
Peer reviewed version

License (if available): CC BY

Link to published version (if available): 10.1016/j.jecp.2023.105836

Link to publication record in Explore Bristol Research

PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Elsevier at https://doi.org/10.1016/j.jecp.2023.105836. Please refer to any applicable terms of use of the publisher.

**University of Bristol - Explore Bristol Research**

**General rights**

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/
Age-related differences in the removal of information from working memory

Siyuan Mei a*, Yang Gao a*, Wanna He a, Christopher Jarrold b, Tengfei Wang a

a Department of Psychology and Behavior Sciences, Zhejiang University

b School of Psychological Science, University of Bristol

* These authors contributed equally as first authors.

Author Note

This research was supported by National Natural Science Foundation of China (No. 32000761), Science and Technology Innovation 2030-“Brain Science and Brain-like Research” Major Project (2022ZD0210800), Ministry of education of Humanities and Social Science project (No. 20YJC880095), and Fundamental Research Funds for the Central Universities (2020QNA3005).

Correspondence concerning this article should be addressed to Tengfei Wang, Department of Psychology and Behavioral Sciences, Zhejiang University, Yuhangtang Road 866, 310058 Hangzhou, China. E-mail: tfwang@zju.edu.cn.

Data is available from the corresponding author upon reasonable request.
Abstract

Removal has been assumed to be a core mechanism in working memory. However, it remains unclear whether children can actively remove outdated information from working memory and how this ability develops as children age. The current study aims to examine age-related differences in removal ability and its relations with cognitive control and working memory capacity. Children aged 7, 9, and 11 years performed a modified working memory updating task assessing removal efficiency. In addition, a battery of cognitive control and working memory capacity tasks were administered. Results indicated that updating response times reduced considerably when a longer time was given for removal suggesting that children aged 7 to 11 years can actively remove outdated items from working memory prior to encoding the new ones, and that removal efficiency increased with age. More importantly, age-related increases in removal efficiency occurred concurrently with the development of working memory capacity. Proactive control predicted removal efficiency over and beyond age and working memory capacity. The findings shed new light on the mechanisms underlying the development of working memory updating.

Keywords: removal, working memory updating, cognitive control, age-related difference.
Age-related differences in the removal of information from working memory

Introduction

Working memory (WM) is a cognitive system responsible for the simultaneous maintenance and manipulation of information (Baddeley, 1992). It plays crucial roles in higher cognitive functions, such as reasoning (Unsworth et al., 2014), reading comprehension (Peng et al., 2018), and multi-tasking (Redick et al., 2016). A critical feature of WM is that it has a quite limited capacity (Cowan, 2001). As a result, WM requires a mechanism that removes irrelevant items effectively and replaces them with goal-relevant ones to keep its contents updated (Oberauer et al., 2012). Therefore, removal has been assumed as a core mechanism in WM (Lewis-Peacock et al., 2018; Shipstead et al., 2016). Removal refers to the process of excluding irrelevant or outdated information from WM in service of the current goal (Lewis-Peacock et al., 2018). Although this removal process has been examined in adults (e.g., Dames & Oberauer, 2022; Ecker, Lewandowsky, et al., 2014; Oberauer, 2018; Singh et al., 2018), it remains unclear whether children can actively remove outdated information from WM and how this ability develops in children. Therefore, the current study set out to investigate the development of removal across different age groups by using a well-established paradigm that enables the isolation of the removal process (Ecker, Lewandowsky, et al., 2014; Ecker, Oberauer, et al., 2014). In addition, the potential mechanisms underlying the development of removal were also explored.

Removal as an important component of WM updating

WM updating is not a unitary process. In a seminal study, Ecker et al. (2010) identified three subprocesses underlying WM updating, including retrieval, transformation, and substitution. In order to replace no-longer relevant information in WM, this information has to be firstly retrieved, and then, if necessary, transformed according to the task requirements. Finally, outdated information is substituted by new information. All three subprocesses make independent contributions to updating performance. However, only substitution appears to be a distinct process of WM updating since it alone showed a weak correlation with WM capacity (WMC) while retrieval and transformation were the processes common to
both WM updating and WMC (Ecker et al., 2010).

In a subsequent study, substitution was further decomposed into two processes, i.e., the removal of outdated information and the encoding of updated information (Ecker, Lewandowsky, et al., 2014). In order to separate removal from encoding processes, Ecker, Lewandowsky, et al. (2014) developed a new WM updating task. Figure 1 presents a schematic illustration of this task. The task started with the presentation of three unrelated letters in three individual squares which had to be memorized. Then, the letters disappeared and one of the squares was cued, indicating that the letter in this square would be replaced by a new one (i.e., a target). After the target appeared, participants were instructed to press a key as soon as they completed letter updating, which provided a measure of updating response time. This updating procedure repeated several times in each trial before a final recall of the most recent letters. It is assumed that there are two processes underlying the substitution of an old item with a new one: the removal process which starts as soon as the cue is presented, and the encoding process which starts from the appearance of the new letter. In order to distinguish these two processes, the cue-target interval (CTI) was manipulated to be either short (200 ms) or long (1500 ms). If the removal process occurs as soon as the cue is presented, participants in the long-CTI condition would update the new item faster since they have already removed the to-be-updated letter prior to the display of the target, and only need to encode the new letter when it is presented. The short-CTI condition, however, only allows participants to focus their attention on the cued frame without being able to initiate the removal process. Therefore, they have to complete both the removal and encoding processes when the target is presented, resulting in a longer updating response time. Consistent with this hypothesis, they found that adults’ mean response time for the short CTI was substantially longer than that for the long CTI, clearly demonstrating the presence of active removal during WM updating (Ecker, Oberauer, et al., 2014; Oberauer, 2018).

In addition to the experimental evidence, based on Ecker and colleagues’ WM updating task individual differences studies have identified a reliable latent variable of removal, which has been shown to be significantly related to adults’ WMC and fluid intelligence (Singh et al., 2018), but not related to more traditional inhibition tasks (Rey-Mermet et al., 2020). Other evidence also suggests the existence of
removal in WM (Lewis-Peacock et al., 2018): (1) removing some contents from WM reduces its load, leading to faster and more accurate access to the remaining relevant information (Ecker, Lewandowsky, et al., 2014; Souza et al., 2014); (2) The removed contents in WM become less accessible (e.g., Williams et al., 2012; Dames & Oberauer, 2022); (3) The removed contents in WM become neurally silent (e.g., Sprague et al., 2016).

**Age-related differences in WM updating and removal**

Though the accumulated evidence clearly suggests that adults can actively remove irrelevant information prior to encoding new information during WM updating (Ecker, Lewandowsky, et al., 2014; Ecker, Oberauer, et al., 2014), it remains unclear whether children engage in a similar process to adults and how this ability develops as children get older. The limited research on the development of updating suggests that children are able to replace old items with new ones and continuously keep track of the most recent items in updating tasks. That updating ability develops from early childhood and continues to improve until adolescence is shown in different studies using different paradigms (e.g., Carriedo et al., 2016; Gathercole et al., 2004; Huizinga et al., 2006; Lee et al., 2013; Lendínez et al., 2015; Pelegrina et al., 2015; Schleepen & Jonkman, 2009). For example, using the n-back task, Pelegrina et al. (2015) examined the development of updating in a large sample of children and adolescents aged 7-13 years. They found age-related increases in performance for different levels of difficulty including 1-back, 2-back, and 3-back. Another task thought to require participants to simultaneously remove outdated items and encode new ones is the running memory task, in which participants report the most recent $n$ items in a list of unpredictable length (Morris & Jones, 1990, though see Palladino & Jarrold, 2008). Lee et al. (2013) used this task to investigate the development of WM updating from kindergarten to sixth grade. They showed a linear increase in WM updating performance with age. Similar research was conducted by Huizinga et al. (2006), revealing that running memory performance continues to develop from 7 to 11 years old.

In addition to behavioral studies, there is also evidence on the neural development of updating (e.g., Myatchin & Lagae, 2013; Pelegrina et al., 2020). For example, Pelegrina et al. (2020) drew on the P300
to assess age-related differences in WM updating, using 0-back and 1-back tasks. They found that d-prime scores for the 1-back task significantly increased as age increased from 6 to 20 years old. The amplitude of the P300 in both tasks decreased with age, which was assumed to reflect maturation of the updating process. In addition, the amplitude difference (0-back vs 1-back) of the late P300 was larger for 6-8 year old children than for older participants, suggesting that the 1-back task exerts larger demands (versus 0-back) on the 6-8 year old children than the older participants. **Despite these studies showing that updating develops with age, it remains unclear how the removal process develops during childhood.**

Although WM updating requires the removal of information, performance on these updating tasks in fact depends on the functioning of several cognitive processes such as encoding, maintenance, and retrieval (Ecker et al., 2010), and separating the effect of removal from that of the others is not straightforward. In a more recent study, Linares et al. (2016) adapted Ecker et al. (2010)’s WM updating task to examine age-related differences in the subprocesses of updating (i.e., retrieval, transformation, and substitution) among children aged 8, 11, 14, and 21 years. Although an increase in accuracy and a decrease in reaction times across all tasks was observed with age, only the retrieval component interacted with age to influence task performance, whereas transformation and substitution did not. This result seems to suggest that the substitution process does not undergo any changes within the age ranges that Linares et al. (2016) considered. However, a closer scrutiny of the data indicates that the substitution component actually developed between 8 and 11 years old, but did not improve substantially in older individuals. Because only substitution involves the removal of outdated information from WM (Ecker, Lawandowsky, et al., 2014), this finding suggests that removal may undergo considerable development between the ages of 8 and 11. However, since the task used in Linares et al. (2016) did not separate the removal process from other potentially confounding processes (i.e., encoding), it cannot unequivocally reveal the development of removal. In order to further elucidate whether children can actively remove information from WM and how this process develops during this age range, one needs to use a paradigm developed especially for this purpose (e.g., Ecker, Lewandowsky, et al., 2014).

**The relationship between removal and cognitive control**
Removal has been assumed to be an active process that operates in the service of current goals (Lewis-Peacock et al., 2018). That is, as individuals are aware of the incoming new information, they prepare themselves in advance to engage in the removal of old information. This kind of proactive cognitive operation has been well-documented within the dual mechanisms of control framework (Braver, 2012; Braver et al., 2007). According to this framework, there are two qualitatively distinct types of cognitive control: proactive control and reactive control. Under proactive control, goal-relevant representations are activated before their implementation, and actively maintained during the preparation period until execution. Conversely, under reactive control, goal-relevant representations are only triggered when they are required. Proactive control is generally more effective than reactive control in carrying out task goals (Braver, 2012). The task most frequently used to assess the two types of cognitive control is the AX-Continuous Performance Task (AX-CPT, Braver, 2012). On each trial of the AX-CPT, participants first encounter a cue letter and then a probe letter. They must respond to the probe only if the cue is A and the probe is X (AX trials). Due to the large proportion of AX trials, participants using proactive or reactive control exhibit different patterns of response in the other cue-probe conditions. If the cue is an A, but the probe is not an X (AY trials), participants using proactive control tend to prepare an erroneous target response when A appears, which leads to worse performance compared to participants using reactive control. On the contrary, when the cue is not an A, but the probe is an X (BX trials), participants using reactive control are prone to make an erroneous target response when they see an X probe, resulting in worse performance compared to participants using proactive control.

Proactive control may be essential for implementing the removal of outdated information before encountering new information. Specifically, in the WM updating task (Ecker, Lewandowsky, et al., 2014), participants are asked to continuously update three letters in WM. When the cue indicates which letter will be replaced, participants using proactive control are more likely to remove the to-be-updated letter beforehand as soon as the cue is presented. By contrast, participants using reactive control may be more prone to initiate the removal process only when the target letter appears. Therefore, it is expected that removal efficiency should be associated with proactive control.
Cognitive control undergoes substantial increase as children develop (e.g., Chatham et al., 2009; Chevalier et al., 2015; Gonthier et al., 2019; Lorsbach & Reimer, 2010; Munakata et al., 2012). For example, Gonthier et al. (2019) indicated that there was a progressive shift from reactive control to proactive control around 6 years of age. Chatham et al. (2009) examined the difference in cognitive control between 3.5- and 8-year-old children using the AX-CPT. Their results revealed an obvious age-related increase in the use of proactive control. Given the hypothesized relationship between proactive control and removal efficiency, we expected that the development of cognitive control would be linked to the development of removal efficiency.

**The present study**

The goal of the current study was to examine age-related differences in removal efficiency among school-aged children and to explore the link between age-related changes in removal efficiency and the development of cognitive control and WMC. We focused on school-aged children because previous studies have indicated that WM updating develops substantially in this age range (e.g., Lee et al., 2013; Linares et al., 2016; Pelegrina et al., 2015). In particular, the results of Linares et al. (2016) suggest that substitution undergoes substantial development between 8 and 11 years old.

Ecker, Lewandowsky, et al. (2014)’s paradigm was adapted to make it appropriate for assessing children’s removal process. Instead of using letters, we used single-digit numbers familiar to primary school-aged children. The CTI was manipulated across three conditions: 200 ms, 1500 ms, and 2800 ms. Previous studies suggest that 1500 ms is sufficient for adults to remove irrelevant information, and 200 ms is enough to focus attention on the to-be-updated frame without permitting removal (Ecker, Oberauer, et al., 2014). We included an additional 2800 ms condition since children might need more time to remove irrelevant information. Between CTI-condition comparisons would reveal whether children could engage in the removal process and its time course. Possible interactions of CTI with age would imply age-related changes in the removal process. However, the evaluation of removal efficiency by comparing response times of the long- and short-CTI conditions might be confounded by differences in general processing speed (Rey-Mermet et al., 2019) across different baseline levels of performance in different
age groups. Such scaling differences can lead to erroneous conclusions in development studies (e.g., the development of rehearsal, Jarrold & Citroën, 2013). Therefore, removal efficiency is best assessed by the proportional-gain score (computed as the updating RT difference between the short and long CTI conditions divided by the updating RT of the short-CTI condition) rather than the absolute difference score between two CTI conditions (Ecker, Lewandowsky, et al., 2014; Singh et al., 2018). We expected that removal efficiency would increase with age.

Furthermore, since removal has been assumed to be an active process (Lewis-Peacock et al., 2018), it is plausible that the use of proactive control would be associated with more efficient engagement of removal processes. Thus, we would expect removal efficiency to be positively related to proactive control. Finally, due to the importance of removal for keeping relevant information in WM, we would expect that removal efficiency is associated with WMC. Two previous studies have examined the relationship between removal efficiency and WMC in adults, but with contradictory conclusions. Ecker, Lewandowsky, et al. (2014) found a negligible link between removal efficiency and WMC. In contrast, a more recent study using a larger sample and several measures of removal revealed a significant relationship between removal efficiency and WMC (Singh et al., 2018). This question therefore warrants further examination, and we expected a positive relationship between WMC and removal efficiency among children.

Method

Participants

Since there are no directly relevant prior data available for estimating the expected effect size of age on removal, data collection was planned based on previous studies using a similar design (e.g., Linares et al., 2016; Gonthier et al., 2019), which included around 24 participants per age group. In addition, a priori power analysis was conducted using G*Power 3.1.9.7 (Faul et al., 2007) for sample size estimation. Based on Linares et al. (2016), which compared substitution accuracy of children aged 8 and 11 years, the effect size was estimated as $f = 0.43$. With this effect size, the minimum sample size required is $N = 54$ for a three-group ANOVA test with a significance criterion of $\alpha = .05$ and power = .80.
In our study, a total of 72 children were recruited from three school grades in an elementary school at [blinded]: 7-year age group \((n = 24, \text{mean age} = 7.03 \text{ years}, SD = 0.34)\), 9-year age group \((n = 24, \text{mean age} = 8.82 \text{ years}, SD = 0.32)\), and 11-year age group \((n = 24, \text{mean age} = 10.82 \text{ years}, SD = 0.31)\). One participant from the 7-year age group and one from the 9-year age group failed to complete the WM updating task. In addition, participants whose updating accuracy was below 85% in the WM updating task were excluded from data analyses. The 85% threshold was determined after data collection based on the overall distribution of correctness among all participants. It is worth noting that 85% represents a fairly high standard, and only four participants (two participants from the 7-year age group and two from the 9-year age group) fell below this criterion. These participants were excluded, leaving a final sample of 66 participants.\(^1\)

Written informed consent describing the purpose and procedure of the study was obtained from the children’s parents. Children received a small gift after completing the experiment. Ethical approval for this study was obtained from the appropriate institutional review board – [blinded] University (approval number: 2021028; title: “The development of children’s removal of information from working memory”).

**Materials**

**Removal: the WM updating task**

The WM updating task developed by Ecker, Lewandowsky, et al. (2014) was adapted to measure children’s removal process. Numbers (1-9) were used as the stimuli since children have learned Arabic numbers in kindergarten and are quite familiar with numbers (Linares et al. 2016). During the task, participants were initially presented with a 1000-ms fixation cross in the center of the screen. Then, three different numbers were presented simultaneously for 4000 ms, each in one of the three black rectangular frames. This timing was sufficient for participants to fully encode all three numbers. This was followed by a set of updating steps. In each step, three empty black rectangles were first presented. Then, one of

\(^1\) An analysis conducted using all participants \((N = 70)\) produced results that were entirely consistent with the current findings (see SM1 in the supplementary materials).
the frames turned bold and red, indicating that the number in the frame was going to be replaced by a new number. Afterwards, a new number was presented in the cued frame, and participants were instructed to press the space bar to indicate that they had encoded the new number and successfully updated their WM set. The maximum response time was 5000 ms. The digits that needed to be remembered in each updating step were never in either an ascending or descending order (e.g., 1, 2, 3 or 6, 5, 4). It was less likely that children used the grouping/chunking strategy to memorize the digits since this strategy might have been less advantageous for children in flexibly replacing individual digit during each updating step, compared to maintaining them as separate numbers. The number of updating steps per trial varied between 1 and 9, with a constant 10% stopping probability after each updating step. The unpredictable stopping procedure prevented children from only remembering the last few numbers, and provided participants with a consistent incentive to perform each updating step regardless of the trial's duration.

At the end of each trial, participants had to recall all three numbers. The recall phase was prompted by a black question mark appearing in each frame one by one from left to right. Recall prompts were presented until a response was given. Upon all answers being given, feedback (i.e., x out of 3 is/are correct) was presented. The updating accuracy was the percentage of digits recalled correctly in the corresponding frames. Participants pressed a key to start the next trial.

**Figure 1**

_A schematic illustration of the working memory updating task_
Note. CTI = cue-target interval, RCI = response-cue interval.

There were three possible CTIs: 200 ms, 1500 ms, and 2800 ms. The CTI was chosen randomly at each updating step within a trial. The empty rectangles appeared in the response-cue interval prior to the CTI phase. This response-cue interval was complementary to the CTI, such that it ensured a constant 3000-ms overall retention interval in all cases (following Ecker, Lewandowsky, et al., 2014; Ecker, Oberauer, et al., 2014). Thus, the response-cue interval was 2800 ms, 1500 ms, and 200 ms, respectively.

Prior to the experiment, we clarified the task procedure using a figure akin to Figure 1. Participants were instructed to promptly mentally replace one of the three digits with a new one upon its appearance and to press the space bar once the substitution was completed. Participants needed to press the key as quickly as possible while ensuring that the substitution was completed. Following the instructions, each participant completed a practice session consisting of three trials, each with 3, 5, or 7 updating steps, respectively. The practice repeated once if recall accuracy was below 80%. After practice,
they completed a total of 24 trials, with an average of 6 to 7 updating steps per trial. There were approximately 52 updating steps per CTI condition. The task took approximately 35 min to complete.

Removal efficiency was evaluated by the proportional gain score following Ecker, Lewandowsky, et al., (2014). Specifically, we computed two proportional gain scores: the short-gain score was calculated as the mean updating RT of 200 ms-CTI minus the mean updating RT of 1500 ms-CTI, then divided by the mean updating RT of 200 ms-CTI. The long-gain score was calculated by replacing the mean updating RT of 1500 ms-CTI in the short gain score with the mean updating RT of 2800 ms-CTI.

**Cognitive control: the AX-CPT task**

The child-adapted version of the AX-CPT task (Gonthier et al., 2019) was used to tap cognitive control. The stimuli were animal pictures, consisting of hen, cat, horse, crocodile, snail, elephant, giraffe, rabbit, lion, sheep, snake, mouse, turtle, and cow. Participants were asked to press “J” with the right index finger only when they saw a cat (X, target) following a hen (A, cue). Otherwise, they had to press “F” with the left index finger.

Children’s knowledge of all these animals used in the task was firstly verified. Then, they were acquainted with the task and performed 16 practice trials, followed by three blocks of 30 formal trials (18 AX, 4 AY, 4 BX, and 4 BY). Trials within each block were pseudo-randomly arranged such that each block started with two AX trials; the same response occurred in no more than three consecutive trials; the sequences of 1 AX trial, 2 AX trials, and 3 AX trials appeared three times each; and AX trials were followed equally often by AY, BY and BX trials.

Each trial began with a fixation point centrally displayed on the screen for 500 ms. A cue animal picture was then presented for 1000 ms, followed by a 1500-ms interstimulus interval. Afterwards, a target animal picture was presented until children responded or until the time limit was reached. The response keys (“F” and “J”) appeared beneath the target picture to remind children to give a response. The time limit was set to 6000 ms during the practice trials but was calculated as the mean response time over all the preceding blocks plus one standard deviation for each individual during the formal trials. Feedback indicating whether each response was correct, incorrect, or too slow was displayed immediately after each trial.
response for 800 ms.

The average error rate and reaction time (RT) were firstly computed for each of the four trial types (AX, AY, BX, and BY). The proactive behavior index (PBI) was additionally computed as \((AY - BX)/(AY + BX)\) for error rate and RT, respectively. This index reflects the relative balance of interference between AY and BX trials, with a higher score representing more proactive control. In addition, a composite PBI was calculated by averaging the standardized PBIs obtained for error rate and RT. This index has the advantage of taking the trade-off between speed and accuracy into consideration (Gonthier et al., 2019). In cases where the error rate was equal to zero, a log-linear correction, i.e., \((\text{number of errors} + 0.5) / (\text{number of trials} + 1)\), was made to the error rate data before computing the error-rate PBI (see Braver et al., 2009; Gonthier et al., 2019 for a similar procedure).

**WMC tasks**

**Verbal WM.** The backward digit span task was used to tap verbal WMC (Gathercole et al., 2004; Prencipe et al., 2011). Participants were instructed to remember a sequence of digits (1-9) and then recall them in reverse order. The digits were presented at the rate of one per 750ms followed by 500ms blank screen intervals. At the end of each trial, participants were asked to recall the digits in reverse order by typing them into an input box. Participants first completed three practice trials of set size 3, and then performed trials of set sizes 3 to 7, with three trials per set size. The score was the number of correct digits recalled in the correct position divided by the number of trials.

**Visuospatial WM.** The Corsi block tapping task (Corsi, 1972) was adapted to tap visuospatial WMC by requiring backwards recall (Gonthier et al., 2019). Participants were asked to recall a series of red squares displayed within a \(4 \times 4\) matrix in reverse order. In each trial, the squares were presented successively in different cells of the matrix each for 750 ms, followed by a 500 ms inter-stimulus interval. At recall, a blank matrix appeared and participants had to recall the red squares in the opposite order to that which they were presented in by clicking the corresponding locations. Participants first performed three practice trials, followed by trials of set sizes 2 to 6, with two trials per set size. The score was the number of red squares recalled in the correct position divided by the number of trials.
Procedure

Participants were tested individually in a quiet room in their school. Two sessions were carried out on different days. In the first session, participants performed the WM updating task and the AX-CPT task, which were implemented with PsychoPy 3 (Peirce et al., 2019). In the second session, participants completed the WMC tasks, which were implemented with E-Prime 2.0 (Psychology Software Tools). The tasks were administered in the following order across all participants: the WM updating task, the AX-CPT, the backward digit span task, and the backward Corsi span task. Keeping the order of different tasks identical for all participants is a conventional way to reduce the contamination of additional method variance in differential research (e.g., Singh et al., 2018; Unsworth et al., 2014). Children had an opportunity to take a short break between tasks.

Results

For the WM updating task, trials with RTs < 300 ms were discarded, as were RTs more than three standard deviations away from the individual participant’s mean RT (following Singh et al., 2018). No more than 2% of the observations were affected by these trimming procedures. For the AX-CPT task, trials with RTs < 200ms and trials where children did not respond within the time limit were excluded. This represented between 5% and 15% of trials for all trial types. The average error rate was calculated without taking these trials into account (following Gonthier et al., 2019). RTs were computed on correct trials only. A composite WMC score was computed by averaging the standardized scores of the backward digit span and the backward Corsi span tasks. Descriptive statistics for all measures are shown in Table 1.

Table 1

Descriptive statistics for all measures as a function of age group

<table>
<thead>
<tr>
<th>Task</th>
<th>Measure</th>
<th>7-year-olds</th>
<th>9-year-olds</th>
<th>11-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Age</td>
<td>7.03</td>
<td>0.34</td>
<td>8.82</td>
<td>0.32</td>
</tr>
<tr>
<td>WM Updating</td>
<td>Accuracy</td>
<td>0.93</td>
<td>0.04</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>200ms-CTI RT</td>
<td>2.14</td>
<td>0.52</td>
<td>1.75</td>
</tr>
</tbody>
</table>
Age differences in the WM updating task

Updating Accuracy

Updating accuracy was high for the 7-year-olds ($M = 0.93, SD = 0.04$), 9-year-olds ($M = 0.95, SD = 0.05$), and 11-year-olds ($M = 0.97, SD = 0.03$), suggesting that children successfully remembered and updated the targets in most trials. The near-ceiling performance in this task was anticipated, given that the task merely required children to memorize three digits, a load well within the limits of their WMC. Moreover, each updating step was under their control, allowing children to press the key when they had successfully encoded the new digit into the number set. An analysis of variance (ANOVA) was performed on accuracy, with age group as the between-participants factor. The result revealed that updating accuracy increased significantly with age, $F(2, 63) = 7.59, p = .001, \eta^2_p = .19$. Bonferroni corrected post-hoc tests revealed that the updating accuracy of the 11-year-olds was significantly higher than that of the 7-year-olds ($p < .001$). However, the differences between the 7-year-olds and the 9-year-olds ($p = .137$), and between the 9-year-olds and the 11-year-olds ($p = .148$) were non-significant.

Updating RTs

Mean updating RTs for each age group are provided in Figure 2A. A $3 \times 3$ mixed ANOVA on updating RTs, with age group (7, 9, and 11 years old) as the between-participants factor and CTI (200 ms,
1500 ms, and 2800 ms) as the within-participants factor, yielded a significant main effect of CTI, $F(2,126) = 98.47, p < .001, \eta^2_p = .61$, and a significant main effect of age group, $F(2,63) = 20.64, p < .001, \eta^2_p = .40$. However, the interaction between age group and CTI was not significant, $F(4,126) = 0.20, p = .936, \eta^2_p = .01$, suggesting that CTI duration had similar impacts on updating RTs across the three age groups.

The effect of CTI on updating RTs found by Ecker, Lewandowsky, et al. (2014) was replicated in all age groups. That is, updating took significantly longer with a short CTI (200 ms) as compared to long CTIs (1500 ms and 2800 ms, all Bonferroni-corrected $p$s < .001), suggesting that children utilized the long CTIs to remove the outdated numbers. There was no significant difference between updating RTs of the 1500-ms CTI and the 2800-ms CTI (Bonferroni-corrected $p = .999$). In addition, there were significant differences in updating RTs between each of the two age groups (all Bonferroni-corrected $p$s < .01). That is, younger children were much slower in updating than older children.

**Figure 2**

(A) Updating response time for the three age groups per CTI condition
(B) Removal efficiency for the three age groups

Note. Vertical bars denote standard error.

**Removal efficiency scores**

We calculated two removal efficiency scores, specifically the proportional gain score between 200-ms CTI and 1500-ms CTI (short-gain score), and 200-ms CTI and 2800-ms CTI (long-gain score). Age differences in these short- and long-gain scores were analyzed separately. Figure 2B displays the mean removal efficiency scores for all age groups. ANOVA yielded a significant effect of age group on the long-gain score, $F(2,63) = 3.29$, $p = .044$, $\eta^2_p = .10$. Bonferroni-corrected post-hoc tests showed that the long-gain score of the 7-years-olds was significantly lower than that of the 11-years-olds ($p = .038$), while the difference between the 7-year-olds and the 9-year-olds ($p = .644$), and between the 9-year-olds and the 11-year-olds ($p = .627$) did not reach significance. However, when using the short-gain score as an index of removal efficiency, the main effect of age group was non-significant, $F(2,63) = 1.303$, $p = .279$, $\eta^2_p = .04$.

**Relations of the development of removal efficiency with cognitive control and WMC**

In this section, we examine whether the development of removal efficiency was accompanied by age-related increases in cognitive control and WMC. Table 2 presents the inter-correlations among age
(age was a continuous variable), the accuracy of the WM updating task, removal efficiency, PBI indices, and WMC, as well as the partial correlations among these variables with age controlled. First, we observed that there were significant correlations of the long-gain score with updating accuracy, error-rate PBI, composite PBI, and WMC. Second, the long-gain score remained associated with error-rate PBI and composite PBI even when age was statistically controlled for. However, the partial correlation between the long-gain score and WMC became non-significant as age was partialled out, suggesting that there is no obvious causal relationship between long-gain score and WMC, but that both concurrently improve with age. Third, the correlations of the short-gain score with PBIs and WMC were non-significant.

Therefore, subsequent analyses only took the long-gain score into consideration.

Table 2
Zero-order and partial correlations among removal efficiency, updating accuracy, PBIs, WMC, and age

<table>
<thead>
<tr>
<th></th>
<th>Long-gain score</th>
<th>Short-gain score</th>
<th>Updating Accuracy</th>
<th>Error-rate PBI</th>
<th>RT PBI</th>
<th>Composite PBI</th>
<th>WMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-gain score</td>
<td>—</td>
<td>.70***</td>
<td>.17</td>
<td>.40***</td>
<td>-.04</td>
<td>.25*</td>
<td>.17</td>
</tr>
<tr>
<td>Short-gain score</td>
<td>.72***</td>
<td>—</td>
<td>.21</td>
<td>.20</td>
<td>.01</td>
<td>.14</td>
<td>.08</td>
</tr>
<tr>
<td>Updating Accuracy</td>
<td>.29*</td>
<td>.27*</td>
<td>—</td>
<td>−.05</td>
<td>.10</td>
<td>.03</td>
<td>.37**</td>
</tr>
<tr>
<td>Error-rate PBI</td>
<td>.41***</td>
<td>.21</td>
<td>−.00</td>
<td>—</td>
<td>.12</td>
<td>.76***</td>
<td>−.05</td>
</tr>
<tr>
<td>RT PBI</td>
<td>−.01</td>
<td>.03</td>
<td>.12</td>
<td>.13</td>
<td>—</td>
<td>.74***</td>
<td>.14</td>
</tr>
<tr>
<td>Composite PBI</td>
<td>.27*</td>
<td>.16</td>
<td>.08</td>
<td>.76***</td>
<td>.74***</td>
<td>—</td>
<td>.06</td>
</tr>
<tr>
<td>WMC</td>
<td>.32**</td>
<td>.18</td>
<td>.55***</td>
<td>.02</td>
<td>.16</td>
<td>.12</td>
<td>—</td>
</tr>
<tr>
<td>Age</td>
<td>.30*</td>
<td>.19</td>
<td>.48***</td>
<td>.09</td>
<td>.08</td>
<td>.11</td>
<td>.62***</td>
</tr>
</tbody>
</table>

*Note. The upper triangle represents partial correlations, and the lower triangle represents zero-order Pearson correlations. PBI = proactive behavioral index, WMC = Working memory capacity.

*** p < .001, ** p < .01, * p < .05

In order to further clarify the relationship between the development of removal efficiency and cognitive control, we implemented a variance partitioning technique using multiple linear regression (following Gonthier et al., 2019; Unsworth & Spillers, 2010) to compute the shared and unique contributions of age and cognitive control to removal efficiency (see Figure 3A). While age and error-rate PBI explained 9.13% and 16.65% of the variance in the long-gain score respectively, the combination of
the two accounted for 23.76% of the variance. These results suggest that only 2.02% of the variance 
(2.02% = 9.13% + 16.65% – 23.76%) was explained in common by age and error-rate PBI. Both age 
(7.11% = 9.13% – 2.02%, p = .018) and error-rate PBI (14.63% = 16.65% – 2.02%, p < .001) accounted 
for a unique portion of variations in the long-gain score even when the other one was partialled out. These 
results suggest that cognitive control explained removal efficiency above and beyond age. That is, 
children who were more proficient in using proactive control could remove irrelevant information more 
efficiently than their counterparts who were less proficient.

To confirm the stability of these findings, we also conducted a comparable set of analyses using 
composite PBI. Specifically, we computed the shared and unique contributions of age and composite PBI 
to the long-gain score. While age and composite PBI explained 9.13% and 7.14% of the variance in the 
long-gain score respectively, the combination of the two accounted for 14.68% of the variance. These 
results indicate that only 1.59% of the variance (1.59% = 9.13% + 7.14% – 14.68%) was explained by the 
combination of age and composite PBI. Both age (7.54% = 9.13% – 1.59%, p = .021) and composite PBI 
(5.55% = 7.14% – 1.59%, p = .047) explained a unique portion of variance in the long-gain score even 
when the other one was partialled out.

Similar analyses were conducted to clarify the relationship between the development of removal 
efficiency and WMC (see Figure 3B). While age and WMC explained 9.13% and 10.09% of the variance 
in the long-gain score, respectively, the combination of the two accounted for 11.89% of the variance. 
These results show that the majority of variance (7.33% = 9.13% + 10.09% – 11.89%) was accounted for 
by the combination of age and WMC. When the shared variance was partialled out, neither the unique 
contribution of age (1.80% = 9.13% – 7.33%, p = .260) nor that of WMC (2.76% = 10.09% – 7.33%, p 
= .165) was significant. These results suggest that the majority of the age-related variance (80.28% = 
7.33%/9.13%) in removal efficiency is accounted for by the increase of WMC.

Figure 3

Percentages of variance in removal efficiency accounted for by age, cognitive control (A) and WMC (B)
Since there is also the possibility that the development of removal efficiency promotes the increase of WMC, we ran the regression analysis in which WMC was regressed on age and removal efficiency. When WMC was considered as the dependent variable, age and the long-gain score explained 38.34% and 10.09% of the variance in WMC respectively; the combination of the two accounted for 40.21% of the variance. The shared variance between age and removal efficiency therefore accounted for 8.22% (38.34% + 10.09% – 40.21%) of the variance in WMC. While age uniquely accounted for the majority of variance (30.12% = 38.34% – 8.22%, \( p < .001 \)), the unique effect of removal efficiency was non-significant (1.87% = 10.09% – 8.22%, \( p = .165 \)). These results suggest that only a small portion of the age-related growth (21.44% = 8.22%/38.34%) in WMC is explained by the improvement of removal efficiency.

**Discussion**

The present study aimed to examine age-related differences in the ability to actively remove irrelevant information from WM and its relation with cognitive control and WMC. To these ends, an established WM updating task (Ecker, Lewandowsky, et al., 2014) was modified to assess the removal efficiency of school-aged children. In addition, a battery of cognitive control and WMC tasks was administered. In line with previous findings with adults (Ecker, Lewandowsky, et al., 2014; Ecker, Oberauer, et al., 2014; Singh et al., 2018), participants updated information faster when CTIs were long than when CTIs were short, suggesting that children aged 7 to 11 can actively remove items from WM.
before encoding new ones. As expected, our results showed an improvement in removal efficiency with age. Age-related increases in removal efficiency occurred concurrently with age-related growth in WMC. In addition, cognitive control predicted removal efficiency over and beyond age and WMC.

The results demonstrated that the manipulation of the CTI had a considerable effect on updating response times across all age groups. This result can be added to those obtained with adults and provides a broader picture of the role of active removal in children’s WM updating. In the present study, even the youngest children (7 years old) showed shorter response times in the long-CTI conditions than in the short-CTI condition, suggesting that children are ready to remove the to-be-updated items from the focus of attention before encoding the new ones. Moreover, like adults, children were able to remove information within 1500 ms because providing more time for removal did not reduce updating response time further (see Figure 2A).

An alternative explanation for the RT difference between short and long CTIs is that participants may not actively remove outdated information, but instead simply stop rehearsing the cued items and let them passively decay. Long CTIs would allow for more decay than the short CTI, thus the difference in updating time between long and short CTIs may reflect time for passive decay instead of active removal. This hypothesis seems reasonable in theory, but has been falsified by previous experimental evidence. First, in Experiment 1 of Ecker, Lewandowsky, et al. (2014), the outdated letter was mostly replaced with a new one, but occasionally this letter was the same as the previous one (i.e., repeating trials). They found that a short CTI led to reduced updating RTs when the old and new items were identical, indicating a faster updating process. However, this benefit in updating RTs significantly diminished with a long CTI. This finding suggests that the extended CTI is actively used to remove the old item from WM.

Second, in Experiment 3 of Ecker, Oberauer, et al. (2014), where two frames were updated at the same time, the RT benefit associated with occasional item repetition vanished with a long CTI, but only in the left updated frame, not the right. Ecker and colleagues claimed that people follow a left-to-right scanning pattern when updating a list and remove only one item during the long CTI, even when two items need to be updated. This strategy prevents the costs linked with frame switches and processing-
mode switches. Consequently, the benefit of item repetition remained in the right to-be-updated frame, which was not affected by removal, but disappeared in the left to-be-updated frame because the memorized item had been removed. The idea that selectively rehearsing relevant items pushes irrelevant items out of the active WM system or allows them to decay cannot explain why one irrelevant item remains in WM while another does not. In other words, selective rehearsal of to-be-remembered items (combined with decay) implies that all non-rehearsed items should disappear from WM at the same rate. However, this contradicts the observed differences in repetition benefits for items to be removed on the left and right sides of the memory set after a long CTI. Thus, the findings of this experiment bolster the hypothesis of active removal and set it apart from the rehearsal (passive decay) hypothesis.

Although such repeat trials therefore help clarify the removal process, we did not include them in our study because of several considerations: (1) the primary measure of removal efficiency is calculated based on the comparison of updating times between long and short CTIs (following Ecker, Lewandowsky, et al., 2014; Singh et al., 2018; Rey-Mermet et al., 2020). In these studies, repeat trials were not included when measuring removal efficiency. We therefore adopted a similar experimental design to these well-established studies with adults. (2) The inclusion of repeat trials may alter participants’ strategy when they notice that occasionally they do not need to remove the items proactively, which may impair the validity of the measurement of removal efficiency. (3) As stated by Ecker, Lewandowsky, et al. (2014): “In repetition trials with short CTI, participants do not remove the old item, and their RTs reflect the time for detecting the identity between the old and the new item, for stopping the default updating process, and perhaps for carrying out a refreshing operation on the old (and still current) representation in WM instead.” Therefore, repeat trials do not provide a valid measure of the removal process. Having said all this, there would clearly be considerable value in conducting more experiments to definitively exclude the rehearsal hypothesis in children.

We did not find a meaningful interaction between CTI conditions and age. At first glance, this result seems to suggest that removal ability does not increase with age. However, the difference between the long and short CTI conditions is not an appropriate index for estimating removal efficiency since it is
confounded by general processing speed (Rey-Mermet et al., 2019). The difference between RTs is inherently proportionally larger for slower than faster participants. For example, suppose a faster participant has RTs of 100 and 200 ms in the long and short CTI conditions respectively, while a slower participant has RTs of 200 and 300 ms respectively. Subtracting RTs would result in the same removal score for each participant, but there is actually a smaller proportional increase for the slower individuals (of a factor of 1.5) than for the faster individuals (of a factor of 2). In order to avoid such problems, we calculated the proportional gain score to index removal efficiency. With this index (i.e., long-gain score), we found that removal efficiency improved from 7 to 11 years of age.

A novel finding of the current study is that proactive control accounted for individual differences in removal efficiency, even when age was statistically partialled out. This result indicates that children manage to remove outdated information more efficiently when they are more prone to use proactive control, confirming the “active” nature of removal in WM among children. As previous studies indicate, children older than 5 years old are able to use proactive control to complete cognitive tasks (Chatham et al., 2009; Gonthier et al., 2019; Yanaoka et al., 2022). Proactive control seems to be a prerequisite for engaging in efficient active removal since participants are able to prepare themselves to sweep away the to-be-updated items as they see the cue. Otherwise, if individuals use reactive control, or employ proactive control inefficiently, they will only start to remove outdated items at the point when targets are detected, resulting in smaller or even no proportional changes in RTs between the long and short CTI conditions. However, it should be noted that longitudinal research is needed to properly elucidate the causal relationship between cognitive control and removal.

The finding that age was no longer predictive of removal efficiency when the shared variance of age and WMC was partialled out suggests that the growth of removal efficiency might be accounted for by age-related improvements in WMC. One of the prominent roles of WM is to hold task-relevant goals in mind in the face of interference (Unsworth et al., 2014). As WMC increases with age, children are more able to proactively maintain the goal of replacing the cued number with a new one, leading to improved removal efficiency. When WMC was taken as the dependent variable, age remained a
significant unique predictor of a large portion of variance in WMC, in contrast to removal efficiency which did not account for significant unique variance in WMC. This result is not in accordance with the theoretical suggestion that removal plays a central role in protecting to-be-remembered items from interference thereby leading to the maintenance of more relevant information (Oberauer & Lewandowsky, 2016). Rather, this finding is more aligned with previous reports of a non-significant correlation between removal efficiency and WMC (Ecker, Lewandowsky, et al., 2014). Our study extends this result from adults to children by revealing that the development of WMC might not be primarily driven by the improvement in removal efficiency. Again, our cross-sectional data do not allow for a causal inference regarding the precise relationship between removal efficiency and WMC.

Removal efficiency was weakly associated with updating accuracy, and this correlation became non-significant after controlling for age. This result should not be interpreted as indicating that removal is unimportant for updating. The WM updating task employed here was not complicated, and updating accuracy was high across all age groups of children (≥ .93), potentially restricting the correlation between updating accuracy and removal efficiency. In addition, updating accuracy is not only influenced by removal efficiency, but also by other factors, such as encoding, maintenance, and retrieval processes.

One might question the extent to which children are fully aware of when they have successfully updated target numbers, which would raise potential questions about the reliability of the RTs associated with their key presses in the updating task. Indeed meta-cognition is known to still be developing during childhood, but children as young as 5 years old show robust metacognitive monitoring skills (Destan et al., 2014). It is again worth noting that recall accuracy was consistently high across all age groups, suggesting that children did update the targets effectively as they pressed the key to proceed. In addition, we observed a clear and consistent RT difference between long- and short-CTI conditions, aligning with findings in adult research. These findings provide compelling evidence for the validity of this approach in

An analysis conducted using only perfect trials with no recall errors produced results that were entirely consistent with the current findings (see SM2 in the supplementary materials).
measuring updating efficiency in children. Moreover, the bootstrap split-half reliability of removal efficiency scores ($\rho = .59$ and .50 for the long- and short-gain score, respectively) was similar to the reliability estimates reported in other individual difference studies using a similar task (e.g., Ecker, Lewandowsky, et al., 2014; Rey-Mermet et al., 2020; Singh et al., 2018). Difference scores are inherently less reliable than raw scores (Draheim et al., 2019), and although our removal efficiency scores were proportional scores, rather than being simple differences, they still depend on the difference in RTs across conditions. As a result, this may have influenced the estimates of the correlations between these scores and other variables. Indeed, the absence of statistically significant correlations between the short-gain score and other variables might be due to the somewhat poorer reliability of the short-gain score.

It should also be noted that we did not observe a significant age-related increase in cognitive control in this study. This result is compatible with that of Gonthier et al. (2019), who only showed a shift from reactive control to proactive control between pre-kindergarten (4-5-year old) and kindergarten (5-6-year old) children, without observing any quantitative change in PBI scores between kindergarten and first grade (6-7-year old) children. This might be because the AX-CPT task adjusted the time allowed to respond according to each child’s ability, which necessarily reduces age-related variations. That said, we observed considerable individual differences in PBI scores, which were predictive of removal efficiency even when age was partialled out.

Several limitations of the current study should be mentioned. First, the current study employed numbers as the sole type of experimental stimuli in the WM updating task. However, it would be important to test the generalization of our findings by using other types of stimuli such as letters, words, and figures to tap removal efficiency more broadly in future studies. Second, we did not conduct a priori power analysis to estimate sample size for the correlation and regression analyses since there are no obviously relevant prior data available for evaluating expected effect sizes. We therefore instead conducted post-hoc analyses based on the effect sizes obtained by the current study. For the correlation analysis, the post-hoc power analysis showed that with an estimated effect size of .41 and .32 (the correlations of removal efficiency with cognitive control and WMC were .41 and .32, respectively), the
power was .94 and .76, with a sample size of $N = 66$ and a significance criterion of $\alpha = .05$. For the regression analysis, the power was .98 when age and cognitive control were used to predict removal efficiency, considering a sample size of $N = 66$ and a significance criterion of $\alpha = .05$. When age and WMC were the predictors for removal efficiency, the power was .75 with the same sample size and significance criterion. Although these analyses suggest that the correlation and regression analyses were carried out with an acceptable/good statistical power given the current sample size, a more controlled study with a larger sample could usefully be conducted to replicate and extend the current findings in children.

To conclude, we have, for the first time, provided direct empirical evidence that children can engage in active removal of the contents of their WM, as indexed by the well-established WM updating paradigm. Children showed considerable growth in removal efficiency with age from 7 to 11 years. The clear majority of the age-related increase in removal efficiency was accounted for by the development of WMC. Crucially, individual differences in removal efficiency were closely related to the efficient use of proactive control, showing that at least some aspects of the removal of the outdated contents of working memory are open to strategic control, even in children.

References


JASP Team (2022). JASP (Version 0.16.1.0)[Computer software].


Pelegrina, S., Molina, R., Rodríguez-Martínez, E. I., Linares, R., & Gómez, C. M. (2020). Age-related changes in selection, recognition, updating and maintenance information in WM. An ERP study


SM1: An analysis conducted using all participants (N = 70)

To further ensure the robustness of our results, we conducted additional analyses involving all participants (N = 70) and found that the results were consistent with our findings reported in the manuscript. The results from the ANOVA, correlation, and regression analyses, encompassing all participants (N = 70), are outlined below.

A 3 × 3 mixed ANOVA on updating RTs, with age group (7, 9, and 11 years old) as the between-participants factor and CTI (200 ms, 1500 ms, and 2800 ms) as the within-participants factor, yielded a significant main effect of CTI, $F(2,134) = 104.85, p < .001, \eta^2_p = .61$, and a significant main effect of age group, $F(2,67) = 18.11, p < .001, \eta^2_p = .35$, However, the interaction between age group and CTI was nonsignificant, $F(4, 134) = 0.16, p = .961, \eta^2_p = .01$, suggesting that CTI duration had similar impacts on updating RTs across the three age groups (see Figure S1 below).

**Figure S1**

*Updating response time for the three age groups per CTI condition (N = 70)*

![Graph showing updating response time for three age groups per CTI condition](image-url)
The table S1 below presents the inter-correlations among age, updating accuracy, removal efficiency, PBI indices, and WMC, as well as the partial correlations among these variables with age controlled. First, we observed that there were significant correlations of the long-gain score with error-rate PBI, composite PBI, and WMC. Second, the long-gain score remained associated with error-rate PBI and composite PBI even when age was statistically controlled for. However, the partial correlation between the long-gain score and WMC became non-significant as age was partialled out.

**Table S1**

*Zero-order and partial correlations among removal efficiency, updating accuracy, PBIs, WMC, and age (N = 70)*

<table>
<thead>
<tr>
<th></th>
<th>Long-gain score</th>
<th>Short-gain score</th>
<th>Updating Accuracy</th>
<th>Error-rate PBI</th>
<th>RT PBI</th>
<th>Composite PBI</th>
<th>WMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-gain score</td>
<td>—</td>
<td>.70***</td>
<td>.10</td>
<td>.41***</td>
<td>−.03</td>
<td>.25*</td>
<td>.17</td>
</tr>
<tr>
<td>Short-gain score</td>
<td>.71***</td>
<td>—</td>
<td>.17</td>
<td>.20</td>
<td>.03</td>
<td>.16</td>
<td>.10</td>
</tr>
<tr>
<td>Updating Accuracy</td>
<td>.23</td>
<td>.24*</td>
<td>—</td>
<td>−.08</td>
<td>−.05</td>
<td>−.08</td>
<td>.41**</td>
</tr>
<tr>
<td>Error-rate PBI</td>
<td>.41***</td>
<td>.21</td>
<td>−.03</td>
<td>—</td>
<td>.11</td>
<td>.75***</td>
<td>−.02</td>
</tr>
<tr>
<td>RT PBI</td>
<td>−.01</td>
<td>.05</td>
<td>−.02</td>
<td>.12</td>
<td>—</td>
<td>.75***</td>
<td>.10</td>
</tr>
<tr>
<td>Composite PBI</td>
<td>.27*</td>
<td>.17</td>
<td>−.03</td>
<td>.75***</td>
<td>.75***</td>
<td>—</td>
<td>.06</td>
</tr>
<tr>
<td>WMC</td>
<td>.32**</td>
<td>.21</td>
<td>.57***</td>
<td>.04</td>
<td>.12</td>
<td>.11</td>
<td>—</td>
</tr>
<tr>
<td>Age</td>
<td>.31*</td>
<td>.20</td>
<td>.46***</td>
<td>.09</td>
<td>.06</td>
<td>.10</td>
<td>.64***</td>
</tr>
</tbody>
</table>

*Note.* The upper triangle represents partial correlations, and the lower triangle represents zero-order Pearson correlations. PBI = proactive behavioral index, WMC = Working memory capacity.

***p < .001, **p < .01, *p < .05

**Regression analyses:** While age and error-rate PBI explained 9.40% and 17.10% of the variance in the long-gain score respectively, the combination of the two accounted for 24.40% of the variance. These results suggest that only 2.10% of the variance (2.10% = 9.40% + 17.10% − 24.40%) was explained in common by age and error-rate PBI. Both age (p = .013) and error-rate PBI (p < .001) accounted for a unique portion of variations in the long-gain score even when the other one was partialled out.
While age and WMC explained 9.40% and 10.30% of the variance in the long-gain score, respectively, the combination of the two accounted for 12.00% of the variance. These results show that the majority of variance (7.70% = 9.40% + 10.30% – 12.00%) was accounted for by the combination of age and WMC. When the shared variance was partialed out, neither the unique contribution of age ($p = .251$) nor that of WMC ($p = .160$) was significant. These results suggest that the majority of the age-related variance (81.91% = 7.70%/9.40%) in removal efficiency is accounted for by the increase of WMC.

When WMC was considered as the dependent variable, age and the long-gain score explained 40.50% and 10.30% of the variance in WMC respectively; the combination of the two accounted for 42.20% of the variance. The shared variance between age and removal efficiency therefore accounted for 8.60% (40.50% + 10.30% – 42.20%) of the variance in WMC. While age uniquely accounted for the majority of variance (31.90% = 40.50% – 8.60%, $p < .001$), the unique effect of removal efficiency was non-significant ($p = .160$). These results suggest that only a small portion of the age-related growth (21.23% = 8.60%/40.50%) in WMC is explained by the improvement of removal efficiency.

**SM2: An analysis conducted using only perfect trials with no recall errors**

In this section, we conducted the analyses solely based on these perfect trials (four participants from the 11-year-old group were excluded due to the absence of recorded recall accuracy on each trial, and their overall accuracy did not reach 100%. This issue did not occur with other participants). The results remained consistent with our prior findings, reaffirming the robustness of our conclusions. The results from the ANOVA, correlation, and regression analyses, encompassing only perfect trials, are outlined below.

A $3 \times 3$ mixed ANOVA on updating RTs, with age group (7, 9, and 11 years old) as the between-participants factor and CTI (200 ms, 1500 ms, and 2800 ms) as the within-participants factor, yielded a significant main effect of CTI, $F(2,118) = 64.15, p < .001, \eta^2_p = .52$, and a significant main effect of age group, $F(2,59) = 22.00, p < .001, \eta^2_p = .43$. However, the interaction between age group and CTI was
nonsignificant, $F(4,118) = 0.41, p = .804, \eta^2_p = .01$, suggesting that CTI duration had similar impacts on updating RTs across the three age groups. (see Figure S2 below).

**Figure S2**

*Updating response time for the three age groups per CTI condition (only perfect trials)*

The table S2 below presents the inter-correlations among age, updating accuracy, removal efficiency, PBI indices, and WMC, as well as the partial correlations among these variables with age controlled. First, we observed that there were significant correlations of the long-gain score with updating accuracy, error-rate PBI, composite PBI, and WMC. Second, the long-gain score remained associated with error-rate PBI and composite PBI even when age was statistically controlled for. However, the partial correlation between the long-gain score and WMC became non-significant as age was partialled out.

**Table S2**

*Zero-order and partial correlations among removal efficiency, updating accuracy, PBIs, WMC, and age (only perfect trials)*

<table>
<thead>
<tr>
<th></th>
<th>Long-gain score</th>
<th>Short-gain score</th>
<th>Updating Accuracy</th>
<th>Error-rate PBI</th>
<th>RT PBI</th>
<th>Composite PBI</th>
<th>WMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-gain score</td>
<td>—</td>
<td>.61***</td>
<td>.16</td>
<td>.41***</td>
<td>.00</td>
<td>.28*</td>
<td>.17</td>
</tr>
</tbody>
</table>

Short-gain score  .65*** — .24 .17 −.03 .09 .08
Updating Accuracy .29* .33** — −.06 .10 .03 .37**
Error-rate PBI .42*** .19 −.01 — .12 .75*** −.05
RT PBI .02 −.02 .11 .13 — .75*** .17
Composite PBI .29* .11 .07 .75*** .75*** — .08
WMC .33** .22 .56*** .02 .16 .12 —
Age .33* .27* .49*** .10 .05 .10 .61***

Note. The upper triangle represents partial correlations, and the lower triangle represents zero-order Pearson correlations. PBI = proactive behavioral index, WMC = Working memory capacity.

*** p < .001, ** p < .01, * p < .05

**Regression analyses:** While age and error-rate PBI explained 10.60% and 17.60% of the variance in the long-gain score respectively, the combination of the two accounted for 25.80% of the variance. These results suggest that only 2.40% of the variance (2.00% = 10.60% + 17.60% − 25.80%) was explained in common by age and error-rate PBI. Both age (8.20%, p = .013) and error-rate PBI (15.20%, p < .001) explained a unique portion of variations in the long-gain score even when the other one was partialled out.

While age and WMC explained 10.60% and 10.80% of the variance in the long-gain score respectively, the combination of the two accounted for 13.30% of the variance. These results show that the majority of variance (8.10% = 10.60% + 10.80% − 13.30%) was accounted for by the combination of age and WMC. When the shared variance was partialled out, neither the unique contribution of age (2.50%, p = .197) nor that of WMC (2.70%, p = .184) was significant. These results suggest that the majority of the age-related variance (76.42% = 8.10%/10.60%) in removal efficiency is accounted for by the increase of WMC.

When WMC was considered as the dependent variable, age and the long-gain score explained 37.30% and 10.80% of the variance in WMC respectively; the combination of the two accounted for 39.20% of the variance. The shared variance between age and removal efficiency therefore accounted for 8.90% (37.30% + 10.80% − 39.20%) of the variance in WMC. While age uniquely accounted for the majority of variance (28.40%, p < .001), the unique effect of removal efficiency was non-significant.
(1.90\%, p = .184). These results suggest that only a small portion of the age-related growth (23.86\% = 8.90%/37.30\%) in WMC is explained by the improvement of removal efficiency.