**Training Semantic Long-Term Memory Retrieval transfers to executive function and reading fluency**

# **Abstract:**

The retrieval of information from long-term memory is a fundamental cognitive ability, crucial for most aspects of successful human functioning. Whether and how long-term memory retrieval (LTMR) can be improved with training has clear societal importance but also theoretical value for furthering our understanding of underlying mechanisms. Here, we provide electrophysiological evidence for the plasticity of semantic LTMR. Thirty-five university students were randomly assigned to adaptive semantic LTMR training (using a Posner task) or to a non-adaptive version of the training. Before and after training they were assessed on measures of semantic LTMR, working memory, central executive function (interference control, switching), reading fluency, and fluid intelligence. Adaptive LTMR training (relative to non-adaptive training) led to significant improvements in semantic LTMR. The intervention group (in contrast to the control group) also showed a significant reduction in the mean amplitude of the N400 ERP component and 700-1000ms measured during a semantic LTMR task, suggesting that changes in retrieval occurred at an early/automatic point and retrieval processing in semantic processing. Moreover, transfer effects were observed for switching, working memory and reading fluency, but not for interference control or fluid intelligence. These results point to the plasticity of semantic LTMR, and suggest that improvement in this ability can transfer to other domains for which LTMR is key.

**Key words:** long-term memory retrieval (LTMR), working memory, transfer, training

# **Introduction**

Long-term memory retrieval (LTMR) is an essential component of higher-order cognitive ability, it plays a fundamental role in many cognitive tasks (Cowan, 2010) and deficient LTMR is the leading cause of forgetfulness (Shiffrin & Atkinson, 1969; Juola et al., 1971; Nobel & Shiffrin, 2001). Over the last two decades, there has been continued interest in the value of cognitive training to improve long-term memory retrieval (LTMR) (Anderson et al., 2018; Finn & McDonald, 2015; Jennings et al., 2005; Ma et al., 2020; Stamenova et al., 2014). However, extant research is not comprehensive and thus the plasticity and generalisability of training LTMR remain unclear.

LTMR has long been categorized into different types. Tulving (1972), for example, proposed that LTMR should be divided into episodic (for events) and semantic (for facts). The majority of previous LTMR training studies have specifically targeted episodic LTMR (Jennings et al., 2005; Stamenova et al., 2014; Finn et al., 2015; Anderson et al., 2018). For example, Jennings et al (2005) trained “consciously-controlled recollection” via a continuous recognition task, in which participants were required to recollect repeating concrete nouns over increasing lag intervals, twice a week for three weeks. In each session, participants had the task of indicating which nouns were encountered in a study phase by responding “yes” to old nouns and “no” to new nouns, both at their first and presentations during a test phase. To accurately respond, participants had to rely on recollection, a consciously controlled process that involves retrieving specific details from a previous event. More specifically, they had to recollect that a repeated word was first presented at test rather than at the study phase, and recollect whether they had already encountered it at test. Importantly, the number of intervening nouns between the first and second presentation was “adaptively” increased as performance improved. Following training, recollection performance improved relative to a recognition practice and no-contact control group. Further to this, performance on “near transfer” tasks (i.e., n-back (working memory) and self-ordered pointing tasks, which entailed retrieval of contextual information similar to the training task) and “far transfer” tasks (i.e., digit symbol substitution, a speed of processing measure which has less overlap with the training task) also improved relative to the control groups. Based on this, the authors argued that the procedure improved the act of recollection rather than teaching a specific task-based strategy. Demonstration of such transfer effects is important for clinical and/or pedagogical utility: Training a particular skill should not only benefit that skill but also produce generalizable benefits that extend to proximal and more distal contexts.

Few studies have explicitly focused on training *semantic* LTMR, which involves automatic processes of retrieval that are more under the control of stimuli rather than intention (Jacoby, 1991). In one exception, Ma et al. (2020) used the Posner mixed task (i.e., adapted from Wang et al., 2017 and see Posner & Mitchell (1967) for details of the original posner task) to adaptively train semantic LTMR over 12 days. The Posner task comprised 4 levels: Levels 1 and 2 include lower grade retrieval tasks (i.e, retrieving physical properties or implications concerning items, such as whether the items share a meaning), whereas Levels 3 and 4 require retrieval at a higher grade (i.e., retrieving conceptual information, such as whether the items have the same odd-even property). Level 1 and 2 tasks were used at training; Level 3 and 4 tasks were used as a measure of semantic LTMR at test, as a way to measure near transfer effects. To capture far transfer effects, Ma et al assessed a range of cognitive abilities previously linked to semantic LTMR before and after the start of training, including fluid intelligence (Raven’s Matrices), updating (working memory), interference control (Stroop), and a Switching task. Post training improvements were observed for semantic LTMR as well as for interference control, updating, and fluid intelligence. In fact, most cognitive tasks have the potential to draw semantic LTMR to some extent, but in the selection of transfer tasks in the cognitive training research, transfer tasks are mainly focused on other abilities that have established a clear association with the trained ability. Given that fluid intelligence, interference control, and switching abilities have been closely associated to semantic LTMR (Cowan，2010; Unsworth & Engle,2007; Wang et al., 2017), Ma et al.(2022) argued that the semantic LTMR training not only improved performance on direct measures of the same skill but also improved performance on tasks that draw on this skill in different contexts. For example, interference control not only involves volitional efforts to exclude or suppress task-irrelevant information it also requires efficient retrieval of task relevant information (Nigg, 2000; Unsworth et al., 2013; Shiffrin et al., 1969; Juola et al., 1971).

However, the design of Ma et al (2020) has key limitations that constrain claims that the semantic LTMR intervention led to training or transfer effects. First, Ma et al utilized a control intervention that involved a vastly different sand painting task. It is therefore possible that the observed improvements following Posner training were a consequence of training reaction time (given this was not a component of the control intervention) rather than training LTMR *per se*. A recent meta-analysis on working memory training argued that using an active control condition is crucial for making claims as to the locus of the improvement (Melby-Lervåg et al., 2016). A good active control intervention should adopt the same materials and procedures as the training task, without the “adaptive” element (i.e., the difficulty level increasing with improvements in performance), and/or maintain consistency in all aspects of the training with the exception of non-involvement of the trained ability.

Another key limitation of Ma et al (2020) is that purely behavioural measures of the effect of the intervention were included (e.g., reaction time and correctness) which are arguably prone to practice and/or strategic effects. Studies of working memory training have used neuroimaging and electrophysiological measures alongside behavioural tasks, to capture more objective changes in performance and reveal information on underlying neurological mechanisms. Such studies have found that behavioural improvements on working memory tasks following training are associated with enhanced activation in brain regions associated with working memory performance (Westerberg & Klingberg, 2007; Olesen et al., 2004; Jolles et al., 2010). Other studies have explored changes in electrophysiological performance when completing specific tasks before and after the training by capturing event-related potentials (ERPs) with electroencephalography (EEG). For example, Zhao et al. (2020) found changes in electrophysiological indicators such as mean wave amplitude of P1, N170, P2, and N2 in individuals after systematic working memory training than compared to pre-training. Based on these findings the authors concluded that the WM training enhanced individuals’ attentional control mechanisms which underlined the training and transfer effects. Thus, the ERP technique, capturing changes in the brain responses involved in trained and transfer tasks, has been argued to provide an important direct measure of brain responses with high temporal resolution, as a complement to behavioural measures.

***The present study***

This study examined the extent to which semantic LTMR can be trained and crucially whether the effects have the potential to transfer to near and far domains. Following Ma et al (2020), we used the Posner task to train LTMR. To assess near-transfer, we used the Semantic Object Retrieval Test (SORT). The SORT was selected as it has been argued to rely on semantic LTMR (Fratantoni et al., 2017; Wieser & Wieser, 2003), requiring participants to indicate whether a presented word pair elicits the retrieval of an object name. Brier et al. (2008) first explored ERP metrics for characterizing semantic memory retrieval as captured by the SORT. The SORT required subjects to judge whether the presented word pairs pointed to the same object (e.g., “desert” and “humps”). The trials that pointed to the same object were referred to as “semantic retrieval trials”, whereas trials that did not point to the same object were termed “non-semantic retrieval trials”. The ERP markers characterizing semantic memory retrieval were obtained by comparing the ERPs of subjects completing semantic and non-semantic retrieval trials. The results revealed that the mean wave amplitude of semantic and non-semantic retrieval trials were significantly different in the left frontotemporal electrodes (F7, FT7, T7) starting from 700 ms until 1000 ms. The authors concluded that this component characterized semantic memory retrieval ability. Subsequently, studies on specific subject conditions have reiterated the validity of this ERP component. Chiang et al. (2015) compared the ERPs of patients with amnestic mild cognitive impairment to the ERPs of normal individuals whilst they completed the SORT. They did this to verify whether the semantic memory deficit in patients with mild cognitive impairment is due to semantic memory retrieval deficits. The key findings were that, in contrast to normal controls, the patients showed lower correctness on the SORT and an increase in the frontal-parietal scalp potential that differentiated retrieval from non-retrieval trials between 900-1000ms post-stimulus onset. This suggests, therefore, that the mean wave amplitude in the 900-1000ms time window may capture semantic memory retrieval ability, but leaves open the question of whether this ERP is sensitive to semantic LTMR training.

As discussed above, working memory, interference control and fluid intelligence have been the focus of previous studies examining transfer effects in the context of semantic LTMR training (e.g., Ma et al., 2020), with studies claiming that LTMR is key to these abilities. Beyond this, there is clear value in assessing real-world skills for which LTMR plays a well-established role, such as reading fluency. Ericsson and Kintsch (1995) suggested that highly-complex skills such as reading require the support of semantic LTMR, and this claim has gained support from subsequent studies. For example, Brenner (2010) found that deficits in LTMR can lead to a reduction in reading speed. Fernández et al (2014) attributed suboptimal eye movements during reading in Alzheimer’s patients to semantic LTMR impairments. Nash and Unsworth (2019), in a systematic review, also highlighted the relationship between semantic LTMR and reading fluency. Therefore, there is an accumulation of research pointing to a link between semantic LTMR and reading fluency, eliciting the hypothesis that improving semantic LTMR may have benefits for this vital component of reading.

To address these empirical gaps, in the present study, participants were randomly assigned to an adaptive training condition or an active control condition. The training group completed 12 days of adaptive Posner task training (which involved Levels 1 and 2, as described above) whilst the active control condition performed the same computerized task but without the adaptive element (in keeping with the recommendations of Melby-Lervåg et al., 2016). While it may be debatable that the training task only involved levels 1 and 2, our aim was to ensure consistency with the training task used in the study by Ma et al. (2020), as one of our main objectives was to test the reliability of their findings. Additionally, the different levels of the Posner task were considered homogeneous but with different levels of difficulty. Using only levels 1 and 2 as training tasks would, on the one hand, somewhat avoid the influence of the practice effect on the training effect, and, on the other hand, if improvement can be observed at higher levels of the task by training at lower levels, which better supporting that the LTMR ability can be improving by training. The training task was administered by the same experimenter in the same environment for 12 consecutive days. Transfer effects were assessed using the higher levels of semantic LTMR Posner task, the SORT (captured by parallel behavioural and ERP responses), working memory, reading fluency, and a general fluid intelligence task. By comparing performance on these tasks pre and post training in each group, and the intervention vs control groups post training, we tested the following hypotheses:

For the behavioural measures, 1) The semantic LTMR ability would improve through training, which would be reflected by greater pre-/post-test improvements in performance on the semantic LTMR task (i.e., the Posner task); Semantic LTMR training will have significant transfer effects to 2) working memory ability; 3) switching ability; 4) interference control ability; 5) reading fluency; 6) and fluid intelligence.

For ERP measures, we expect to observe significant training effects in the 700-1000ms time window, whereas no training-specific effects are observed for the other two-time windows (300-500ms and 500-700ms) and for the P2 components.

# **Methods**

## ***Participants***

## The minimum number of participants required was 28-66, as calculated by G\*power (selected a F test as ‘ANOVA: Repeated measures, within-between interaction’ and parameters: effect size=0.4-0.25, α=0.05, power=0.8). We recruited 60 college students (Mean age=22.8, SD age =1.98; male: 26, female: 34) whose native language was Chinese. Subjects were randomly assigned to the training condition (*n* = 30) and the control group (*n* = 30). Crucially, there were no significant group differences in any of the pre-training tasks (**Table 2)**. Four participants in the training condition and two participants in the control condition were excluded from behavioral data analysis, because they failed to comply with the experimental instructions. Additionally, six participants in the training condition and ten participants in the control condition were excluded from ERP data analysis due to noisy (coughing or muscle artifacts), incomplete data or not enough trials for ERP averaging (Fewer than 15 trials). Therefore, for the analysis of the behavioural data, there were 54 subjects (training condition: *n* = 26, Mean age =20.62, SD age = 1.98, male = 7; control condition: *n* = 28, Mean age = 20.18, SD age = 1.74, male = 13), while for the ERP data, there were 40 subjects (training condition: *n* = 20; control condition: *n* = 18). All participants reported normal or corrected vision and no history of psychosis or neurological disorders. The study was approved by the local ethics committee and all experimental operations were conducted by following the approved guidelines. All participants provided informed consent and received cash compensation.

## ***Materials***

## ***Computerized training programs***

The training task from Ma et al. (2020) was used. Materials comprised two rules and character-pairing consisting of Arabic numerals (1~9) or Chinese-characters (一~九). The task involved responding to the character-pairing according to the given rules. If the character-pairing conformed to the rules participants were instructed to press the “J” key on a keyboard, if the pairing did not conform to the rules, they were instructed to press the “F” key. The rules were as follows: (1) Whether the physical properties are the same (e.g., 1-9, correct response “J”; 1-九, correct response “F”; (2) Whether the meanings are the same (e.g., 4–四, correct response “J”; 4-3, correct response “F”). The training task contained ten difficulty levels, which were determined by stimulus presentation time (i.e., presentation time decreased with each difficulty level). There were 32 trials in each difficulty level (16 trials in each of the two rules, and half of them conformed to rules). On each trial, a rule was first presented in the center of the screen. After reading the rule, the participants had to press the “space” key, then a fixation cross was presented for 500 ms, followed by a random character-pairing, which as shown in **Fig.1**. Participants had to respond while the stimulus was on the screen, otherwise the trial was regarded as an incorrect response. The next trial began immediately thereafter. Each subject had 10 opportunities to pass the task that have specific difficulty level every day, which means everyone must and only complete 10 times training (a total of 320 trials) everyday, and the entire task lasted about 25 minutes. Importantly, the training task was adaptive, such that task difficulty varied according to the performance of the participant (as captured by the number of incorrect trials at each difficulty level). See **Table 1** for specific difficulty levels. The score standard was as follows: Each participant in the training condition would begin training at the first difficulty level (1200ms). In each difficulty level, if the number of wrong responses was less than or equal to 3, they would progress to the next difficulty level. If the number of wrong responses was greater than or equal to 4, they would remain at that difficulty level for retraining (If they fail three time in a row, which also means that they have used 3 training opportunities and have 7 training opportunities remaining on this day, they would drop to the previous difficulty level). There were 10 opportunities for each participant every day. Finally, the last difficulty level of the participant would be recorded. The next time the same participant was trained, they would start from the last recorded difficulty level. The results of the participants’ daily training were determined by their final difficulty level on that day. For example, on the first day, if a participant progressed to the third difficulty level (1000ms) by 10 opportunities, his or her score was 3. On the second day, if he or she progressed from the third difficulty level to the fifth difficulty level (800ms) by 10 opportunities, his or her score was 5.



**Fig 1.** Schematic diagram of the trial flow in the training task.

**Table 1** Specific difficulty level

|  |  |
| --- | --- |
| Difficulty level | Duration (ms) |
| 1（easiest） | 1200 |
| 2 | 1100 |
| …… | …… |
| 9 | 400 |
| 10（hardest） | 300 |

The control task used the same procedures and materials as the training task. The only difference was that task difficulty did not vary with participant performance. The participants in the active control condition were required to complete the same number of trials at the same time and place every day as the training condition. During each trial, a rule was first displayed in the center of the screen. After reading the rule, participants had to press the “spacebar” key. Upon pressing, a 500 ms fixed cross was presented, followed by a random character pairing (response window). After the participants responded, they would receive feedback on whether they were right or wrong. The next trial started immediately after the feedback. The participants were given a break after every 32 trials. Thus, there was a total of 320 trials per participant every day. Participant in active control group would know their performance after every task day, including correctness of the reaction and the mean reaction time (RT) of the correct reaction.

### ***Pre- and post-test tasks***

This study used six tasks to measure transfer effects, all of which were prepared using the E-Prime 2.0 software (Psychology Soft-ware Tools, Pittsburgh, PA).

*Near Transfer*

*Posner task.* This task measured LTMR, referring to Level 3 and 4 tasks used in Wang et al., (2017). Stimuli comprised a rule and character-pairing consisting of Arabic numerals 1~9 and Chinese number 一~九. Subjects had to make a response in accordance with the character-pairing by pressing the “J” or “F” on the keyboard if the character-pairing conformed the rule or not, respectively. The Level 3 task involved making a parity (odd/even) judgement with pairs comprising Chinese characters or numbers (e.g. “1,2”; “一,二”); the Level 4 task involved the same judgement but for pairs consisting of Chinese characters and numbers (e.g. “1, 二”; “一, 2”). When the character-pairings were both odd or even, participants pressed the “J” key (e.g. 1,9; 1,九); when there was an odd number and an even number, participants pressed the “F” key (e.g. 1, 2; 1, 二). On each trial, a rule was first presented in the center of the screen; after reading the rule, the subjects had to press the “space” key, then a red fixation cross was presented for 500ms, followed by a character-pairing stimuli (response window). After the response, the next trial started immediately. After 16 practice trials, the main task was initiated, which comprised 32 trials. The stimulus presentation sequence was random. The dependent measure was the correctness of the reaction and the mean RT of the correct reaction, the higher the correctness and the faster the RT, the better the long-term memory retrieval ability.

*SORT.* This task was used to obtain ERP indicators of semantic LTMR. The task material consisted of word pairs that represent the object’s features (e.g., “desert” and “hump”). Participants were asked to determine whether the combination of the features depicted by the two words led to the memory concerning the same object (e.g., “desert” and “hump,” which together point to “camel”) or not (e.g., “mane” and “wings”). Moreover, the participants were informed that the object must be a specific object and not represent an association between the two words (e.g., “mane” and “wings” both pertain to living creatures). Trials that pointed to the same object were considered to be able to induce a semantic retrieval process (Assaf et al., 2006a). As such, they are referred to as “semantic retrieval trials”. Trials that could not point to the same object were called “non-semantic retrieval trials”. All word pairs used can be found in the supplementary material. The whole task consisted of 100 trials. There were 50 semantic retrieval trials and 50 non-semantic retrieval trials, presented randomly. Participants were required to press “J” for semantic retrieval trials and “F” for non-semantic retrieval trials. In each trial, a random gaze point “+” was presented for 2000–3000 ms. The word pair stimuli were then presented for 3000 ms, during which participants respond by pressing the appropriate key. The subsequent trial started after the presentation ended or after the participants pressed the key. There was a break opportunity (1.5 minutes) after every 25 trials, and the whole task lasted for approximately 10 minutes.

*Stroop task.* This task was used to measure interference control. The stimuli comprised Chinese characters (Hanzi) and strings (# # # #), each printed in a color that may or may not correspond to the actual meaning of the word (e.g., the word "red" printed in blue ink). The participant's task is to ignore the semantic content of the word and instead focus on responding to the ink color in which the word is presented as quickly and accurately as possible. Pressing “F” for red and “J” for green on a keyboard. The task contained three kinds of trials: congruent, incongruent, and neutral. On congruent trials, the Hanzi referring to the word “green” was printed in green and the character representing “red” was printed in red. On incongruent trials, the “green” Hanzi was printed in red and the “red” Hanzi was printed in green. Finally, on neutral trials, the symbols “####” were either printed in green or red. Each trial started with a fixation cross that was presented for 500ms, followed by a blank screen presented for 1000ms. Thereafter, the target stimulus (colored Hanzi or strings) was presented for 1500ms, followed by a blank screen that was presented for a variable duration lasting between 600ms and 1000ms. The next trial began immediately thereafter. The task comprised 18 practice trials and three blocks, each of which has 36 trials. The participants were invited to take a break between the blocks. Each block contained 12 congruent, 12 incongruent, and 12 neutral trials. The task lasted roughly 15 minutes. The dependent measure was the mean RT of correct responses, we computed the interference index by computing proportional cost (Verhaeghen, 2011 & 2014): (the mean RT of incongruent trials minus the mean RT of neutral trials) divided by the mean RT of neutral trials. A high index score represents weak interference control.

*Flanker task.* This task measured interference control ability. Stimuli comprised five arrows presented in the center of the screen. If the direction of the middle arrow was pointing to the left, the subjects had to press the “F” key, and if the direction was pointing to the right, the subjects had to press the “J” key. The task comprised trials of two kinds: congruent (→→→→→or←←←←←) and incongruent (→→←→→or←←→←←). For each trial, a fixation cross was first presented for 500ms, followed by a blank screen that was presented for a variable duration lasting between 300ms and 500ms, then the stimulus was presented for 1500 ms (response window), after that was a 1000 ms blank screen. The next trial began immediately thereafter. The task comprised 5 blocks, each of which had 32 trials (16 congruent), the first block was the practice block, and the latter four blocks comprised the main task. Participants were invited to take short breaks between blocks. This task lasted ~15 minutes. The interference index was computed by (the mean RT of incongruent trials minus the mean RT of congruent trials) divided by the mean RT of congruent trials. The higher the index score, the weaker the interference control ability.

*Number Updating task.* This task was used to assess working memory (updating) ability. A series of single numbers from 0 to 9 were presented in the center of the screen. The length of the digital sequence varied with the type of trial; there were four types of trials with 5, 7, 9, or 11 digits. The number of occurrences of each trial type was the same and presented in random order. Participants had to remember the last three numbers in sequence, regardless of the length of the digital sequence. For example, if the number presented was 7-3-2-4-6-2-1, participants should remember 7-73-732-324-246-462-621. Finally, when blank bars were presented on the screen, participants had to use the keyboard to enter the last three digits presented (that is, 6-2-1). Each trial began with the “+” fixation point, which was presented for 500ms. Subsequently, the number was displayed, and the empty screen was displayed during the trial. The duration of the empty screen varied from between 800 ms to 1200 ms. The presenting time (PT) of each number distinguishes the difficulty of the task. We defined 750ms as a difficult task that we used. The tasks consisted of three blocks; the first block was a practice block, comprising eight trials, each of which occurred twice randomly. The main task included two test blocks, each comprising 12 trials. The evaluation index was proportion of correct responses (only all of the numbers were correct, the response was regard as correct).

*Switching tasks.* This task was used to evaluate the ability to flexibly switch between tasks. The stimuli comprised red or blue numbers 1–9 (excluding number 5). Participants were asked to judge the size or parity of each number. The specific rules depended on the color of the number. In Task A, if the color of the number was red, a size judgment should be made (magnitude task). If the number was > 5, then “A” should be pressed. If the number was < 5, then “L” should be pressed. In Task B, if the color of the stimulus number was blue, a parity judgment should be made (parity task). If the number was odd, then “A” should be pressed. If the number was even, “L” should be pressed. The participants first performed two task blocks (task A and task B) until the correct rate reached 75%. Then, they completed 20 experimental blocks: 10 single-task blocks, each comprising 8 trials, and 10 mixed-task blocks, each comprising 17 trials. During mixed-task blocks, the participant had to switch between tasks A and B every second trial. The order of blocks was random, but there were restrictions on combining two single and two mixed task blocks. On each trial, the fixation cross was first presented for 1400 ms, then, the stimulus was presented until the subject responded. The next trial began immediately thereafter. The task lasted about 20 minutes. The dependent measure was the switching cost and mixed cost. The switching cost was computed by (mean RT on switch trials minus mean RT on non-switch trials) divided by mean RT on non-switch trials. The mixed cost was computed by mean RT on single-task blocks minus mean RT on repeat trials in switching blocks. The greater the dependent measures, the worse the switching ability.

*Reading Fluency tasks.* Reading fluency was measured with a three-minute rapid reading task (Lei et al., 2011). The task material consisted of 100 simple sentences, sorted in an ascending order with respect to the number of words in the sentence. Participants read aloud as many sentences as possible in three minutes. They were further asked to determine whether the meaning of the sentence was correct (i.e., consistent with objective facts or patterns). They had to press “J” if the meaning was correct and “F” if the meaning was incorrect. For example, in response to “the sun rises in the west” (by), participants should have pressed the “F” key; in response to “the swallow flies” (√), participants should have pressed the “J” key. The procedure ended automatically after three minutes. The number of sentences correctly answered by each participant and the average response time for correct responses were calculated.

*Raven's Advanced Progressive Matrices.* The Raven’s Advanced Progressive Matrix (Raven, 1983) test was used to measure fluid intelligence. The test comprised a series of diagrams or designs with a missing part. The participant needed to select the correct part from multiple options provided to complete the presented picture. Sixty items were divided into two equal parts by title. Participants were asked to complete the even-numbered items in the pre-training test and the odd-numbered items in the post-training test. Each participant was given a limited time (20 minutes) to complete the test. The dependent measure was the proportion of correctly solved items.

## ***Procedures***

The design of the present study was double-blind. Specifically, we used the following to ensure that the double-blind was performed correctly: 1) the experimenters responsible for supervising the subjects' completion of the training task and the control task each day were different, while these experimenters did not know whether they were responsible for the subjects in the training or the control group, as there was no difference between the names of the two tasks; 2) the subjects in the control group were also informed each day after completing the task that they had completed 10 training sets that day, although there was no specific level of difficulty; and 3) above all, the experimenters who performed the daily training and those who performed the pre and post-tests were different, and all of them did not come into direct contact with the handling of any of the data. Participants completed the pre-training tests in the following order: Posner task, Stroop task, Flanker task, Number Updating task, Reading Fluency task, Raven’s advanced progressive matrices, and SORT. They completed these in a span of 48 hours with adequate rest between each task. The training condition then completed an adaptive training process consisting of 12 working days. This took roughly 25 minutes per day in the school’s standard computer behaviour lab. The control condition completed the control task at the same time and place each day. All participants completed the post-training test in the same order as in the pre-training test immediately after completing the training or control process.

## ***EEG data acquisition and processing***

EEG data were acquired form 64 K/Cl electrodes positioned in a salt water cap from the EGI system (EGI, Electrical Geodesics, Inc., America) according to the extended international 10–20 systems and the amp300 amplifier of EGI company was used. Voltage from all electrodes during recording was referred to the CZ. Signals from all channels (sampling rate 1000 Hz) were filtered online with a 0.01–100 Hz bandpass and the impedance was maintained below 50 kΩ.

Data analysis was performed offline in Matlab 9.2.0 platform using EEGLAB 14.1.2 (Herron & Wilding, 2005). The original EEG signal was re-referenced to the averaged recordings of mastoids and the continuous data was band-pass filtered from 0.1 Hz to 30 Hz. ERPs were time locked to the word pairs presented and epochs were created from 100 ms prior to the occurrence of word pairs for 1000 ms. Independent component analysis (ICA) combined with the Adjust toolbox plug-in (Mognon et al., 2011) was applied to manually remove artifacts such as ocular artifact, ECG and EMG (Delorme & Makeig, 2004). Trials containing voltages exceeding the range ±75 μV were excluded from the analyses and baseline correction was applied again.

## ***Statistical analysis***

*Behavioral data.*For the Stroop and Flanker task, error trials and trials with an RT < 200 ms were excluded (because this is somewhat below the range of a biologically plausible response time for these tasks) prior to the RT analyses (Zhao & Jia, 2019). For the switching task, RT ≥ 4000 ms (This approach aimed to remove responses that were likely due to technical errors, participant disengagement, or other non-cognitive factors, and ensure the reliability of the cognitive switching task results) and error trials were excluded prior to the analyses. Prior to performing the main analyses, we first tested whether there was a difference between the trained and control participants on any of the pre-training tests, which was not the case (**Table2**). We used mixed design analysis of variance (ANOVA) to examine interactions between session and group, and ascertain the effectiveness of the intervention. Finally, for the transfer effect, group (training vs. control) and session (pre- vs. post-training) were used in the mixed design ANOVA. Significant interactions were followed up by analysis of simple effect analysis. An alpha level of 0.05 was adopted throughout. Partial eta-squared (*ηp2*) was used as an estimate of effect size. Post hoc pairwise comparisons with the Bonferroni adjustment were applied when significant interaction effects were observed. The Greenhouse–Geisser correction was used when Mauchly's test of sphericity was violated. It should be noted that the comparison correction is done through SPSS's built-in algorithm, so the *p*-value reported in this case is the corrected *p*-value.

*ERP data.* The trials with incorrect responses were excluded in the Sort task’s ERP data analysis. The number of trials in the averaged ERP for each condition are as following: for training group in pre-session: semantic retrieval condition, Mean = 43.1, non-semantic retrieval condition, Mean = 43.5; in post-session: semantic retrieval condition, Mean = 39.4, non-semantic retrieval condition, Mean = 40.9. For control group in pre-session: semantic retrieval condition, Mean = 42.1, non-semantic retrieval condition, Mean = 42.9; in post-session: semantic retrieval condition, Mean = 38.78, non-semantic retrieval condition, Mean = 42.3. The latency and electrode choices were based on existing research (Wieser et al., 2003) and visual inspection of condition-averaged waveform maps. Four-time windows of 230-290ms (P2 component), 300–500 ms, 500–700 ms, and 700–1000 ms were selected for data analysis. The main electrode sites in the left frontotemporal region were selected for analysis: F7, FT7, and T7. The amplitudes of the three selected electrode sites were averaged for statistical analysis. To perform statistical analysis in each of the four-time windows a three-way repeated-measures analysis of variance (ANOVA) was used: a 2 (group: training vs. control) by 2 (session: pre- vs. post-training) by2(retrieval condition: semantic retrieval vs. non-semantic retrieval), and dependent measure is mean amplitude.

Some of the subjects included in the analysis of behavioural data were not able to be analyzed for ERP data due to EEG data poor quality, which resulted in different sample sizes in the two analyses. Therefore, we have also conducted parallel behavioral analyses with only the participants that were included in the ERP analyses. Relevant results can be found in the support information (Table S1 and Table S2).

# **Results**

## ***Effects of training on LTMR***

The daily mean score of participants in the training condition and the mean difficulty levels reached after training are shown in **Fig.2**. The results showed that with the increase of training sessions, the mean difficulty level increased from 4.3 to 8.3, *F*(1,25) = 31.78, *p*< 0.001, *ηp2* = 0.566, indicating that the participants in training condition completed the training as required and made progress. Additionally, we assessed the performance of the control group at 12 sessions (**Support Information, Figure S1**). Note the point that we observed a significant drop in mean ACC in session 9 and then a return to normal levels on session 10. This may indicate that the control group was conscientious in completing the task in most of the sessions, but in some sessions the control group may not have been conscientious in completing the task due to a variety of reasons (e.g., fatigue and reduced attractiveness of the control task).

G:\攻击性投稿\QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY投稿\修改\8.21\figure\毕设训练组（实验2） [Mean Difficulty Level].tif

**Fig 2.** Mean (±SEM) difficulty level reached by partici-pants in training condition on each of the 12 training sessions.

## ***Transfer effects***

**Table 2** shows the training groups’ mean scores on the near- and far-transfer tasks during the pre- and post-training sessions. Details of the overall ANOVAs on the pre- and post-training performance measures are presented in **Table 3**. All tables are described in turn below.

**Table 2** Descriptive statistics (mean and SD) for the conditions’ pre- and post-training transfer task performance measures

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Condition** | **Pre-training** | **Post-training** |
| Posner Task  (response time, ms) | Training  Control | 988.64（108.00）  967.52（134.07） | **788.04#**（152.28）  860.24（144.61） |
| Posner Task  (correct proportion) | Training  Control | .89（.06）  .91（.05） | .89（.09）  .91（.05） |
| Raven | Training  Control | .68（.15）  .69（.12） | .63（.16）  .68（.14） |
| Reading Fluency  (number of sentence) | Training  Control | 72.35（8.29）  74.43（9.73） | **81.92\***（7.25）  77.25（9.18） |
| Reading Fluency  (RT) | Training  Control | 2402.51（425.82）  2298.16（426.47） | 2155.61（506.77）  2227.90（371.77） |
| Stroop | Training  Control | .09（.11）  .09（.10） | .06（.14）  .05（.11） |
| Flanker | Training  Control | .06（.04）  .06（.05） | .07（.06）  .08（.04） |
| WM updating | Training  Control | .85（.11）  .87（.15） | **.92\***（.09）  .87（.13） |
| Switching  (Switching Cost) | Training  Control | .24（.17）  .23（.13） | **.16\***（.11）  .21（.11） |
| Switching  (Mixed Cost) | Training  Control | 156.72（112.84）  158.84（94.19） | **98.42\***（79.75）  136.82（96.09） |

\*: The training group was significantly better behaved in the post-test than in the pre-test compared to the control group; #: Borderline significant.

**Table 3** Results of statistical analyses (ANOVAs) of transfer test data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Task  (dependent measure) | Factor | *dfs* | *F* | *p* | *ηp*2 |
| posner task  (response time) | Condition  Session  Condition by session | 1,52  1,52  1,52 | <1  87.53  8.04 | 0.445  **9.85×10-13**  **0.007** | 0.011  0.627  0.134 |
| posner task  (correct proportion) | Condition  Session  Condition by session | 1,52  1,52  1,52 | <1  <1  <1 | 0.366  0.968  0.886 | 0.016  0.000  0.000 |
| Raven | Condition  Session  Condition by session n | 1,52  1,52  1,52 | <1  2.68  <1 | 0.444  0.108  0.347 | 0.011  0.049  0.017 |
| Reading Fluency  (number of sentence) | Condition  Session  Condition by session | 1,52  1,52  1,52 | <1  87.58  26.00 | 0.571  **9.75**×**10-13**  **0.05**×**10-4** | 0.006  0.627  0.333 |
| Reading Fluency  (RT) | Condition  Session  Condition by session | 1,52  1,52  1,52 | <1  9.12  2.83 | 0.880  **0.004**  **0.099** | 0.000  0.149  0.052 |
| Stroop | Condition  Session  Condition by session | 1,52  1,52  1,52 | <1  4.04  <1 | 0.898  0.060  0.616 | 0.000  0.072  0.005 |
| Flanker | Condition  Session  Condition by session | 1,52  1,52  1,52 | <1  3.54  <1 | 0.662  0.076  0.804 | 0.004  0.064  0.001 |
| WM updating | Condition  Session  Condition by session | 1,52  1,52  1,52 | <1  5.94  7.39 | 0.708  **0.018**  **0.009** | 0.003  0.102  0.124 |
| Switching  (Switching Cost) | Condition  Session  Condition by session | 1,52  1,52  1,52 | <1  12.39  3.75 | 0.675  **0.001**  **0.058** | 0.003  0.196  0.067 |
| Switching  (Mixed Cost) | Condition  Session  Condition by session | 1,52  1,52  1,52 | <1  24.64  5.03 | 0.421  **0.08**×**10-4**  **0.029** | 0.012  0.322  0.088 |

### ***Posner task.*** A repeated-measures ANOVA was conducted on the Posner task reaction time metric for both intervention groups. There were no significant main effects of condition, but main effect of session, and significant interactions between group and session. Further simple effects analysis revealed that the two groups were significantly faster in the post-training test (Training group: *M*=788.04, *SD*=152.28; Control group: *M*=860.24, *SD*=144.61) than in the pre-training test (Training group: *M*=988.64, *SD*=108.00; Control group: *M*=967.52, *SD*=134.07). In the post-training session, the training group showed a faster trend than the control group, although it did not reach the significant level, *F*(1,52) = 3.19, *p* = 0.080, *ηp2* = 0.058, 95% CI = [-8.88, 153.27].

A repeated-measures ANOVA on the Posner task correctness metric for both the conditions of subjects showed no significant test-phase main effect. Similarly, there were no significant condition main effect nor significant interaction between the training condition and the test phase.

### ***SORT.*** A three-way repeated-measures ANOVA with 2 (condition: training condition, control condition) × 2 (session: pre-training, post-training) × 2 (retrieval condition: semantic retrieval, non-semantic retrieval) was applied within the four time windows previously selected. **Table 4** shows results of the ANOVA. **Fig. 3 shows** the average waveforms for semantic retrieval condition andthe topography of the effects for semantic retrieval condition. **Fig. 4** presents the average waveforms for the non-semantic retrieval condition and the topography for the non-semantic retrieval condition.

In the 300-500 ms time window, statistical analysis revealed that only the interaction effects of training condition by session and was significant. Simple effects analysis for training condition by session interaction effect revealed that the mean amplitude in the training group was significantly lower in the post-training test (*M*=0.13, *SD*=1.03) than in the pre-training test (*M*=3.44, *SD*=1.04), *F*(1, 36) = 8.86, *p =* 0.005, *ηp2* = 0.197, 95% CI = [1.06, 5.57]. However, the mean amplitude difference between the post-training test (*M*=1.80, *SD*=1.08) and the pre-training test (*M*=1.09, *SD*=1.09) was not significant for the control group, *F*s < 1.

In the 500-700 ms time window, a statistical analysis revealed that only the main effect of retrieval type was significant. Main effect analysis showed that in the post-training test, the mean amplitude under the semantic retrieval condition (*M* = -3.65, *SD* = 0.64) was significantly lower than that of the non-semantic retrieval condition (*M* = -1.68, *SD* = 0.61), *F*(1, 36) = 38.32, *p* < 0.001, *ηp2* = 0.516, 95% CI = [-2.62, -1.33].

In the 700-1000 ms time window, A repeated-measures ANOVA revealed that the interaction effect of training condition by session by retrieval type was significant. For training group, the mean amplitude under the semantic retrieval condition was significantly lower in the post-training test (*M*=-0.45, *SD*=1.05) than in the pre-training test (*M*=3.44, *SD*=1.35), *p* = 0.012, 95% CI = [0.95, 6.84]. However, the mean amplitude difference between the post-training test (*M*=0.71, *SD*=0.98) and the pre-training test (*M*=3.44, *SD*=1.15) was not significant for the non-semantic retrieval condition, *p* = 0.055, 95% CI = [-0.07, 5.52]. For control group, we did not found any significant effect (main effect of session: *F*s < 1; main effect of retrieval type: *F*(1, 17) = 1.24, *p* = 0.28; the interaction effect: *F*s < 1).

A statistical analysis of the mean P2 component amplitude revealed that the session main effect and retrieval type were significant. Additionally, the interaction effect of training condition by session and training condition by session by retrieval type also were significant. To explore the three-way interaction, follow-up simple effects analyses were conducted. For training group, two factor ANOVA revealed that the main effect of session (*F*(1, 19) = 10.56, *p* = 0.004) and retrieval type (*F*(1, 19) = 7.04, *p* = 0.002) were significant. Additionally, the interaction effect was also significant, *F*(1, 19) = 4.06, *p* = 0.050. Further simple effects analysis found that the mean amplitude under the semantic retrieval condition was significantly lower in the post-training test (*M*=3.56, *SD*=0.74) than in the pre-training test (*M*=5.73, *SD*=0.63), *p <* 0.001, 95% CI = [0.99, 3.36]. However, the mean amplitude difference between the post-training test (*M*=3.33, *SD*=0.71) and the pre-training test (*M*=4.52, *SD*=0.77) was not significant for the non-semantic retrieval condition, *p* = 0.053, 95% CI = [-0.2, 2.4]. For control group, we did not find any significant effect (main effect of session: *F*s < 1; main effect of retrieval type: *F*s < 1; the interaction effect: *F*(1, 17) = 1.73, *p* = 0.21).

### ***Stroop task.*** A repeated-measures ANOVA and did not reveal the presence of any significant main or interaction effects.

### ***Flanker task.*** A repeated-measures ANOVA did not reveal the presence of any significant main or interaction effects.

### ***Number Updating task.*** A repeated-measures ANOVA showeda significant main effect of session, no effect of training group, and a significant interaction between session and training group. Further simple effects analysis revealed that the accuracy in the training group was significantly higher in the post-training test (*M*=0.92, *SD*=0.09) than in the pre-training test (*M*=0.85, *SD*=0.12), *F*(1,52) = 12.81, *p* < 0.001 *ηp2* = 0.198, 95% CI = [0.03, 0.0.11]. However, there was no significant difference between the post-training test (*M*=0.87, *SD*=0.13) and the pre-training test (*M*=0.87, *SD*=0.15) for the control group, *F* < 1*.*

**Table 4** Results of statistical analyses (ANOVAs) of each time window for Sort task

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Time Window  /measures | Condition | Session | Retrieval Type | Condition by  Retrieval Type | Session by  Retrieval Type | Condition by  Session | Condition by Session  By Retrieval Type |
| 300-500 | *F*(1,36)=0.07  *p*=0.787  *ηp2*=0.002 | *F*(1,36)=2.60  *p*=0.116  *ηp2*=0.067 | *F*(1,36)=0.15  *p*=0.704  *ηp2*=0.004 | *F*(1,36)=3.60  *p*=0.066  *ηp2*=0.091 | *F*(1,36)=0.43  *p*=0.515  *ηp2*=0.012 | ***F*(1,36)=6.17**  ***p*=0.018**  ***ηp2*=0.146** | *F*(1,36)=3.27  *p*=0.079  *ηp2*=0.083 |
| 500-700 | *F*(1,36)=1.21  *p*=0.278  *ηp2*=0.033 | *F*(1,36)=0.09  *p*=0.760  *ηp2*=0.003 | ***F*(1,36)=38.32**  ***P* = 3.87×10-7**  ***ηp2*=0.516** | *F*(1,36)=3.74  *p*=0.061  *ηp2*=0.094 | *F*(1,36)=1.27  *p*=0.267  *ηp2*=0.034 | *F*(1,36)=2.86  *p*=0.099  *ηp2*=0.074 | *F*(1,36)=0.03  *p*=0.856  *ηp2*=0.001 |
| 700-1000 | *F*(1,36)=3.04  *p*=0.090  *ηp2*=0.078 | *F*(1,36)=0.66  *p*=0.423  *ηp2*=0.018 | *F*(1,36)=0.36  *p*=0.554  *ηp2*=0.010 | *F*(1,36)=0.11  *p*=0.746  *ηp2*=0.003 | *F*(1,36)=1.23  *p*=0.274  *ηp2*=0.033 | *F*(1,36)=1.33  *p*=0.256  *ηp2*=0.036 | ***F*(1,36)=4.25**  ***p*=0.047**  ***ηp2*=0.106** |
| P2 | *F*(1,36)=1.03  *p*=0.317  *ηp2*=0.028 | *F*(1,36)=4.26  *p*=0.046  *ηp2*=0.106 | *F*(1,36)=4.29  *p*=0.046  *ηp2*=0.106 | *F*(1,36)=0.80  *p*=0.377  *ηp2*=0.022 | *F*(1,36)=0  *p*=0.993  *ηp2*=0.00 | *F*(1,36)=6.61  *p*=0.014  *ηp2*=0.155 | ***F*(1,36)=5.03**  ***p*=0.031**  ***ηp2*=0.123** |



**Fig 3.** EEG data under the semantic retrieval condition. A. the average ERP waveforms at electrodes F7, FT7 and T7 of the two conditions pre- and post-training. B. the topographic maps of the difference between the two conditions post- and pre training (post - pre).



**Fig 4.** EEG data under the non-semantic retrieval condition. A. the average ERP waveforms at electrodes F7, FT7 and T7 of the two conditions pre- and post-training. B. the topographic maps of the difference between the two conditions post- and pre training (post - pre).

### ***Switching.*** A repeated-measures ANOVA on the Switching Cost for both the conditions of subjects showeda significant main effect of session, no effect of training group, and a significant interaction between session and training group. Further simple effects analysis revealed that the cost of conversion in the training group was significantly lower in the post-training test (*M*=0.16, *SD*=0.11) than in the pre-training test (*M*=0.24, *SD*=0.17), *F*(1,52) = 14.58, *p* < 0.001, *ηp2* = 0.219, 95% CI = [0.04, 0.12]. However, there was no significant difference in the cost of conversion between the post-training test (*M*=0.21, *SD*=0.11) and the pre-training test (*M*=0.23, *SD*=0.13) for the control group,*F*(1,52) = 1.38, *p* = 0.246, *ηp2* = 0.026, 95% CI = [-0.02, 0.06].

A repeated-measures ANOVA on the Mixed Cost also showed a significant main effect of session, no effect of training group, and a significant interaction between session and training group. Further simple effects analysis revealed that the cost of conversion in the training group was significantly lower in the post-training test (*M*=98.42, *SD*=79.75) than in the pre-training test (*M*=156.72, *SD*=112.84), *F*(1,52) = 25.03, *p* < 0.001, *ηp2* = 0.325, 95% CI = [34.92, 81.69]. However, there was no significant difference in the cost of conversion between the post-training test (*M*=136.82, *SD*=96.09) and the pre-training test (*M*=158.84, *SD*=94.19) for the control group, *F*(1,52) = 3.85, *p* = 0.055, *ηp2* = 0.069, 95% CI = [-0.50, 44.56].

### ***Fluid intelligence.*** A repeated-measures ANOVA did not reveal the presence of any significant main or interaction effects.

### ***Reading Fluency task.***A repeated-measures ANOVA on the sentence count metric for the rapid reading task revealed the main effect of session and interaction between session and training group were significant, but the main effect of training group was not. Further simple effects analysis revealed that the two groups read significantly more sentences in the post-training test (Training group: *M*=81.92, *SD*=7.25; Control group: *M*=77.25, *SD*=9.18) than in the pre-training test (Training group: *M*=72.35, *SD*=8.29; Control group: *M*=74.43, *SD*=9.73). However, in the post-training session, the number of sentences in the training group is significantly more than that in the control group, *F*(1,52) = 4.26, *p* = 0.044, *ηp2* = 0.076, 95% CI = [0.13, 9.22].

### A repeated-measures ANOVA on the Reaction time showed a significant main effect of session but no main effect of training group and no interaction between training group and session. A test-session main effect analysis revealed that participants were significantly faster in the post-training test (*M*=2193.09, *SD*=438.95) than in the pre-training test (*M*=2348.40, *SD*=425.38), *F*(1,52) = 9.12, *p* =0.004, *ηp2* = 0.149, 95% CI = [53.20, 264.00]

# **Discussion**

The key aims of this study were to examine the plasticity of semantic LTMR following 12 days of training and ascertain the extent to which any benefits transfer to near and far domains. Behavioural assessments of semantic LTMR, working memory, reading fluency, and fluid intelligence were administered at pre-test and post-test. In addition, ERPs were recorded during the semantic LTMR task, to index neural changes in automatic retrieval processes. The key findings were that LTMR improved over training and transferred to other measures of semantic LTMR. This “near transfer” effect was apparent both behaviourally and electrophysiologically. There was also evidence of “far transfer” to measures of switching, updating and reading fluency (but not to interference control or fluid intelligence). These findings will be discussed in detail below, and in relation to theories of semantic LTMR.

Training on the Posner task required individuals to quickly retrieve the desired target information from semantic long-term memory to make judgements about whether number pairs were presented in the same format (i.e., Chinese/numerical) and whether the numbers matched. For the adaptive intervention group there were clear improvements in retrieval time over the 12 days of the study. Specifically, it was found that the speed by which participants made accurate responses increased each day, from Level 4.3 at the beginning of training to Level 8.3 at the end of training. Therefore, individuals could retrieve the required target information from the semantic long-term memory system more efficiently after (than before) training without a cost to correctness (consistent with Ma et al., 2020).

Participants who completed the adaptive LTMR training also showed significant gains in the harder levels of the Posner task that were used in the pre- and post-tests, in contrast to participants who received the non-adaptive control training. The parity judgements that formed the Posner test task required similar cognitive processes as in training (i.e., rapid and accurate retrieval of semantic information in response to the target information), but went beyond simply retrieving physical properties or specific meanings to incorporate a semantic decision regarding the relationship between the stimulus pair. Specifically, participants were required to decide whether the digits were both odd/even or not, when the pair were both presented in the same format (i.e., numerical or Chinese) and when the pair were presented in differing formats (i.e., numerical and Chinese). These results suggest that the gains in semantic LTMR elicited “near transfer” to different semantic LTMR tasks that required a more elaborate form of retrieval. Indeed, whilst Levels 1 and 2 demanded a simple level of extraction (extracting the number of physical attributes and representations), levels 3 and 4 required participants to perform complex extractions based on their understanding of parity.

Similarly, near transfer effects were also observed for the EEG SORT, a well-established measure of semantic LTMR (e.g., Fratantoni et al., 2017), thus providing the direct and further electrophysiological support for the plasticity of semantic LTMR. Consistent with the hypotheses, we observed a significant change in mean amplitude within the 700–1000 ms time window from pre to post training for the semantic but not for non- semantic retrieval conditions in training group, which may reflect the changes in the semantic retrieval process (Assaf et al., 2006a; Fratantoni et al., 2017). Additionally, we also find that in the post-training test, the mean amplitude in the 300-500ms time window was significantly reduced for the training condition than at the pre-test. Conversely, there was no significant difference between the pre-training and post-training tests for the control group. One potential explanation for semantic retrieval trials is that the change in mean wave amplitude within this early time window may be signaling a change in the N400 ERP component, which has been claimed to provide an index of the ease with which semantic information is automatically retrieved, with larger N400 amplitudes associated with a greater resource cost during semantic access (Kutas & Federmeier, 2011; Voss & Federmeier, 2011). In terms of non-semantic retrieval trials, the explanation maybe that individuals become more sensitive to word pairs with obvious semantic surprise or conflict. This aligns with our interpretation of the Posner test data, further supporting the claim that participants’ semantic retrieval ability was becoming more efficient following adaptive training.

It is also important to note that we found a significant effect of session on the amplitude of the P2 component for the training group, but not for the control group. This suggests that participants in the training condition exhibited improved attentional control in the post-test compared to the pre-test. This result is perhaps not surprising. The SORT, as a cognitive task, requires individuals to focus their attention throughout the task to complete it. However, we did not observe any improvement in the flanker or stroop tasks, which may be due to P2 reflecting more on the recruitment of attention resources rather than control in the face of conflicting environments (i.e. reaction inhibition, Campbell & Sharma, 2013; Geisler & Murphy, 2000; Sugimoto & Katayama, 2013).

One point that needs to be emphasized about the ERP results is that although the topographic maps indicated that there may have been training-induced effects at other scalp locations, given the main purpose of this study and the clear hypotheses we have made by combing through the literature. The analysis and discussion of other electrode points is beyond the scope of this paper, and can be fully analyzed in future related studies, which may deepen the understanding of semantic long term memory plasticity.

There was also evidence of far transfer for the adaptive training group (in comparison to the control group), namely for reading fluency and switching. Turning first to the benefits of semantic LTMR for reading fluency, we found that after adaptive (but no control) training, individuals responded more quickly and accurately to the meaning of the presented utterance in the reading task, and they read more sentences within the same time constraints as compared to their pre-training performance. Moreover, the average response time for a single sentence was reduced. These findings can be explained if we assume that the cognitive processing involved in the LTMR training task shares common processing mechanisms with the rapid reading task. Reading involves bottom-up processes (i.e., the literal apprehension of the text-based information) and top-down processing, involving the use of prior knowledge to predict and construct the information and ultimately derive meaning. Thus, it is plausible that speeding up processes of semantic retrieval could work to improve both bottom-up and top-down mechanisms during reading. Previous studies have also claimed that reading requires efficient access to existing long-term memory content in the service of successful comprehension (Just & Carpenter, 1992; Ericsson et al., 1995). Despite this promising transfer effect to reading fluency however, it is important to caveat our interpretation with the limited nature of the reading measure used here. Reading is a multi-component skill, involving fluent and accurate word and sentence production but also comprehension (Nation & Snowling, 1998). Whilst the improvements in reading fluency observed are notable and provide important evidence of far-transfer effects following LTMR training, we do not know whether the reading fluency benefits are transient or long-term, nor do we know whether they are pedagogically meaningful. Furthermore, based on the present results alone we cannot claim that LTMR may be a promising form of reading intervention, since it is likely that well-validated evidence-based reading interventions that tackle reading directly are likely to be more effective than indirect training in semantic LTMR.

In addition to observing far transfer effects on the reading fluency task, we also observed improvements from pre- to post-training for the Switching task (both switching cost and mixed cost) and updating task. For switching task, participants had to retrieve the corresponding task rule given before the start of the task based on different stimulus information, and make a response after retrieving stimulus-related properties (e.g., parity properties) from long-term memory. This finding thus suggests that the cost of switching at least partly reflects the time required to retrieve the appropriate task rule from long-term memory (aligning with previous studies e.g., Mayr et al., 2000). For updating task, subjects need to maintain a constant focus of attention (similar to active control) and to quickly retrieve the target information from memory system efficiently. The transfer effect to the updating task is understandable, if the P2 component represents recruitment and retention of attentional resources and changes in 700-1000ms EEG amplitude characterise the ability to retrieve semantic memory. However, it is important to note that the results of the updating task in this study are not consistent with Ma et al. (2020). Given the similarities between the training task material and the updating task, the influence of the practice effect still cannot be completely dismissed.

In the above we discussed the possibility that increased efficiency of semantic LTMR may have led to a wide range of transfer effects, but we must also acknowledge the possibility that the observed effects may not be exclusively due to increased efficiency of LTMR. The LTMR training task used in the present study is a complex cognitive task, other cognitive abilities (such as working memory, executive function, and processing speed) may inevitably be involved to some extent. One of the most obvious debates may be whether the training effect is due to changes in processing speed, as the training task clearly forced an acceleration of processing speed by altering the presentation time of stimuli. However, considering the transfer effects of the training effect observed in other tasks, it is unlikely that this is solely attributable to an improvement in processing speed. Additionally, as we highlighted that the topographic maps indicated that there may have been training-induced effects at other scalp locations, which also may be related to other potential training effects. Thus, we suggest that the observed transfer effects, although primarily related to the increased efficiency of semantic LTMR, may also be influenced by other underlying cognitive abilities, which we are willing to refer to as an “additive effect”. Clearly, future research should aim to use a more refined design to distinguish between them.

Conversely there were no far-transfer effects for interference control (measured by flanker and stroop task) or fluid intelligence. The result of fluid intelligence is consistent with previous studies’ results (Ma et al., 2020), but the result of interference control is not consistent. A potential reason is the The dependent measure of flanker and stroop task the present study used is different from Ma et al. (2020). By computing the proportional cost, the mixed effect of processing speed can be well avoided (Verhaeghen, 2011 & 2014). Another explanation is that stroop and flanker rely more on response inhibition ability, which is a form of passive attentional control, whereas our training may have improved active control (the P2 component of the training group changed before and after training). However, given that semantic LTMR does correlate highly with fluid intelligence and WM (Wang et al., 2017; Nash et al., 2019), further research is required to better understand why transfer did not occur to these domains. One possible explanation is that perhaps the training is speeding up LTMR, and fast LTMR is not required for fluid intelligence at least. Meanwhile perhaps speed isn't so important for some specific WM task.

It is important to acknowledge a number of limitations of the present study. First, the present study only conducted an immediate post-test at the end of the 12 day training period and did not monitor longer term maintenance of any of the above reported improvements. As such, it is unclear whether the improvements in semantic LTMR efficiency and its transfer effects on other cognitive abilities persist in the longer term. Second, a strength of the study is the use of a non-adaptive control condition. However, the control task in the present study might have been more boring or less motivating (The control group's a significant drop in mean ACC in session 9 seems to indicate this), and less likely to lead to improvements in response time. Future research could better localise the mechanisms of improvement in LTMR and transfer through the use of different control conditions. For example, using the control task whereby the procedure is exactly the same as the training task, including different difficulty level, with only materials and task requirements set to tasks that do not involve semantic LTMR. Finally, the current study’s population was dominated by healthy, college-aged adult participants and although the study was adequately powered the sample size was small and not sufficient for examining individual differences in the response to intervention. Thus, future studies could broaden the study population’s age range, sample size and background demographics to more fully examine the plasticity of semantic LTMR and the generalizability of the effects in a broader population.

# **Conclusion**

The present study used an adaptive training procedure with the aim of enhancing semantic LTMR. Behavioural and electrophysiological measures clearly demonstrated the plasticity of semantic LTMR over a 12 day training period. Furthermore, improvements in semantic LTMR transferred to tasks that were both closely and more distantly related to the trained task, including switching, updating, and reading fluency. These data therefore provide novel empirical support for the value of training semantic LTMR, and suggest that such training effects can generalize to distal domains. Future research is required to ascertain the longer-term value of LTMR training, and the extent to which it can have societal value across a broader population.

**References**

Anderson, N. D., Ebert, P. L., Grady, C. L., & Jennings, J. M. (2018). Repetition lag training eliminates age-related recollection deficits (and gains are maintained after three months) but does not transfer: Implications for the fractionation of recollection. *Psychology & Aging*, *33*(1), 93-108. https://doi.org/10.1037/pag0000214

Assaf, M., Rivkin, P. R., Kuzu, C. H., Calhoun, V. D., Kraut, M. A., Groth, K. M., Yassa, M. A., Jr, J. H., & Pearlson, G. D. (2006). Abnormal Object Recall and Anterior Cingulate Overactivation Correlate with Formal Thought Disorder in Schizophrenia. *Biological Psychiatry*, *59*(5), 452-459. https://doi.org/10.1016/j.biopsych.2005.07.039

Brier, M. R., Maguire, M. J., Tillman, G. D., Hart, J., & Kraut, M. A. (2008). Event-related potentials in semantic memory retrieval. *Journal of the International Neuropsychological Society*, *14*(5), 815-822. <https://doi.org/10.1017/S135561770808096X>

Campbell, J., & Sharma, A. (2013). Compensatory changes in cortical resource allocation in adults with hearing loss. *Frontiers in systems neuroscience*, *7*, 71. <https://doi.org/10.3389/fnsys.2013.00071>

Chiang, H. S., Mudar, R. A., Pudhiyidath, A., Spence, J. S., & Hart, J. (2015). Altered Neural Activity during Semantic Object Memory Retrieval in Amnestic Mild Cognitive Impairment as Measured by Event-Related Potentials. *Journal of Alzheimers Disease Jad*, *46*(3), 703-717. https://doi.org/10.3233/JAD-142781

Cowan, N. (2010). The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, *19*(1), 51-57. https://doi.org/10.1177/0963721409359277

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Norris D. (2017). Short-term memory and long-term memory are still different.*Psychological bulletin*, *143*(9), 992–1009. https://doi.org/10.1037/bul0000108

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211-245. https://doi.org/10.1037/0033-295x.102.2.211

Fernández, G., Laubrock, J., Mandolesi, P., Colombo, O., & Agamennoni, O. (2014). Registering eye movements during reading in Alzheimer's disease: Difficulties in predicting upcoming words. *Journal of Clinical & Experimental Neuropsychology*, *36*(3), 302-316. https://doi.org/10.1080/13803395.2014.892060

Finn, M., & McDonald, S. (2015). Repetition-lag training to improve recollection memory in older people with amnestic mild cognitive impairment. A randomized controlled trial. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition*, *22*(2), 244–258. https://doi.org/10.1080/13825585.2014.915918

Fratantoni, J. M., Delarosa, B. L., Didehbani, N., Hart, J., & Kraut, M. A. (2017). Electrophysiological Correlates of Word Retrieval in Traumatic Brain Injury. *Journal of Neurotrauma*, *34*(5), 1017-1021. <https://doi.org/10.1089/neu.2016.4651>

Geisler, M. W., & Murphy, C. (2000). Event-related brain potentials to attended and ignored olfactory and trigeminal stimuli.*International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, *37*(3), 309–315. <https://doi.org/10.1016/s0167-8760(00)00111-2>

Herron, J. E., & Wilding, E. L. (2005). An electrophysiological investigation of factors facilitating strategic recollection. *Journal of Cognitive Neuroscience*, *17*(5), 777-787. https://doi.org/10.1162/0898929053747649

Jacoby, L. L. . (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory & Language*, *30*(5), 513-541. https://doi.org/ 10.1016/0749-596X(91)90025-F

Jennings, J. M., Webster, L. M., Kleykamp, B. A., & Dagenbach, D. (2005). Recollection Training and Transfer Effects in Older Adults: Successful Use of a Repetition-Lag Procedure. *Aging Neuropsychology & Cognition*, *12*(3), 278-298. https://doi.org/10.1080/138255890968312

Jolles, D. D., Grol, M. J., Buchem, M., Rombouts, S., & Crone, E. A. (2010). Practice effects in the brain: Changes in cerebral activation after working memory practice depend on task demands. *Neuroimage*, *52*(2), 658-668. https://doi.org/10.1016/j.neuroimage.2010.04.028

Juola, J. F., Fischler, I., Wood, C. T., & Atkinson, R. C. (1971). Recognition time for information stored in long-term memory. *Perception & Psychophysics*, *10*(1), 8-14. https://10.3758/BF03205757

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, *99*(1), 122–149. https://doi.org/10.1037/0033-295x.99.1.122

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Lei，L ., Pan, J., Liu, H., Chang，C.，, M.-., Li, H. Z., & Y., e. a. (2011). Developmental trajectories of reading development and impairment from ages 3 to 8 years in Chinese children. *Journal of Child Psychology & Psychiatry*, *52*(2), 212-220. https://doi.org/10.1111/j.1469-7610.2010.02311.x

Ma, X., Zhang, H., Zhao, X., & Zhou, A. (2020). Training and transfer effects of long-term memory retrieval training. *Journal of Cognitive Psychology*(1), 1-12. https://doi.org/10.1080/20445911.2020.1814306

Mayr, Ulrich, Kliegl, & Reinhold. (2000). Task-Set Switching and Long-Term Memory Retrieval. *Journal of Experimental Psychology Learning Memory & Cognition, 26*(5), 1124-1140*.* https://doi.org/10.1037//0278-7393.26.5.1124

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229-240. https://doi.org/10.1111/j.1469-8986.2010.01061.x

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

Nation, K. , & Snowling, M. J. . (1998). Semantic processing and the development of word-recognition skills: evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, *39*(1), 85-101. https://doi.org/10.1006/jmla.1998.2564

Nash, & Unsworth. (2019). Individual differences in long-term memory. *Psychological Bulletin*, *145*(1), 79-139. https://doi.org/10.1037/bul0000176

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220-246. https://doi.org/10.1037/0033-2909.126.2.220

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology Learning Memory & Cognition*, *27*(2), 384-413. https://doi.org/10.1037/0278-7393.27.2.384

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*(1), 75-79. https://doi.org/10.1038/nn1165

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. Psychological review, 74(5), 392–409. https://doi.org/10.1037/h0024913

Raven, J. (1983). *The Progressive Matrices and Mill Hill Vocabulary Scale in Western Societies*. https://doi.org/10.1007/978-1-4899-2151-2\_8

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, *76*(2), 179-193. https://doi.org/10.1037/h0027277

Stamenova, V., Jennings, J. M., Cook, S. P., Walker, L. A., Smith, A. M., & Davidson, P. S. (2014). Training Recollection in Healthy Older Adults: Clear Improvements on the Training Task, but Little Evidence of Transfer. *Frontiers in human neuroscience*, *8*, 898. <https://doi.org/10.3389/fnhum.2014.00898>

Sugimoto, F., & Katayama, J. (2013). Somatosensory P2 reflects resource allocation in a game task: assessment with an irrelevant probe technique using electrical probe stimuli to shoulders. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, *87*(2), 200–204. https://doi.org/10.1016/j.ijpsycho.2013.01.007

Tulving , E. ( 1972 ). Episodic and semantic memory . In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381 –403). New York : Academic Press .

Unsworth, N., & Engle, R. W. (2007). The nature of individual difffferences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114(1)*, 104–132. https://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: the role of controlled search. *Memory & Cognition*, *41*(2), 242-254. https://doi.org/10.3758/s13421-012-0261-x

Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science*, *20*(3), 174–180. https://doi.org/10.1177/0963721411408772

Verhaeghen, P. (2014). The elements of cognitive aging: Meta-analyses of age-related differences in processing speed and their consequences. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195368697.001.0001

Voss, J. L., & Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology*, *48*(4), 532-546. https://doi.org/10.1111/j.1469-8986.2010.01085.x

Wang, T., Ren, X., & Schweizer, K. (2017). Learning and retrieval processes predict fluid intelligence over and above working memory. *Intelligence*, *61*, 29-36. https://doi.org/10.1016/j.intell.2016.12.005

Westerberg, H., & Klingberg, T. (2007). Changes in cortical activity after training of working memory--a single-subject analysis. *Physiology & Behavior*, *92*(1-2), 186-192. https://doi.org/10.1016/j.physbeh.2007.05.041

Wieser, S., & Wieser, H. G. (2003). Event-related brain potentials in memory: correlates of episodic, semantic and implicit memory. *Clinical Neurophysiology*, *114*(6), 1144-1152. https://doi.org/10.1016/s1388-2457(03)00044-0

Zhao, X., Dang, C., & Maes, J. H. R. (2020). Effects of working memory training on EEG, cognitive performance, and self-report indices potentially relevant for social anxiety. *Biological Psychology*, *150*, 107840. https://doi.org/10.1016/j.biopsycho.2019.107840

Zhao, X., & Jia, L. (2019). Training and transfer effects of interference control training in children and young adults. *Psychological research*, *83*(7), 1519–1530. https://doi.org/10.1007/s00426-018-1007-6