cognitive performance, task engagement, and potential motivation of at least a subset of older adults in the present study.

Despite these considerations, it is worth noting that our results are in the same trend as traditional research on cognitive training programmes (see Gobet & Sala, 2023, and von Bastian et al., 2023). Initially, studies yielded impressive positive results, but these were often reported in studies with poor methodological rigor. However, when replicating the effects with more rigorous methods, the training effect often diminishes (Gobet & Sala, 2023; von Bastian et al., 2023). One of the strengths of this study is its rigorous methodology, including an active control condition, random assignment, the use of multiple outcome measures, and a relatively large sample size. By comparing SMART to another established cognitive training programme, we could draw more specific conclusions about the effectiveness of SMART while controlling for spontaneous, non-training related improvement. While we did not reach the planned sample size, we still had a larger sample than most previous research on SMART (May et al., 2022), and a sensitivity analysis indicated that we still had adequate power to observe a small-to-moderate effect size.

Given that the current study used a more rigorous method and yielded null results, this suggest that the effectiveness of SMART might be overestimated in the previous literature (May et al., 2022). Nevertheless, it is important to note that our study targeted a different population, whereas many previous studies focused on children or younger adults. Therefore, no definitive conclusions can yet be drawn about the overall effectiveness of SMART. Further methodological rigorous research is needed to validate the training effects of SMART across different outcome measures and populations.

Despite the methodological strengths of this study, several limitations should be considered. First, by using self-report measures to assess executive functioning, depression and quality of life, we cannot conclude whether the observed (lack of) training effects on these measures are representative of objective changes in cognitive or mental changes as self-reports could reflect self-perception rather than actual performance. Hence, future research should aim to implement objective behavioural measures alongside self-reports. A second limitation to consider is the fact that the study was partially conducted online. Since older adults can be a hard population to reach, we also gave the option to complete assessments online. This lead to less control during the assessments, and also some data being lost in the process. Also, not all older adults have access to a computer or are comfortable using one, which was necessary for participating in the SMART or Dual N-Back training. This limits the representativeness of our sample. Third, the absence of a long-term follow-up means we cannot determine whether any observed effects would be sustained over time.

Taken together, the findings of this study highlight the need for careful consideration when implementing cognitive training programs such as SMART, especially in older populations. Although relational reasoning improved, the lack of transfer to other cognitive abilities suggests that the effectiveness of SMART is limited in this population. Future research could explore whether adaptations in SMART might enhance the programme's effectiveness in this group by facilitating generalizability and cognitive flexibility. For example, a new version of SMART could incorporate contextual cues that signal how stimuli need to be related (e.g., as being similar or opposite). This would allow researchers to assess how flexibly participants can shift relational responses on the basis of contextual cues.

To conclude, this study aimed to evaluate the effects of SMART compared to Dual N-Back training in older adults, but only relational reasoning seemed to benefit from SMART specifically, and no transfer effects were observed in untrained tasks. This suggests that the programme's effectiveness in older adults is limited. These findings follow a similar course as traditional research on cognitive training which shows that initial cognitive training findings often subside under more rigorous research methods (Gobet & Sala, 2023; von Bastian et al., 2023). This poses questions on the application of the programme, especially in older adults. Future research should explore adaptations to SMART that might enhance its effectiveness, while prioritizing methodological rigour. Overall, SMART's potential as a cognitive training programme remains unconfirmed and should be studied further in different populations and with different adaptations.

References

Amd, M., & Roche, B. (2018). Assessing the effects of a relational training intervention on fluid intelligence among a sample of socially disadvantaged children in Bangladesh. *The Psychological Record*, *68*(2), 141–149. https://doi.org/10.1007/s40732-018-0273-4

Bezdicek, O., Ferreira, J., Fellows, R., & Liepelt-Scarfone, I. (2023). Editorial: Activities of daily living and everyday functioning: From normal aging to neurodegenerative diseases. *Frontiers in Aging Neuroscience*, *15*. https://doi.org/10.3389/fnagi.2023.1161736

Birtwistle, E., Chernikova, O., Wünsch, M., & Niklas, F. (2025). Training of executive functions in children: A meta-analysis of cognitive training interventions. *SAGE Open*, *15*(1), 21582440241311060. https://doi.org/10.1177/21582440241311060

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, *36*(4), 1065–1105. https://doi.org/10.1177/0149206309352880

Bopp, K. L., & Verhaeghen, P. (2020). Aging and N-back performance: A meta-analysis. *The Journals of Gerontology: Series B*, *75*(2), 229–240. https://doi.org/10.1093/geronb/gby024

Cassidy, S., Roche, B., Colbert, D., Stewart, I., & Grey, I. M. (2016). A relational frame skills training intervention to increase general intelligence and scholastic aptitude. *Learning and Individual Differences*, *47*, 222–235. https://doi.org/10.1016/j.lindif.2016.03.001

Cassidy, S., Roche, B., & Hayes, S. C. (2011). A relational frame training intervention to raise intelligence quotients: A pilot study. *The Psychological Record*, *61*(2), 173–198. https://doi.org/10.1007/BF03395755

Champely, S. (2020). *pwr: Basic functions for power analysis* (Version 1.3-0) [Computer software]. https://CRAN.R-project.org/package=pwr

Clay, O. J., Wadley, V. G., Edwards, J. D., Roth, D. L., Roenker, D. L., & Ball, K. K. (2005). Cumulative meta-analysis of the relationship between useful field of view and driving performance in older adults: Current and future implications. *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, *82*(8), 724–731. https://doi.org/10.1097/01.opx.0000175009.08626.65

Colbert, D., Dobutowitsch, M., Roche, B., & Brophy, C. (2017). The proxy-measurement of intelligence quotients using a relational skills abilities index. *Learning and Individual Differences*, *57*, 114–122. https://doi.org/10.1016/j.lindif.2017.03.010

Colbert, D., Tyndall, I., Roche, B., & Cassidy, S. (2018). Can SMART training really increase intelligence? A replication study. *Journal of Behavioral Education*, *27*(4), 509–531. https://doi.org/10.1007/s10864-018-9302-2

Cummins, J. (2023). On the measurement of relational responding. *Journal of Contextual Behavioral Science*, *30*, 155–168. https://doi.org/10.1016/j.jcbs.2023.10.003

Cummins, J., Nevejans, M., Colbert, D., & De Houwer, J. (2023). On the structure of relational responding. *Journal of Contextual Behavioral Science*, *27*, 16–25. https://doi.org/10.1016/j.jcbs.2022.11.006

Czaja, S. J., Sharit, J., Lee, C. C., Nair, S. N., Hernández, M. A., Arana, N., & Fu, S. H. (2013). Factors influencing use of an e-health website in a community sample of older adults. *Journal of the American Medical Informatics Association : JAMIA*, *20*(2), 277–284. https://doi.org/10.1136/amiajnl-2012-000876

Deary, I. J., Whalley, L. J., & Crawford, J. R. (2004). An ‘instantaneous’ estimate of a lifetime’s cognitive change. *Intelligence*, *32*(2), 113–119. https://doi.org/10.1016/j.intell.2003.06.001

Derafshi, R., Babulal, G. M., & Bayat, S. (2024). Impact of cognitive impairment on driving behaviour and route choices of older drivers: A real-world driving study. *Scientific Reports*, *14*(1), 14174. https://doi.org/10.1038/s41598-024-63663-y

Duff, K., Clark, H. J. D., O'Bryant, S. E., Mold, J. W., Schiffer, R. B., & Sutker, P. B. (2008). Utility of the RBANS in detecting cognitive impairment associated with Alzheimer's disease: Sensitivity, specificity, and positive and negative predictive powers. *Archives of Clinical Neuropsychology*, *23*(5), 603–612. https://doi.org/10.1016/j.acn.2008.06.004

Fox, J., & Weisberg, S. (2019). *An {R} companion to applied regression*. Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gioia, G., Isquith, P., Guy, S. C., & Kenworthy, L. (2000). Behavior rating inventory of executive function. *Child Neuropsychology*, *6*, 235–238.

Giuli, C., Papa, R., Lattanzio, F., & Postacchini, D. (2016). The effects of cognitive training for elderly: Results from my mind project. *Rejuvenation Research*, *19*(6), 485–494. https://doi.org/10.1089/rej.2015.1791

Glisky, E. L. (2007). Changes in Cognitive Function in Human Aging. In *Brain Aging: Models, Methods, and Mechanisms*. CRC Press/Taylor & Francis. https://www.ncbi.nlm.nih.gov/books/NBK3885/

Gobet, F., & Sala, G. (2023). Cognitive training: A field in search of a phenomenon. *Perspectives on Psychological Science*, *18*(1), 125–141. https://doi.org/10.1177/17456916221091830

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, *14*(11), 497–505. https://doi.org/10.1016/j.tics.2010.08.005

Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in Geriatric Medicine*, *29*(4), 737–752. https://doi.org/10.1016/j.cger.2013.07.002

Hawes, Z. C. K., Gilligan-Lee, K. A., & Mix, K. S. (2022). Effects of spatial training on mathematics performance: A meta-analysis. *Developmental Psychology*, *58*(1), 112–137. https://doi.org/10.1037/dev0001281

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-skinnerian account of human language and cognition*. Springer US. https://doi.org/10.1007/b108413

Hayes, S. C., & Hayes, L. J. (1992). Verbal relations and the evolution of behavior analysis. *American Psychologist*, *47*(11), 1383–1395. https://doi.org/10.1037/0003-066X.47.11.1383

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573. https://doi.org/10.3758/s13428-019-01283-5

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *20*(10), 1727–1736. https://doi.org/10.1007/s11136-011-9903-x

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Kamin, S. T., & Lang, F. R. (2020). Internet use and cognitive functioning in late adulthood: Longitudinal findings from the survey of health, ageing and retirement in europe

(SHARE). *The Journals of Gerontology: Series B*, *75*(3), 534–539. https://doi.org/10.1093/geronb/gby123

Kelly, M. E. (2020). The potential of a relational training intervention to improve older adults' cognition. *Behavior Analysis in Practice*, *13*(3), 684–697. https://doi.org/10.1007/s40617-020-00415-0

Krivanek, T. J., Gale, S. A., McFeeley, B. M., Nicastri, C. M., & Daffner, K. R. (2021). Promoting successful cognitive aging: A ten-year update. *Journal of Alzheimer's Disease*, *81*(3), 871–920. https://doi.org/10.3233/JAD-201462

Lobo, E., Gracia-García, P., Lobo, A., Saz, P., & De-la-Cámara, C. (2021). Differences in trajectories and predictive factors of cognition over time in a sample of cognitively healthy adults, in Zaragoza, Spain. *International Journal of Environmental Research and Public Health*, *18*(13), 7092. https://doi.org/10.3390/ijerph18137092

May, R. J., Tyndall, I., McTiernan, A., Roderique-Davies, G., & McLoughlin, S. (2022). The impact of the SMART program on cognitive and academic skills: A systematic review and meta-analysis. *British Journal of Educational Technology*, *53*(5), 1244–1261. https://doi.org/10.1111/bjet.13192

McLoughlin, S., Tyndall, I., & Pereira, A. (2020). Convergence of multiple fields on a relational reasoning approach to cognition. *Intelligence*, *83*, 101491. https://doi.org/10.1016/j.intell.2020.101491

McLoughlin, S., Tyndall, I., & Pereira, A. (2022). Relational operant skills training increases standardized matrices scores in adolescents: A stratified active-controlled trial. *Journal of Behavioral Education*, *31*(2), 298–325. https://doi.org/10.1007/s10864-020-09399-x

McLoughlin, S., Tyndall, I., Pereira, A., & Mulhern, T. (2021). Non-verbal IQ gains from relational operant training explain variance in educational attainment: An active-controlled feasibility study. *Journal of Cognitive Enhancement*, *5*(1), 35–50. https://doi.org/10.1007/s41465-020-00187-z

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. https://doi.org/10.1037/a0028228

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *11*(4), 512–534. https://doi.org/10.1177/1745691616635612

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The repeatable battery for the assessment of neuropsychological status (RBANS): Preliminary clinical validity. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 310–319. https://doi.org/10.1076/jcen.20.3.310.823

Raven, J. C. (1936). *Raven Standard Progressive Matrices (RSPM)*. APA PsycTests.

Roche, B., Cummins, J., Cassidy, S., Dillon, A., Moore, L., & Grey, I. (2023). The effect of SMART relational skills training on intelligence quotients: Controlling for individual differences in attentional skills and baseline IQ. *Journal of Contextual Behavioral Science*, *28*, 185–197. https://doi.org/10.1016/j.jcbs.2023.03.012

Sala, G., Aksayli, N. D., Tatlidil, K. S., Tatsumi, T., Gondo, Y., & Gobet, F. (2019). Near and far transfer in cognitive training: A second-order meta-analysis. *Collabra: Psychology*, *5*(1), 18. https://doi.org/10.1525/collabra.203

Salthouse, T. A. (2019). Trajectories of normal cognitive aging. *Psychology and Aging*, *34*(1), 17–24. https://doi.org/10.1037/pag0000288

Saraçlı, Ö., Akca, A. S. D., Atasoy, N., Önder, Ö., Şenormancı, Ö., Kaygisız, İ., & Atik, L. (2015). The relationship between quality of life and cognitive functions, anxiety and depression among hospitalized elderly patients. *Clinical Psychopharmacology and Neuroscience*, *13*(2), 194–200. https://doi.org/10.9758/cpn.2015.13.2.194

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Sun, H.-L., Chen, P., Bai, W., Zhang, Q., Sha, S., Su, Z., Cheung, T., Ungvari, G. S., Jackson, T., Feng, Y., & Xiang, Y.-T. (2024). Depressive symptoms and cognitive function in older adults: A cross-lagged network analysis. *Depression and Anxiety*, *2024*, 6166775. https://doi.org/10.1155/2024/6166775

Thirus, J., Starbrink, M., & Jansson, B. (2016). Relational frame theory, mathematical and logical skills: A multiple exemplar training intervention to enhance intellectual performance. *International Journal of Psychology & Psychological Therapy / Revista Internacional de Psicologia y Terapia Psicologica*, *16*(2), 141–155.

Tucker-Drob, E. M. (2011). Neurocognitive functions and everyday functions change together in old age. *Neuropsychology*, *25*(3), 368–377. https://doi.org/10.1037/a0022348

Velloso, V., Latgé-Tovar, S., Bomilcar, I., & Mograbi, D. C. (2025). Cognitive interventions for healthy older adults: A systematic meta-review. *International Journal of Clinical and Health Psychology : IJCHP*, *25*(1), 100538. https://doi.org/10.1016/j.ijchp.2024.100538 von Bastian, C. C., Belleville, S., Udale, R. C., Reinhartz, A., Essounni, M., & Strobach, T. (2022). Mechanisms underlying training-induced cognitive change. *Nature Reviews Psychology*, *1*(1), 30–41. https://doi.org/10.1038/s44159-021-00001-3

von Bastian, C. C., Hyde, E. R. A., & Jiang, S. (2023). Tackling cognitive decline in late adulthood: Cognitive interventions. *Current Opinion in Psychology*, 101780. https://doi.org/10.1016/j.copsyc.2023.101780

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Yesavage, J. A., & Sheikh, J. I. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*, *5*(1–2), 165–173. https://doi.org/10.1300/J018v05n01_09

Contributions

Contributed to conception and design: MD, JC, JDH

Contributed to acquisition of data: ZM, MD

Contributed to analysis and interpretation of data: ZM, JC

Drafted and/or revised the article: ZM, JC, JDH

Approved the submitted version for publication: ZM, MD, JC, JDH

Acknowledgements

We would like to thank Selma Abd El Motleb Omar, Inne De Backer and Rozerin Bozkurt for their contribution to the data-collection.

Funding Information

This study was funded by the Methusalem grant awarded to Jan De Houwer (reference number: BOF22/MET_V/002 and code 01M00209).

Ethics Statement

This study was approved by the ethical committee of the Faculty of Psychology and Educational Sciences with reference number 2020/47.

Competing Interests

The authors declare that they have no competing interests.

Data Accessibility

Experimental materials (excluding copyrighted measures), data, and analysis scripts (in R) are openly available at https://osf.io/sdxv5/overview?view_only=49940b32ba4643c6a29ee40f0706e5cf