Using relational operant training in older adults with subjective cognitive complaints

Jamie Cummins, Marie Delabie, Selma Abd El Motleb Omar, Inne De Backer, & Jan De Houwer

Ghent University, Belgium

*Pre-registration*

**Author note**

JC, MD, SO, IDB, & JDH, Department of Experimental Clinical and Health Psychology, Ghent University. This research was conducted with the support of Grant BOF16/MET\_V/002 to Jan De Houwer. Correspondence concerning this article should be sent to [jamie.cummins@ugent.be](mailto:jamie.cummins@ugent.be).

Using relational operant training in older adults with subjective cognitive complaints

Recent research has strongly suggested that relational operant training can improve the intellectual abilities of various populations, particularly schoolchildren. Others have further suggested that relational operant training such as SMART (Strengthening Relational Abilities with Relational Training; Cassidy et al., 2011) should be tested in other populations which may benefit from it, such as older adults (Kelly, 2020). Indeed, given the promising results of SMART thus far, it may well be that this training can succeed where many others have failed (cf. Gobet & Sala, 2022). However, SMART studies are not without their pitfalls: they have generally suffered from the issues which plague many contemporary studies in psychological science, such as a lack of preregistered methods and analyses, small sample sizes, and weak or absent control conditions.

With this study, we aim to simultaneously test SMART in a population of older adults, as suggested by Kelly (2020), while also overcoming the issues present in previous research using SMART (as evidenced also by this preregistration document). We focus on older adults who experience self-reported cognitive complaints in the absence of a neuropsychological condition, given that this population (i) represent a risk group for the onset of dementia or other neuropsychological impairments, (ii) experience distress as a function of their self-reported complaints, and (iii) have a possibility for recovery from these issues. We will compare the impact of SMART on a variety of outcomes for this population compared to a “gold-standard” cognitive training, dual n-back training. The range of outcomes will assess both performance-based and self-reported cognitive skills which range on a spectrum from closely- to distally-related from the skills trained in SMART, allowing also for a test of SMART in terms of far-transfer effects. Participants will complete 8 weeks of either SMART or dual n-back, with assessment batteries completed before and after this course of training. Generally, we expect that participants should show improvements on the various outcome measures such that SMART improves scores to a greater extent than dual N-Back.

**Method**

**Sample**

Data will be collected from participants in Belgium, collected via online and in-person advertising. Participants will participate on a voluntary basis.

**Planned sample size & stopping rules.** Our primary analyses will consist of a series of 2 (condition: experimental vs. control) x 2 (timepoint: T1 vs. T2) mixed ANOVAs. These analyses will use (i) overall RAI scores, (ii) fluid IQ scores, and (iii) global RBANS scores as dependent variables. We conduct power analyses for these analyses using the *Superpower* Rpackage (Caldwell et al., 2020).

We expect that interaction effect sizes will differ as a function of the dependent variable. For instance, in comparing relational training and dual n-back on performance on the RBANS, there will likely be a smaller effect in comparison to on the RAI (abilities on which are explicitly trained on SMART vs. dual N-Back). In general, our smallest effect size of interest (SESOI; Anvari & Lakens, 2021) for the critical interaction effect is just under a medium Cohen’s d effect size (i.e., Cohen’s d = 0.4). Assuming at minimum strong correlation between measures at each timepoint (i.e., r = 0.5; this is consistent with previous findings on each of these measures), 100 participants per condition would provide us with 80% power to detect an interaction effect of d = .40, and 94% power to detect a medium interaction effect (i.e., Cohen’s d = 0.50). Based on reviews of the cognitive training literature (e.g., Bahar-Fuchs et al., 2013), we expect an attrition rate of around 10-15%. Therefore, we will endeavor to collect 230 participants in total. Notably, our population of interest is relatively difficult to access (in comparison to undergraduate populations), and the assessments and training required for this study are time-consuming. As such, there is a possibility that we will not reach our planned sample size. However, we will aim to collect data from as many participants as possible in order to reach our planned N and will transparently report if this was not the case in any subsequent publication.

**Inclusion criteria*.*** Aged over 65 years, ability to use a computer.

**Exclusion criteria.**Clinical diagnosis of dementia and/or disorders which may affect ability to complete the training (e.g., attentional dysfunction), or using medication which can influence cognition (either inhibitory or facilitatory). Participants who experienced a mild TBI in their younger years may be 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).

**Design**

A single-blind randomized active control trial, with participants randomly allocated to either the SMART or Dual N-Back treatment group. Our primary outcome measures consist of the overall scores on each of the three performance-based measures used in the study (i.e., the RAI, SPM, and RBANS). Our secondary outcome measures consist of the self-report scales.

**IVs.**

1. Treatment group (SMART vs. Dual N-Back).

**Primary Outcome Measures.**

1. Fluid IQ (assessed using Raven’s Progressive Matrices; RPM).

2. Relational abilities (assessed using the Relational Abilities Index+; RAI)

3. Cognitive functioning (measured using the Repeatable Battery for the Assessment of Neuropsychological Status; RBANS).

**Secondary Outcome Measures.**

1. Daily functioning (measured using the BRIEF-A)
2. Depression (measured using the GDS)
3. Quality of Life (measured using the EuroQol-5D-5L)

**Procedure**

Participants will firstly complete the battery of assessments. Then, participants will be randomly assigned to either the SMART or dual N-Back training conditions. Following 8 weeks of the training to which they have been assigned (3 sessions per week for 30 minutes each), participants will complete the battery of assessments once again (alternate forms will be completed by participants for the RBANS).

**Measures**

**SMART.** Our SMART training is programmed in lab.js and consists of 122 levels of relational training in a syllogistic reasoning-style presentation. Each trial in each level of training consists of between one and three statements regarding the relation between various nonsense stimuli (e.g., CUG is the same as VEK; VEK is opposite to JOM) followed by a question regarding the relation between stimuli within the relational network presents in the preceding sentences (e.g., is CUG the same as JOM?). On each trial, participants are required to respond either “yes” or “no” to this question within a pre-specified timeframe (30s). In every trial, three- or four-letter nonsense syllables are used as sentence subjects/objects. Critically, these nonsense syllables are generated randomly, and nonsense syllables are never repeated within the same level of training.

Across the 122 levels of SMART, the degree of difficulty is varied along a number of dimensions. Firstly, different relations are used across the levels: in Levels 1-29, same/opposite relations are used; in Levels 30-55, more than/less than relations are used; in Levels 56-81 before/after relations are used; in Levels 82-103 hierarchical relations are used; and in Levels 104-122 mathematical relations are used. Within the levels of each of these types of relations, complexity is varied through changing the number of sentences presented, the order in which they are presented (i.e., linearly, or randomly), the variation in the relations presented (e.g., all same or opposite relations vs. a mix of same and opposite of 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?).

Each level of SMART consists of a training phase and a testing phase. In the training phase, participants are provided with corrective feedback after responding on each trial. Participants are presented with trials from this level indefinitely until they achieve a criterion accuracy of 16 correct in-a-row. Once participants have achieved this criterion, they are cycled to a testing phase. This phase consists of 16 trials wherein no feedback is presented to participants. If participants successfully answer all 16 trials of the testing phase correctly, then they are moved on to the next level of training. Otherwise, they are recycled back to the training phase of the current level.

**Dual N-Back.** Dual N-Back training (our version programmed and implemented using lab.js) consists of the simultaneous presentation of (i) a small, colored square in one of 8 locations on a larger 3x3 square grid (i.e., excluding the center square), and (ii) one of several auditory stimuli consisting of a single letter. On a given trial, participants are asked to respond to both the visual stimulus and the auditory stimulus on the basis of whether their position/identity matches that of previous visual/auditory stimuli N trials ago, where N gradually increases across stages of the training. Participants are required to press the “A” key on the keyboard for a match in the position of the squares, and the “L” key for a match in the auditory stimuli. For example, on a 1-back stage, a first trial may present participants with a square in the top-right corner and the spoken word “S”, and the next trial may present participants with a square in the top-left corner and the same spoken word “S”. In this case, participants would be required to press “L” (because the auditory stimuli match) but not “A” (because the position of the squares did not match). On a 2-back stage, participants would need to emit these responses based on the identity of not the previous trial, but the trial before that. Each training stage consists of 20 + N trials, and participants pass on to the next training stage only if they emit less than 3 mistakes in the previous stage (otherwise they are recycled back to the complete the same training stage as the one they had just completed). Unlike SMART (which tracks participants’ progress across sessions), participants begin each session of dual n-back training at the 1-back stage.

**Raven’s Progress Matrices (RPM).** The RPM consist of 60 perceptual-analytic reasoning problems. Each problem consists of a matrix wherein features of different elements of the matrix change in some unspecified manner, which the participant is required to identify (through selecting the “next piece” in the progression of the matrix from one of several options). In each case, as its name implies, the problems increase progressively in difficulty from one problem to the next. Although the RPM can in principle be completed without a time limit, we will impose a 40-minute time limit for the completion of the measure.

**Relational Abilities Index (RAI).** The RAI we will administer (from Cummins et al., 2022) will consist of 128 unique syllogistic reasoning problems (16 trials per relation) probing the following relations in the following order: same/opposite, same/different, more/less, before/after, contains/is within, analogy, perspective-taking, and mathematical (see also Colbert et al., 2020). The task will take approximately 30 minutes to complete. Each trial follows a similar format to trials in SMART: beginning with the presentation of 1-4 relational premises, a question based on these premises presented, and a requirement to respond either “YES” or “NO”. Participants are required to respond on each trial within a 30s time limit.

**RBANS.** The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph et al., 1998) is a battery designed to assess cognitive functioning, consisting of 12 subtests assessment 5 cognitive domains in older adults: Immediate Memory, Delayed Memory, Language, Visuospatial/Construction, and Attention. The entirety of the battery takes approximately 30 minutes to administer.

**BRIEF.** The Behavior Rating Inventory of Executive Functioning is a 75-item scale aimed at assessing the executive functioning of patients in their daily lives.

**GDS.** XXX

**EuroQol-5D-5L.** XXX

**Research Questions**

**Primary.**

*RQ1.*Are increases in relational abilities from T1 to T2 greater in the SMART condition than in the N-Back condition?

*RQ2.*Are increases in fluid IQ scores from T1 to T2 greater in the SMART condition than in the N-Back condition?

*RQ3.* Are increases in overall RBANS scores from T1 to T2 greater in the SMART condition than in the N-Back condition?

**Secondary.**

*RQ4.*What are the correlations between each of the primary and secondary measures at baseline?

*RQ5.*Are increases in BRIEF-A scores from T1 to T2 greater in the SMART condition than in the N-Back condition?

*RQ6.*Are increases in GDS scores from T1 to T2 greater in the SMART condition than in the N-Back condition?

*RQ7.*Are increases in EuroQol scores from T1 to T2 greater in the SMART condition than in the N-Back condition?

*RQ8***.** Are similar expected patterns observed on any of the 5 subscales of the RBANS?

*RQ9.*Do any of the RAI subscales differentially correlate with any of the other primary and secondary measures?

**Analyses**

For all analyses involving ANOVA, we will control for baseline scores on the corresponding dependent variable to control for possible regression to the mean effects. Given that we are uncertain of the causal model underpinning the relationship between all our variables, we do not plan at present to control for any other variables at present (cf. Cinelli et al., 2022). For each ANOVA, we will also likely follow-up the interaction effect with post-hoc tests to unpack the shape of this interaction in more detail. In each instance, we will use Tukey’s HSD for this.

**Analytic strategy**

**Primary analyses.**

For RQ1, we will conduct a mixed between-within ANOVA, using overall RAI scores as DV, condition (SMART vs. dual n-back) as one independent variable, and timepoint of IQ assessment (T1 vs. T2) as the second independent variable. We expect a significant interaction effect, such that the change in IQ from T1 to T2 is greater for SMART compared to dual n-back. For RQ2, we will conduct an identical analysis to RQ1, but now using fluid IQ score (as derived from scores on Raven’s Matrices) as the dependent variable. For RQ3, we will employ similar ANOVAs to those for RQ1 and RQ2, but instead using RBANS overall score as the DV.

**Secondary analyses.**

For RQ5, we will compute Pearson correlations for each combination of the primary and secondary DVs (using only the overall scores in each of the measures). At present we do not plan to conduct hypothesis tests for this RQ. However, if we do so, we will use the *cocor* package in R.

RQs 6 through 8 follow a similar pattern to the tests for RQs 1-3. For RQ8, 5 ANOVAs would need to be conducted. In this case, we will utilise Holm-Bonferroni correction on the alpha levels of these tests to control for Type I error. Finally, for RQ9, we will follow a similar approach to RQ5, but now including each of the subscales of the RAI instead of its overall score.

**Potential Exploratory Analyses.**

In this study, some participants will complete assessments in-person, and some will complete assessments online. One potential exploratory question in this regard would be to assess the impact of this factor on performances: both in terms of whether there are differences at baseline in the magnitude of scores, and whether training effects are greater for those who complete assessments in-person vs. online. We do not have firm predictions of whether such an effect will be present or not, and are unsure of the specific number of participants who will complete assessments online (our preference will be for assessments to be completed in-person wherever possible). Depending on the distribution of this variable, we may conduct exploratory analyses in this regard.

Depending on rates of drop-out, we may also examine whether drop-out rates are greater for SMART vs. dual N-Back.