Working Memory Training in Old Age: An Examination of Transfer and Maintenance Effects

Erika Borella*, Barbara Carretti, Giulia Zanoni, Michela Zavagnin, Rossana De Beni

Department of General Psychology, University of Padova, Padova, Italy

*Corresponding author at: Department of General Psychology, Via Venezia 8, 35131 Padova, Italy. Tel.: +39-049-8276035; fax: +39-049-8276600.
E-mail address: erika.borella@unipd.it (E. Borella) or barbara.carretti@unipd.it (B. Carretti).

Accepted 17 March 2013

Abstract

The present study examined the efficacy of a verbal working memory (WM) training program in old-old individuals (over 75 years of age). Thirty-six adults aged 75–87 took part in the study: 18 were randomly assigned to receive training and the remainder served as active controls. Specific training gains in a verbal WM task (criterion task), and transfer effects on measures of visuospatial WM, short-term memory, inhibition, processing speed, and fluid intelligence were examined. The trained old-old adults performed better than the controls in the criterion task, and this benefit persisted after 8 months; they also showed an increase in the efficiency of inhibitory mechanisms at follow-up compared with pretest. The results of this study suggest that the present WM training program produces benefits maintained over time even in old-old adults. These findings confirm that there is still room for plasticity in the basic mechanisms of cognition in advance old age.

Keywords: working memory; cognitive training; transfer effects; maintenance effects; cognitive flexibility; old-old adults, aging

Introduction

Working memory (WM) is one of the basic cognitive mechanisms that enable us to process and retain temporary information for use in other cognitive tasks (Miyake & Shah, 1999). Its role is particularly crucial because of its involvement in higher-order cognitive abilities also relating to everyday life, such as fluid intelligence, problem-solving, and reading comprehension. Given its central role in cognition, training activities targeting WM have thus been assumed to produce wider effects on the cognitive system. A growing number of studies are therefore focusing on whether WM performance can be modified, and enhanced, by training and whether improving this general mechanism has a beneficial effect on other cognitive abilities that are not trained directly (transfer effects).

Effective WM training and training-related improvements in untrained skills (e.g., fluid intelligence) have been reported in various sample populations, such as children with typical development (e.g., Jaeggi, Buschkuehl, Jonides, & Shah, 2011), young adults (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008), and adults with brain injuries (e.g., Lundqvist, Grundström, Samuelsson, & Rönberg, 2010). Since WM has a well-documented role in explaining age-related changes in older adults (Borella, Carretti, & De Beni, 2008; Park et al., 2002), the possibility of enhancing WM in the elderly by means of training has also been examined in older adults.

These new studies on the effect of training programs are of considerable interest because they aim to examine the extent and the nature of cognitive plasticity, that is, how the age-related decline in cognition might be slowed by improving cognitive skills and existing cognitive resources (e.g., Mayr, 2008). Until recently, cognitive training studies involving older adults focused mainly on episodic memory and consisted in training people to use mnemonic strategies. These studies generally found that teaching strategies could improve the older adults’ memory (e.g., Lustig, Shah, Seider, & Reuter-Lorenz, 2009; Singer, Lindenberger, & Baltes, 2003), meaning that there is a certain potential even in older age to improve performance, acquire new skills, and make more efficient use of available cognitive resources—what is called “cognitive plasticity”—after training.
The main drawback of the above-mentioned research is that it is limited to a given domain or for a particular task, whereas their transfer to other tasks/situations is rare, especially in older adults, who often do not spontaneously use such strategies after completing their training (e.g., Rebk, Carlson, & Langbaum, 2007; Verhaeghen & Marcoen, 1996; Verhaeghen, Marcoen, & Goossens, 1992). Studies on WM training have tried to overcome these limitations (related to the development of task-specific strategies) and focus on a central mechanism of cognition—WM—that: (a) is involved in a wide variety of abilities of higher-order cognition (attention, reasoning, etc.) and (b) relies on the frontoparietal network, which shows a more pronounced decline in aging than other brain networks.

WM training, therefore, has the potential to improve performance on the criterion measure and also to have transfer effects on broader (untrained) abilities that share the same processes, thereby facilitating cognitive plasticity. The general procedure adopted for this training involved asking participants (older adults) to practice repeatedly with a WM task, without teaching them any context-specific strategies that would help them only to complete the task in question. The WM training thus seemed to allow changes in participants’ underlying ability to induce a process-based plasticity, with consequences for other WM-related and the ability to engage similar processes or brain regions (for a review, see Buschkuehl, Jaeggi, & Jonides, 2012). In fact, WM training succeeded in inducing changes in cognitive performance associated with changes in the neural activity of the prefrontal cortex, when young adults (e.g., Takeuchi et al., 2010; Westerberg & Klingberg, 2007) or children (Klingberg, Forssberg, & Westerberg, 2002) are examined, at least. To our knowledge, only one study has examined this issue in aging, showing an increased activation of the striatum in older adults trained with an n-back task (Dahlin, Nyberg, Backman, & Stigsdotter Neely, 2008).

To date, only seven studies have investigated the effects of WM training (only studies using WM tasks or updating WM tasks were selected for the training sessions) in older adults (Table 1) using procedures in which they were trained with complex WM tasks (Borella, Carretti, Riboldi & De Beni, 2010; Brehmer, Westerberg, & Backman, 2012; Buschkuehl et al., 2008; Li et al., 2008; Richmond, Morrison, Chein, & Olson, 2011; Schmiedek, Lövđén, & Lindenberger, 2010) or with updating WM tasks (Dahlin et al., 2008). A review of the studies highlights that training programs differ in many respects (as summarized in Table 1), and this makes it difficult to identify the features and variables capable of facilitating the beneficial effects of WM training in older adults. The studies vary in terms of: (a) the type of task used to train participants (classical WM, updating WM, n-back tasks, or a combination of classical and updating WM tasks); (b) the nature of the training (verbal, visual, spatial, or verbal and visual combined); (c) the duration of the training (ranging from 3 to 45 sessions) and the frequency of the sessions; (d) the testing of any long-term maintenance benefits (which was not done by Schmiedek et al., 2010, or Richmond et al., 2011); (e) the older adults’ age range; and (f) the use of an active control group. Notwithstanding these limitations, they all reported gains in the trained task—the so-called specific effects—which were also maintained (apart from the previously mentioned studies by Schmiedek et al., 2010, and Richmond et al., 2011, which did not include a follow-up session). They also showed transfer effects for tasks very similar to the trained task (i.e., near transfer effects). It should be noted, however, that the near transfer effects cannot be considered general and robust, given that one study failed to find any (Dahlin et al., 2008), and four others found them only for some, not all, of the tasks considered (Buschkuehl et al., 2008; Li et al., 2008; Richmond et al., 2011; Schmiedek et al., 2010). As for the training gains in dissimilar tasks (far transfer effects), either they were not found (Dahlin et al., 2008; Li et al., 2008) or they were seen only for some of the transfer tasks considered (Brehmer et al., 2012; Buschkuehl et al., 2008; Richmond et al., 2011; Schmiedek et al., 2010; Table 1) and, even when they were found, these effects were limited and short-lived (e.g., Buschkuehl et al., 2008).

Different results were obtained by Brehmer and colleagues (2012) and by Borella and colleagues (2010): in both their studies, WM training was offered to a younger sample of elderly adults, aged in the range of 60–70 or 64–74, respectively, that is, the so-called young-old (e.g., Magaziner, 1989). Both these studies also included an active control group and a follow-up session to ascertain any maintenance effects. Borella and colleagues (2012) found both specific and transfer effects on sustained attention, short-term memory, and a measure of cognitive functioning (self-rating scale), while they found no transfer effects on interference control, reasoning, or episodic memory; both the training gains and the transfer effects were maintained at the 3-month follow-up.

Borella and colleagues (2010) also found specific gains in the trained task but, unlike Brehmer and colleague, they identified transfer effects too (with large effect sizes) on short-term memory, reasoning, inhibition, and processing speed; after 8 months, the specific gains and some of the transfer effects (reasoning and processing speed) also reportedly persisted. The specific gains, and their transfer and maintenance, were interpreted by Borella and colleagues as effects of their training procedure that, unlike all the other studies, combined an adaptive procedure (the difficulty of the training task was increased if participants were successful at a given level; if not, the lowest level was presented) with a constant variation in the maintenance and processing requirements of the trained task (to avoid practice effects). The authors suggest that this training regimen may promote transfer effects because it combines the involvement of multiple cognitive processes (attentional control, shifting) with an adaptive procedure (which enables participants to be trained at a level of difficulty coming close to the limits of their capacity),
<table>
<thead>
<tr>
<th>Training task</th>
<th>Length of training session and frequency</th>
<th>Months since training: Follow-up</th>
<th>Age range of older adults</th>
<th>Control group</th>
<th>Specific effects</th>
<th>Near transfer effects</th>
<th>Far transfer effects</th>
<th>Maintenance effects</th>
<th>Specific near far transfer effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li and colleagues (2008)</td>
<td>Spatial n-back task-two-back task</td>
<td>45 daily sessions</td>
<td>3 months</td>
<td>70–80 ($M = 74.5 ± 2.8$)</td>
<td>No</td>
<td>Yes</td>
<td>Yes (the WM tasks were the same as the criterion task with either a different modality or the same modality but a different version of the task, i.e., a three-back task)</td>
<td>Not to complex WM span tasks</td>
<td>Yes</td>
</tr>
<tr>
<td>Buschkuehl and colleagues (2008)</td>
<td>Visual WM tasks</td>
<td>24 sessions, 2 sessions/ week, 45 min each</td>
<td>12 months</td>
<td>NR ($M = 80 ± 3.3$)</td>
<td>Active: Physical training</td>
<td>Yes</td>
<td>Yes: To the Digit Span Task but Not to the visual block span</td>
<td>Yes: To visual free recall task but Not to verbal free recall task</td>
<td>No</td>
</tr>
<tr>
<td>Dahlin and colleagues (2008)</td>
<td>Updating task (letter-memory task)</td>
<td>15 sessions, 3 sessions/week, 45 min each</td>
<td>18 months</td>
<td>NR ($M = 68.38 ± 1.66$)</td>
<td>No-contact/passive*</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Borella and colleagues (2010)</td>
<td>CWMS tasks</td>
<td>3 sessions</td>
<td>8 months</td>
<td>65–75 ($M = 69 ± 3.18$)</td>
<td>Active: Alternative activities (questionnaires)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes: To processing speed, inhibition, and fluid intelligence tasks</td>
<td>Yes</td>
</tr>
<tr>
<td>Schmiedek and colleagues (2010)</td>
<td>Verbal, non-verbal, and updating WM task</td>
<td>10 sessions, 2 h each</td>
<td>NE</td>
<td>65–80 ($M = 71.3 ± 4.1$)</td>
<td>No-contact/passive*</td>
<td>Yes</td>
<td>Mixed: Only to one of the three WM/updating tasks considered</td>
<td>Mixed: Only to one of the three complex span tasks considered; to reasoning, and to episodic memory tasks</td>
<td>NE</td>
</tr>
</tbody>
</table>

(continued on next page)
Table 1.
(continued)

<table>
<thead>
<tr>
<th>Training task</th>
<th>Length of training session and frequency</th>
<th>Months since training: Follow-up</th>
<th>Age range of older adults</th>
<th>Control group</th>
<th>Specific effects</th>
<th>Near transfer effects</th>
<th>Far transfer effects</th>
<th>Maintenance effects</th>
<th>Specific near far Transfer effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brehmer and colleagues (2012)</td>
<td>Verbal and non-verbal WM tasks</td>
<td>25 session, 5 weeks</td>
<td>3 months</td>
<td>60–70 ($M = 63.9 \pm 3.4$)</td>
<td>Active: Same WM training task with no variation in the task difficulty</td>
<td>Yes</td>
<td>Yes</td>
<td>Mixed: To inhibition, sustained attention, and self-rating on cognitive failures but Not to fluid intelligence and episodic memory tasks</td>
<td>Yes for the significant specific and transfer effects found</td>
</tr>
<tr>
<td>Richmond and colleagues (2011)</td>
<td>Verbal and spatial WM tasks</td>
<td>20 sessions, 30 min each</td>
<td>NE</td>
<td>60–80 ($M = 66 \pm NR$)</td>
<td>Active: Trivia training</td>
<td>Yes</td>
<td>Yes: To a verbal WM task but Not to short-term memory tasks (the forward and backward Digit span tests)</td>
<td>Mixed: For repetition for a short- and a long-term verbal and memory task but Not to fluid intelligence, short-term memory, attention tasks</td>
<td>NE</td>
</tr>
</tbody>
</table>

Notes: WM = working memory; CWMS = Categorization Working Memory Span; NR = not reported; NE = not evaluated.

*Participants performed pre- and post-test tasks only; they were, therefore, not involved in any alternative activities, whereas the trained group attended the training sessions.*
which may stimulate participants’ cognitive flexibility and plasticity, as well as their interest, since the requirements of the training task are always novel and challenging. All these aspects may also help to motivate older participants, who feel that they can handle the task, keeping them interested in the training activities (see also Hertzog & Hultsch, 2000; Schmidt & Bjork, 1992). The training schedule, in which sessions were arranged with a fixed interval between them, may be another feature that made the training more successful because it left participants sufficient time to consolidate the skills they acquired while, at the same time, it reduced the risk of their losing any beneficial effects of having practiced with the task (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006).

It should nonetheless be noted that the positive and encouraging results seen with this training, as in Brehmer and colleagues (2012), might also depend on factors other than the characteristics of the training, such as the participants’ age. Indeed, unlike previous WM training studies, Table 1 shows that the sample considered did not include only young-old adults (as in Brehmer et al., 2012, or in Borella et al., 2010), but also individuals over 75 years of age (see Buschkuehl et al., 2008; Li et al., 2008), that is, old-old adults. In some other studies (Table 1), although the participants’ mean age coincided with the young-old age group, the range of ages was rather broad and the sample included some old-old adults too (Richmond et al., 2011; Schmiedek et al., 2010). It may be that including old-old adults affected the magnitude and pattern of the training gains because old-old individuals experience a more accentuated cognitive decline than the young-old (see e.g., Borella, Carretti, Cornoldi, & De Beni, 2007). Fluid abilities decline linearly with increasing age (e.g., Park et al., 2002), and crystallized abilities start to decline beyond the age of 74 (e.g., Schaie & Willis, 1996; Singer, Verhaeghen, Ghisletta, Lindenberger, & Baltes, 2003). Meanwhile, though cognitive and neural plasticity mechanisms can be stimulated in old age by training and new learning (Cabeza, 2002; Greenwood & Parasuraman, 2010; Park & Reuter-Lorenz, 2009), these mechanisms appear to be less effective in very late adulthood (Schmiedek et al., 2010). These changes may therefore affect the extent to which old-old adults can benefit from training, as suggested in the meta-analysis by Verhaeghen and colleagues (1992), which reported that training benefits correlated negatively with age. Therefore, the few or no transfer and maintenance effects reported in most WM training studies may also be explained by the characteristics of the sample involved and their age in particular.

The aim of the present study was to assess in a sample of old-old adults (between 75 and 87 years of age) the efficacy of the verbal WM training schedule already tested by Borella and colleagues (2010) in young-old adults. In particular, we wished to examine whether this training can improve WM performance in the old-old, and to what extent, it can elicit and maintain both near and far transfer effects on tasks other than those trained directly.

The Borella and colleagues training procedure was chosen because: (a) to our knowledge, it is the only WM training study to have shown such promising results; (b) the same training procedure has been used in other studies, showing similar benefits in both normal (Carretti, Borella, Zavagnin, & De Beni, 2012) and pathological aging—older adults with amnestic mild cognitive impairment-(Carretti, Borella, Fostinelli, & Zavagnin, 2013); and (c) it includes only a few sessions (three), and this could make it more suitable for old-old adults, because it favors their motivation and engagement, and—should it prove being effective—it could be of a considerable practical value for clinicians who have to cope with time constraints.

To test the efficacy of the training, the old-old in our study were assigned to either a training group or an active control group, and both were administered a battery of tests assessing processes implicated in WM, such as visuospatial WM, short-term memory, reasoning, inhibition, and processing speed, to analyze changes in their performance after the training and at the follow-up assessment (after 8 months).

Consistently with previous results, we expected the training to improve the old-old participants’ performance in the criterion task (which resembled the trained one), and we expected this gain to be maintained over time. In particular, we expected to obtain a large effect size in the practiced task in line with the results of previous training studies, as, for example, documented by the meta-analysis by Melby-Lervåg and Hulme (2013). In fact, most of the studies measuring the increase in performance in the practiced task (i.e., the so-called “criterion task”) reported large effect size; this is true also in the case of old-old adults (e.g., Li et al., 2008).

Transfer and maintenance effects were expected to be limited in magnitude compared with those seen in the young-old participants due to the more accentuated cognitive decline and more limited cognitive plasticity characteristic of very old age.

**Method**

**Participants**

Participants were recruited at the University of the Third Age or at social clubs in Padova. Plenary sessions were organized to recruit participants. The older adults who attended these sessions were told about the opportunity to take part in a study involving two different programs lasting five sessions. It was explained that participants in one of the programs would be presented with memory tasks and be asked to practice with them, whereas in the other, the participants’ would reflect on several
aspects of memory (e.g., autobiographical memory) after filling in some questionnaires. Our inclusion criteria were: an age of 75 or older; a Mini-Mental State score (Folstein, Folstein, & McHugh, 1975) over 27; Italian mother tongue; and the ability to make the necessary time commitment. Our exclusion criteria (based on a semi-structured interview) were: a history of psychiatric or neurological diseases, or brain fever such as encephalitis or meningitis; any use of benzodiazepines in the previous 3 months; symptomatic cardiovascular conditions, breathing problems, or diseases capable of causing cognitive impairments; visual, auditory, and/or motor impairments (see Crook et al., 1986).

Among the 40 who volunteered for the study, 36 fulfilled the inclusion criteria; the four who did not take part in the control group’s activities but their data were not considered. This was done to obtain a more homogeneous (albeit smaller) sample size, in terms of participants’ demographics and cognitive status, in line with other training studies on older adults. The sample size was nonetheless sufficient to obtain a large effect size with an estimated power of .95 (computed with G power 3.1.3, Faul, Erdfelder, Lang, & Buchner, 2007).

Each of the 36 participants was given a sequential study identification number, then they were randomly assigned to one of the two groups: 18 to receive training (age range = 75–87; 70% were women) and the other 18 to form the active control group (age range = 75–84; 50% were women).

None of the 36 participants dropped out during the study, and valid results were available for all 36 regarding the pretest assessment, the three-session training, and the post-test and 8-month follow-up assessments.

The two groups did not differ in age ($M_T = 79.22 \pm 3.49$; $M_C = 79.17 \pm 2.95$, $M \pm SD$), $F(1, 36) = 1.89$, $p = .18$; or vocabulary score (Wechsler, 1981; $M_T = 42.72 \pm 9.04$; $M_C = 43.79 \pm 10.37$), $F < 1$, $p = .96$. Moreover, their level of formal education (see ISTAT, 2011) and vocabulary scores can be considered representative of the Italian population of old-old adults.

**Design**

The study can be considered a single-blind study because only participants were blinded to the treatment condition.

**Material**

**Pretest, Post-Test, and Follow-Up Assessment**

Before and after the training, and at the 8-month follow-up, all participants were examined using the same battery of tests to identify any training-related gains in performance in the criterion task (the Categorization Working Memory Span [CWMS] task) practiced during the three-session training and in the tasks for ascertaining any transfer effects. Performance in the criterion task was evaluated in terms of both correct recall and intrusion errors. We included this latter measure because it is often considered when assessing one of the inhibitory functions, that is, the ability to dampen the activation of no longer relevant items (see Carretti, Mammarella, & Borella, 2012; Friedman & Miyake, 2004), so analyzing this additional score could shed light on the mechanisms enhanced by the training.

Transfer effects were assessed using the same tasks as in Borella and colleagues (2010), chosen for their theoretical relationship with WM. In particular, the transfer tasks were classified, as proposed by Noack, Lövdén, Schmiedek, and Lindenberger (2009), along a continuum from nearest to far transfer tasks. Thus, a visuospatial WM task, the Dot Matrix Transfer task (adapted from Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001), was used to assess nearest transfer effects because it measures WM but with a different task content from the training task, while for near transfer effects the Forward and Backward Digit Span tests were used. Those tasks were considered short-term memory measures as they differ in their requirements from the classical WM tasks as shown by Bopp and Verhaeghen (2005). Tasks assessing reasoning ability (Cattell test), processing speed (Pattern Comparison task), and inhibition (Stroop Color task) were used to measure far transfer effects, as these abilities/mechanisms have been shown to be related to WM or, when processing speed and inhibition are taken into account, to explain the age-related decline in WM (see Borella, Ghisletta, & de Ribaupierre, 2011).

Parallel versions of each task (with different stimuli) were used for the pre- and post-test assessments, counterbalanced across testing sessions; the pretest versions were used again for the follow-up assessment.

**Criterion Task**

**CWMS test.** This task (Borella et al., 2008) is similar to the classic WM tasks, such as the Listening Span test (Borella et al., 2008). As described below, it requires to maintain information in active memory (maintenance phase) while simultaneously performing distracting or interfering activities (processing phase) (e.g., Daneman & Carpenter, 1980). Lists of words have
to be processed in the CWMS (rather than sentences as in the Listening Span test), and this limits the role of semantic processing. The material consisted of a series of lists containing five medium-high frequency words (from the thesaurus of Italian words provided by Barca, Burani & Arduino, 2002) organized into sets of different lengths (from two to six lists of words). Each series contained no, one, or two animal nouns in any position, including the last. Out of a total number of 200 words in the task, 28% were animal nouns. To give an example, in a trial with two words to recall as follows:

- house, mother, dog, word, night;
- flower, pen, tiger, chair, shoe;
- participants had to tap their hand for “dog” and “tiger”, and remember “night” and “shoe.”

Participants listened to audio-recordings of the lists in which words were presented at a rate of 1 s per word, and they were asked to tap with their hand on the table whenever they heard an animal noun (processing phase). At the end of the set of lists, participants had to recall the last word in each list in serial order (maintenance phase). The interval between the word lists was 2 s. Two practice trials containing two words to remember were administered before the experiment started and, if these practice trials were not answered correctly, they were repeated to ensure that the participant had understood the instructions correctly.

To make sure that participants were not trading-off between processing the animal nouns and remembering the last words in the lists, an 85% accuracy was required on the judgment task: during the administration of the task, the experimenter ticked on a dedicated form if participants tapped the table whenever an animal noun occurred. All participants satisfied this criterion.

The total number of correctly recalled words was considered as the measure of WM performance (maximum score 20). This task has been shown to correlate strongly with classical verbal and visuospatial WM tasks (see Borella et al., 2008; De Beni, Borella, Carretti, Marigo, & Nava, 2008), fluid intelligence (Borella, Carretti, & Mammarella, 2006), and complex cognitive abilities, such as reading comprehension (De Beni, Borella, & Carretti, 2007). The number of intrusion errors, relating to words not in the last position that were recalled by participants, was also computed. This measure is often considered as representing a participant’s ability to inhibit no longer relevant information (see, e.g., Borella et al., 2007).

Half (five) of the sets were used for the pretest session, the other five at the post-test session, and the two sets were counterbalanced across test sessions.

**Tasks Measuring Transfer Effects**

**Nearest transfer effects: Visuospatial WM task**

Dot Matrix task (adapted from Miyake et al., 2001). This task (adapted from Miyake et al., 2001) involves participants verifying a matrix equation consisting of either an addition or a subtraction, presented as lines drawn on a 3 × 3 matrix, and then memorizing sequences of dots presented on a 5 × 5 grid. Participants were given a maximum of 4.5 s to check each equation and say “True” or “False.” Immediately after they gave their answer, they were presented for 3 s with a 5 × 5 grid containing a dot in one of the squares. After seeing sets of two to six equation–grid pairs, they had to indicate the positions of the dots on a blank 5 × 5 grid. There was one practice trial with two equations, each with one dot. The number of dot locations to recall increased from two to six. A total of 28 equations and 28 matrices were presented.

The total number of dot positions correctly recalled was considered as the dependent variable (maximum score 14). Half (14) of the sets were used at the pretest session and the others at the post-test session, counterbalancing them across the testing sessions.

**Near transfer effect**

**Forward and Backward Digit Span tasks.** In this task (e.g. De Beni et al., 2008) of digits were presented at a rate of 1 s per digit, and participants were asked to repeat them in the same (forward) or reverse (backward) order. The series started with three digits and rose to nine for the forward task and ranged from two to eight for the backward task. For each level of difficulty, there were two strings of digits. After two consecutive recall errors, the task was discontinued. A practice sequence with two digits was administered for each task before the test started. One point was awarded for each sequence recalled correctly.

The final score corresponded to the total number of correct trials recalled correctly (maximum score 14 for both tasks). Two versions of each task were created, exchanging the digit strings within each level of difficulty; one was administered at the pretest session and the other at the post-test session, in counterbalanced fashion across the testing sessions.
Far transfer effects

**Culture Fair test.** Scale 3 of the Cattell test (Cattell & Cattell, 1963) consists of two parallel forms (A and B), each containing four subtests to be completed in 2.5–4 min, depending on the subtest presented. Participants were asked to complete a series of abstract shapes/figures (Subset 1), solve a series of problems (Subset 2), or establish the relationship between abstract figures, lines, and so on (Subset 3). The dependent variable was the number of items answered correctly across the four subsets (maximum score 50).

One of the two parallel forms (A or B) was administered at the pretest session and the other at the post-test session, in counterbalanced fashion across the testing sessions.

**Stroop Color task.** The task (adapted from Trenerry, Crosson, De Boe, & Lever, 1989) consists of six cards that list: names of colors printed in incongruent colors (Incongruent condition), names of colors printed in congruent colors (Congruent condition), and color patches (Control condition). There were two cards for each condition. Participants were asked to name the color of each stimulus and process the stimuli as fast as possible while also being as accurate as possible.

The interference effect was calculated as the relative difference to adjust for baseline individual differences (see Ludwig, Borella, Tettamanti, & de Ribaupierre, 2010), between the Incongruent and Control conditions calculated in terms of time and errors, as follows: \[ \frac{(incongruent \text{ condition} - \text{control condition})}{\text{control condition}} \]. A higher score thus implied greater difficulty in controlling the prepotent response in the incongruent condition.

Two versions of the task were created by inverting the card order within each condition; one was administered at the pretest session and the other at the post-test session, in counterbalanced fashion across the testing sessions.

**Pattern Comparison task.** In this task (adapted from Salthouse & Babcock, 1991), participants were asked to decide whether arrangements of line segments, presented on two pages, were identical or not. The experimenter used a stopwatch to record the time it took to complete each page. Three practice trials were run before the experiment started. The dependent variable was the total time taken to provide the answers for the two pages (a modified version of the Pattern Comparison task was used in this study, consisting of one sheet of patterns to compare).

Two versions of the task were created by inverting the page order; one was administered at the pretest session and the other at the post-test session, in counterbalanced fashion across the testing sessions.

Parallel versions of each task (using different stimuli) were thus used for the pre- and post-test sessions and counterbalanced across testing sessions, whereas the pretest versions were re-used for the follow-up session.

**Procedure**

All participants were tested individually at the pretest (Session 1), post-test (Session 5), and follow-up sessions (8 months later). The order of presentation of the tasks at these three sessions remained the same for all participants in both the trained and the control groups, that is, Vocabulary (at pretest only), Forward Digit Span, Backward Digit Span, Pattern Comparison (paper and pencil), CWMS (auditory), Stroop Color (paper), Dot Matrix (computerized), and Cattell (paper and pencil) tests. Sessions 2–4 were completed within a 2-week timeframe with a fixed 2-day break between sessions, and the trained participants received training while the control group completed alternative activities (see below).

The schedule was identical for the two groups, enabling the amount of social interaction to be matched. The experimenter used standardized instructions to describe the rationale and benefits of both programs to keep participants blinded to the goals of the study.

All sessions were individual, lasted about 60 min, and were guided by the experimenter, who explained the activities involved in each session and presented the materials. Sessions ended with the experimenter asking participants how they felt about the activities just completed and reminding them about the date of the following meeting.

To limit the influence of sensory variables (sight and hearing) on outcomes, auditory presentations were adjusted to participants’ hearing levels. For the paper-and-pencil tasks, all participants were also asked whether they found the stimuli easy to read.

**Trained group.** During the three training sessions (Sessions 2–4), participants practiced with a modified version of the CWMS. The experimenter presented participants with lists of audio-recorded words organized in the same way as for the original CWMS task. The basic instructions were the same, that is, participants had to tap with their hand on the table whenever an animal noun was heard and recall target words, but the amount of information to recall and the processing and maintenance requirements were manipulated by comparison with the original CWMS task.
The training involved an adaptive procedure, during which the level of task difficulty was adjusted to participants' performance, increasing the number of words to recall if they were successful, and presenting the lowest memory load if they were not (Session 2), with constant variations in the secondary task requirement, that is, asking participants to recall: (a) the last or the first word in each series (Sessions 2 and 4) or (b) words followed by a “beep” (Session 3) (as presented below). The processing requirement (tapping when an animal noun was heard) was also manipulated by varying the frequency of these animal nouns in the lists (Session 3). Participants were not instructed to use specific strategies and no feedback was provided in order to limit the use of task-specific strategies and thus favor transfer effects.

In Session 2, participants were presented with a series of word lists, for a total of 128 words grouped into sets of different length (from 2 to 5), each with three series of word lists, always starting from the lowest memory load level (i.e., two words). The task was organized in three phases: in the first, the experimenter instructed participants to recall the last word of each list of the series, then the first word in the second phase, and the last word again in the third phase. If participants performed well in a given phase (correctly recalling the words in two of the three series), the difficulty of the task was increased. When the highest level was reached (five words to recall), participants started with the lowest level in the following phase. If they failed to answer correctly, the level of difficulty was not increased further and the following phase started from the lowest level (two words to recall). The experimenter stopped administering the task when a participant failed to recall the last words in two of the three series in phase 3.

In Session 3, participants were asked to recall words followed by a beep, and the frequency of the secondary task (tapping their hand on the table when an animal noun was heard) was manipulated. Sets of different length (from 2 to 5), each consisting of four series of word lists, were presented, for a total of 280 words. Word series of length 2 could contain 2–8 animal nouns, those of length 3 contained 4–9, those of length 4 contained 6–11, and those of length 5 contained 8–17. Participants were asked to remember each word followed by a beep, in serial order. The difficulty of the task was not adjusted to participants’ performance.

In Session 4, participants had to recall words alternately in the last and first positions in a list (the last words of each list in the first series, the first in the second, the last in the third, and the first in the fourth series). Sets of different lengths (from 2 to 5) were presented, each consisting of four series of two word lists (for a total of 280 words); the memory load of the task increased regardless of the participant’s performance.

**Control Group**

In Session 2, participants in the control group were asked to rate the vividness of events they could remember from their childhood and adulthood or recent events (autobiographical memory questionnaire). In Session 3, they recorded the frequency of their behavior dedicated to saving memories of life events (memory sensitivity questionnaire). In Session 4, they rated their personal satisfaction with their life, emotional competence, and coping strategies regarding everyday problems (well-being questionnaire). All these paper-and-pencil questionnaires were drawn from a standardized Italian battery for assessing well-being and memory in adulthood (De Beni et al., 2008).

**Results**

The two groups’ pretest performance was compared first with one-way ANOVA conducted separately for each of the baseline measures, with Group as the independent variable. Results showed no significant differences between the groups’ baseline performance (CWMS task, $F_{\text{s}} < 1$; Dot Matrix task, $F_{\text{s}} = 1.08$; Forward Digit Span test, $F_{\text{s}} < 1$; Backward Digit Span test, $F_{\text{s}} = 1.50$; Cattell test, $F_{\text{s}} < 1$; Pattern Comparison test, $F_{\text{s}} < 1$; Stroop Color interference index on response times, $F_{\text{s}} < 1$; Stroop Color incongruent errors, $F_{\text{s}} < 1$; intrusion errors in the CWMS, $F_{\text{s}} = 1.37$).

Then, to assess the effects of training, 2 (Group: trained, control) × 3 (Session: pretest, post-test, follow-up) mixed-design ANOVA was run for the measures of interest. Significant main effects and interactions were analyzed using pairwise comparisons, with Bonferroni’s adjustment for multiple comparisons. The $\alpha$-value was set at 0.05 for all statistical tests and at 0.01 (for interactions, the $\alpha$-value was set at 0.01 because nine comparisons were conducted [0.05/9 = 0.006]) for the interactions.

Descriptive statistics are given in Table 2, and the results of the ANOVA are summarized in Table 3.
The results revealed main effects of group, with trained participants recalling more words correctly than controls ($M_{\text{Diff}} = 2.44, p = .007$), and of session, with performance at post-test and follow-up (which did not differ from one another) differing significantly from pretest performance ($M_{\text{Diff}} = -2.11, p = .001; M_{\text{Diff}} = -1.72, p = .001$, respectively).

The group $\times$ session interaction was significant. Post hoc comparison showed that trained participants performed better at the post-test ($M_{\text{Diff}} = 3.72, p = .002$) and follow-up ($M_{\text{Diff}} = 3.44, p = .001$) than at the pretest sessions. This group’s performance did not differ significantly between the post-test and follow-up sessions. No significant differences emerged across the sessions for the control group. The trained group significantly outperformed the control group at both the post-test ($M_{\text{Diff}} = 3.44, p = .002$) and follow-up sessions ($M_{\text{Diff}} = 3.67, p = .001$).

As for the intrusion errors in the CWMS task, there were significant main effects of group, with the trained group making fewer intrusion errors than controls ($M_{\text{Diff}} = -1.83, p = .011$), and also of session, since the intrusion errors were more numerous in the pretest session than at the follow-up ($M_{\text{Diff}} = 0.97, p = .01$). The group $\times$ session interaction was significant too. Post hoc comparisons indicated that the trained group’s intrusion errors decreased from the pretest to the follow-up ($M_{\text{Diff}} = 2.17, p = .001$) and to a marginally significant level from the post-test to the follow-up sessions ($M_{\text{Diff}} = 1.56, p = .01$). The control group’s performance did not differ significantly between sessions. The trained group made fewer intrusion errors than controls, but only at the follow-up stage ($M_{\text{Diff}} = -3.39, p = .001$).

Nearest Transfer Effect

For the Dot Matrix task, the results for the total number of dot positions recalled correctly showed main effects of group, with the trained group performing better than the controls ($M_{\text{Diff}} = 1.40, p = .02$), and of session, with participants performing better at the post-test than at the follow-up ($M_{\text{Diff}} = 1.11, p = .02$); pretest performance was not statistically different from the other two sessions. The group $\times$ session interaction was not significant.

Near Transfer Effect

In terms of the total number of correct answers given in the Forward Digit Span test, the main effect of group was not significant. The main effect of session indicated that performance was only better at the post-test session compared with the pretest ($M_{\text{Diff}} = 0.53, p = .001$).

The group $\times$ session interaction was significant. Post hoc comparisons indicated that only the trained group showed a consistent improvement in their performance from the pre- to the post-test ($M_{\text{Diff}} = 0.89, p = .001$), and a marginally significant difference was found from the pretest to the follow-up sessions ($M_{\text{Diff}} = 0.72, p = .01$); performance in the post-test and follow-up sessions did not differ statistically. The control group’s performance did not differ significantly between any of the sessions. The trained group’s performance was marginally better than the control group’s at the follow-up session ($p = .09$).

In the Backward Digit Span test, the main effects of group and session and the group $\times$ session interaction were not significant.
Far Transfer Effect

The main effects of group and session and the group × session interaction were not significant for the Cattell test or the Pattern Comparison task.

As for Stroop Color, in the interference index on response times − [(incongruent condition − control condition)/ control condition] − the main effects of group and session and the group × session interaction were not significant. (Non-parametric [Mann–Whitney] tests confirmed the ANOVA results for both errors and response time [RT] indexes. Concerning errors, the trained group differed significantly from the controls only at follow-up, Z = −2.96, p < .01. No significant difference was found between groups in terms of the RT indexes.)

The main effects of group and session were again not significant when the number of errors in the incongruent condition was examined. The group × session interaction was significant, however. Post hoc comparisons indicated that only the trained group made fewer mistakes at the follow-up than at the pretest stage (MDiff = −3.00, p = .001). For this group, a marginally significant difference was found between the post-test and pretest sessions (MDiff = −2.39, p = .01. Finally, the post-test and follow-up results did not differ to a statistically significant degree. No significant differences were found between the sessions for the controls. The trained group made significantly fewer errors than the control group, but only at the follow-up session (MDiff = −3.22, p = .001).

Table 3. Mixed-design 2 × 3 ANOVA results for the measures of interest, with group (trained, control) as between-subjects factor and session (pretest, post-test, follow-up) as repeated measures

<table>
<thead>
<tr>
<th>Specific effect</th>
<th>F</th>
<th>df</th>
<th>MSE</th>
<th>n²p</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects: Group (G)</td>
<td>8.27</td>
<td>1, 34</td>
<td>3.81</td>
<td>0.20</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Within subjects: Session (S)</td>
<td>17.60</td>
<td>2, 68</td>
<td>2.58</td>
<td>0.34</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>G × S</td>
<td>12.95</td>
<td>2, 68</td>
<td>2.58</td>
<td>0.28</td>
<td>.001</td>
</tr>
<tr>
<td>CWMS intrusion errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects: Group (G)</td>
<td>7.33</td>
<td>1, 34</td>
<td>12.37</td>
<td>0.18</td>
<td>.01</td>
</tr>
<tr>
<td>Within subjects: Session (S)</td>
<td>4.04</td>
<td>2, 68</td>
<td>2.12</td>
<td>0.11</td>
<td>.02</td>
</tr>
<tr>
<td>G × S</td>
<td>7.72</td>
<td>2, 68</td>
<td>2.12</td>
<td>0.18</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Transfer effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dot Matrix</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects: Group (G)</td>
<td>6.32</td>
<td>1, 34</td>
<td>8.24</td>
<td>0.16</td>
<td>.02</td>
</tr>
<tr>
<td>Within subjects: Session (S)</td>
<td>4.12</td>
<td>2, 68</td>
<td>2.87</td>
<td>0.11</td>
<td>.02</td>
</tr>
<tr>
<td>G × S</td>
<td>0.94</td>
<td>2, 68</td>
<td>2.87</td>
<td>0.04</td>
<td>.40</td>
</tr>
<tr>
<td>Digit Span Forward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects: Group (G)</td>
<td>0.88</td>
<td>1, 34</td>
<td>3.79</td>
<td>0.02</td>
<td>.35</td>
</tr>
<tr>
<td>Within subjects: Session (S)</td>
<td>4.61</td>
<td>2, 68</td>
<td>0.55</td>
<td>0.12</td>
<td>.013</td>
</tr>
<tr>
<td>G × S</td>
<td>3.36</td>
<td>2, 68</td>
<td>0.55</td>
<td>0.09</td>
<td>.04</td>
</tr>
<tr>
<td>Digit Span Backward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects: Group (G)</td>
<td>2.72</td>
<td>1, 34</td>
<td>3.71</td>
<td>0.07</td>
<td>.11</td>
</tr>
<tr>
<td>Within subjects: Session (S)</td>
<td>0.69</td>
<td>2, 68</td>
<td>0.85</td>
<td>0.02</td>
<td>.51</td>
</tr>
<tr>
<td>G × S</td>
<td>0.39</td>
<td>2, 68</td>
<td>0.85</td>
<td>0.01</td>
<td>.68</td>
</tr>
<tr>
<td>Cattell test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects: Group (G)</td>
<td>0.13</td>
<td>1, 34</td>
<td>55.13</td>
<td>0.00</td>
<td>.72</td>
</tr>
<tr>
<td>Within subjects: Session (S)</td>
<td>1.20</td>
<td>2, 68</td>
<td>3.47</td>
<td>0.03</td>
<td>.31</td>
</tr>
<tr>
<td>G × S</td>
<td>0.65</td>
<td>2, 68</td>
<td>3.47</td>
<td>0.02</td>
<td>.53</td>
</tr>
<tr>
<td>Stroop Color interference index (RT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects: Group (G)</td>
<td>0.52</td>
<td>1, 34</td>
<td>0.46</td>
<td>0.02</td>
<td>.48</td>
</tr>
<tr>
<td>Within subjects: Session (S)</td>
<td>2.13</td>
<td>2, 68</td>
<td>0.10</td>
<td>0.06</td>
<td>.13</td>
</tr>
<tr>
<td>G × S</td>
<td>0.93</td>
<td>2, 68</td>
<td>0.10</td>
<td>0.03</td>
<td>.40</td>
</tr>
<tr>
<td>Stroop Color incongruent errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects: Group (G)</td>
<td>1.21</td>
<td>1, 34</td>
<td>34.40</td>
<td>0.03</td>
<td>.28</td>
</tr>
<tr>
<td>Within subjects: Session (S)</td>
<td>2.78</td>
<td>2, 68</td>
<td>6.44</td>
<td>0.08</td>
<td>.07</td>
</tr>
<tr>
<td>G × S</td>
<td>5.59</td>
<td>2, 68</td>
<td>6.44</td>
<td>0.14</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Pattern Comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects: Group (G)</td>
<td>0.25</td>
<td>1, 34</td>
<td>4621.46</td>
<td>0.00</td>
<td>.62</td>
</tr>
<tr>
<td>Within subjects: Session (S)</td>
<td>0.95</td>
<td>2, 68</td>
<td>533.02</td>
<td>0.03</td>
<td>.39</td>
</tr>
<tr>
<td>G × S</td>
<td>0.86</td>
<td>2, 68</td>
<td>533.02</td>
<td>0.03</td>
<td>.43</td>
</tr>
</tbody>
</table>

Notes: CWMS = Categorization Working Memory Span test; RT = response time.
To gain a better understanding of the extent of the benefits of training (between the pre- and post-test sessions), the effect size was computed with $d$ Cohen’s (1988), which expresses the effect size of comparisons, on the within-subject data using the methods proposed by Morris and DeShon (2002). Comparing the pre- to post-test gains within each group revealed in the trained group a large effect size (over 0.80) on the criterion task and in the Forward Digit Span task, a medium effect size for errors in the incongruent condition of the Stroop Color test, and a small effect size for the other measures. Only small effect sizes were seen in the control group (Fig. 1).

When the pretest and follow-up data were compared, the effect sizes for the trained group were large for the criterion task ($d = 1.44$), the Stroop Color errors in the incongruent condition ($d = 0.85$) and the intrusion errors in the CWMS ($d = 0.95$), and medium for the Forward Digit Span task ($d = 0.61$).

Comparisons Between Old-Old (Current Study) and Young-Old Adults (previous published study)

To understand the dimension of the benefit of WM training in relation to age differences in elderly people, we compared the present findings with those obtained in young-old adults (Borella et al., 2010) in an effort to answer the question: do the old-old gain as much from the training as young-old individuals?

The dimension of the effects seen in the current study was thus compared with the data in Borella and colleagues (2010) (Table 4). Since our aim was to examine both transfer and maintenance effects, the $d$ index was calculated for the pre-/post-test data in Borella and colleagues (2010) too, using the $n$, mean, and standard deviations reported therein. (We only calculated the $d$-values for the effect sizes of the maintenance effects reported in Borella and colleagues (2010). Provided that Borella and colleagues did not consider the scores for the intrusion errors in the CWMS task, no comparison could be conducted between the two studies for this measure).

Table 4. Benefits of WM training (expressed in Cohen’s $d$ values): Pretest versus post-test sessions in young-old (Borella et al., 2010) and old-old (present study)

<table>
<thead>
<tr>
<th></th>
<th>Young-old</th>
<th>Old-old</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWMS</td>
<td>2.25</td>
<td>1.40$^a$</td>
</tr>
<tr>
<td>Dot Matrix</td>
<td>1.70</td>
<td>0.47</td>
</tr>
<tr>
<td>Forward Digit Span</td>
<td>2.24</td>
<td>1.07$^a$</td>
</tr>
<tr>
<td>Backward Digit Span</td>
<td>2.35</td>
<td>0.09</td>
</tr>
<tr>
<td>Cattell</td>
<td>1.40</td>
<td>0.15</td>
</tr>
<tr>
<td>Stroop Color test interference index (RT)</td>
<td>0.85</td>
<td>$-0.61$</td>
</tr>
<tr>
<td>Stroop Color test incongruent errors</td>
<td>0.86</td>
<td>0.65$^a$</td>
</tr>
<tr>
<td>Pattern comparison$^2$</td>
<td>0.99</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Notes: CWMS = Categorization Working Memory Span test; RT = response time.

$^a$These changes were statistical significant in the ANOVA analyses.
Concerning the immediate benefit of training for the criterion task, the $d$-value was larger for the young-old than for the old-old. As for transfer effects, the $d$-values were large or medium for the young-old, but generally weak or nil for the old-old, except for the Forward Digit Span.

As for any maintenance effects, $d$-values were larger for young-old ($d = 2.01$) than for old-old ($d = 1.44$) in the criterion task. No significant transfer effects were seen in either of the two samples, except in the Cattell and processing speed tasks, but only in the younger age group. On the other hand, only the old-old group showed a significant, medium effect size ($d = 0.61$) in the Forward Digit Span test.

**Discussion**

WM is one of the cognitive mechanisms that decline linearly in aging. The age-related decline in WM is also supported by the more accentuated decline in the frontoparietal regions (crucial to WM) than in other brain regions. Developing procedures to improve WM performance through training regimes and assessing related benefits in other abilities is therefore of interest especially for its potential practical implications, with a view to improving a skill that shares processes with different higher cognitive abilities influencing people’s everyday activities, such as reading comprehension (Daneman & Merikle, 1996) and reasoning (Kane et al., 2004), consequently improving their quality of life.

Given the more accentuated cognitive decline characterizing old-old (adults aged 75 and over), the present study aimed to analyze the efficacy of a WM training program in terms of specific gains, transfer and maintenance effects in a sample of old-old people. In particular, we adopted a WM training program that, in contrast with available studies on WM training for older adults, had already proved effective in terms of training gains and their maintenance in young-old adults (Borella et al., 2010; see also Carretti et al., 2012) and in other sample populations (amnesic Mild Cognitive Impairment - a; Carretti et al., 2013). This was done to shed light on the effect of older age on the efficacy of the training.

Following the recommendations of Shipstead, Rinck, and Engle (2010), the trained group’s performance was compared with that of an active control group, who met the experimenter for the same number of times but did not practice with the WM tasks. Participants were randomly assigned to the training or control groups, preventing the influence of factors such as individual differences in cognitive and non-cognitive variables. It also increases the internal validity of the study.

As expected on the basis of previous studies, our results demonstrate that verbal WM training had a positive effect, since the trained group performed significantly better in the criterion task (correct recall in the CWMS) afterwards, even at the 8-month follow-up. In addition, trained participants were found to produce fewer intrusion errors at follow-up than controls, suggesting that inhibitory mechanisms are involved in the training-related gains.

The pattern of results in the trained group was not seen in the active control group, so it is reasonable to assume that the benefits seen in the trained participants were attributable mainly to their verbal WM training.

The trained group’s gain in the criterion task is consistent with reports in the literature, showing that WM training programs do produce persistent benefits in the criterion task, even in the oldest participants (e.g., Buschkuehl et al., 2008). Most importantly, the maintenance of the training-related improvement (pretest vs. follow-up) in the criterion task was further emphasized in the comparison with young-old adults, as the two groups showed similar maintenance gains when effect sizes were taken into account. This finding confirms that there is still room for improvement, even in very old age, when the basic mechanisms of cognition such as WM are examined.

It would be of interest to ascertain whether such a behavioral change measured at the cognitive performance level also helps older adults’ brains to build scaffolds in response to such improvements (Park & Reuter-Lorenz, 2009). Judging from some recent models, supported by various studies (though no WM training was involved), the brain adapts and becomes reorganized when stimulated with new learning input and cognitive training, improving its ability to build scaffolds and develop new neural circuitry to compensate for the age-related decline (e.g., Goh & Park, 2009; Greenwood & Parasuraman, 2010). To date, however, the extent to which WM training can improve the plasticity of the neural systems underlying WM training and induce transfer effects is largely unknown (Buschkuehl et al., 2012). Future studies should therefore make an effort to combine analyses on the results of WM training from a cognitive viewpoint with imaging data to identify the neural mechanisms that might account for training-related gains in older adults. This would probably enable new and more effective cognitive interventions to be designed in aging.

When looking at the dimension of the transfer effects, the changes induced by the training seemed less robust, suggesting that more time is needed to modify performance in advanced old age. In fact, the ANOVA results showed that the trained group outperformed the controls in some of the near and far transfer measures, but only at the follow-up session. In particular, when the trained participants were compared with the controls at follow-up, the former: (a) improved significantly in one of the measures of short-term memory and (b) were better able to resist interference, producing fewer errors in the Stroop Color incongruent condition. This was not apparent at the post-test stage. In descriptive terms, however, looking at the effect sizes ($d$ of
Cohen) for differences between the pre- and post-test stages within each group, it is noteworthy that a large effect size was only found for the trained group in the Forward Digit Span test ($d = 1.07$), and a medium one for the errors produced in the Stroop Color incongruent condition ($d = 0.65$). In contrast, the effect size for these measures in the control group was nil ($d = 0.16$ and $0.12$, respectively).

The gains in the Forward Digit Span for the trained group suggest that WM training enabled participants to acquire skills associated with verbal WM rather than on an information-processing level, that is, an enhanced functioning of their cognitive mechanisms and, more generally, of their cognitive resources. This might also explain the lack of transfer effects to visuospatial WM, processing speed, and reasoning (Noack et al., 2009; Schmiedek et al., 2010). At the same time, the reason why there were no gains in other short-term memory measures (see Bopp & Verhaeghen, 2005), that is, the Backward Digit Span test, may be that they require a more active information processing than the forward span test.

The gains in the Stroop Color task (as far as incongruent errors were concerned), along with the decrease in intrusion errors in the CWMS task, suggest that trained participants were also better able to resist irrelevant information. It may be, therefore, that training produced some changes in the way participants allocated their attention and focused on the relevant information. The requirements of the task on which they were trained involve multiple processes, for example, controlling attention, suppressing dominant, and prepotent, but irrelevant responses (Stroop Color incongruent errors), and inhibiting no longer relevant information in WM (intrusion errors). On the other hand, no significant benefit emerged when the Stroop Color index was examined in terms of response times. This latter result was rather unexpected, given the significant reduction in the number of errors in the incongruent condition. It may be that our trained participants focused on accuracy rather than speed, though this is only a speculation because no differences emerged between the groups when the interference index was considered. Caution is needed, moreover, in interpreting the results for the interference index because the RTs used to calculate it were recorded for a series of stimuli and reflect the “sum” of all the trials, including those associated with errors and omissions (see Ludwig et al., 2010), making the index scarcely reliable. Indeed, this is one of the limitations of the Stroop Color card version, as opposed to presenting stimuli one item at a time (as in the computerized version), which enables errors to be separated from response times and thus affords a purer measure of susceptibility to interference (Ludwig et al., 2010). It will be of interest to examine training-related gains in the Stroop Color task using an item-by-item version to clarify the benefit in terms of attentional control when RTs are taken into account.

No other transfer effects were identified so, unlike the picture seen in the previously studied young-old group, the transfer effects seemed to be specific to measures of short-term memory, inhibition of dominant but irrelevant information (Stroop Color test). These effects clearly appear at follow-up, however.

The fact that the changes in the transfer tasks only emerged when performance was compared between the pretest and follow-up sessions may indicate that more practice is needed for trainees to adopt a different approach to coping with the requirements of the task. This hypothesis is supported by the effect size obtained for the trained group and by comparison with the control group, suggesting that some changes were underway already at the post-test stage, but were not robust enough to produce significant differences identifiable by ANOVA. Moreover, these findings are in line with the data from Brehmer and colleagues (2008), who reported an increase in older adults’ performance from the first to the second follow-up session in a training study in which memory strategies were taught. This was not the case for young adults, who maintained their performance but did not improve further. Despite the clear differences between the two studies, the results give the impression that the amount of practice and the opportunity to reactivate contextual information relating to the skill acquired could be important to the efficacy of training in older adults. It is also worth mentioning that the particular measures for which the far transfer effects were found are also the ones that show a less steep decline with aging (see Bopp & Verhaeghen, 2005). Both verbal short-term memory (the Forward Digit Span test) and inhibitory mechanisms show a more accentuated decline only in very late adulthood, as some studies examining the inhibition of no longer relevant information (intrusion errors in a WM task) or resistance to dominant but irrelevant information have clearly demonstrated (see Borella et al., 2008, 2007). In this sense, the finding of transfer effects (albeit limited) could be explained by the similarity of these processes/mechanisms to WM processes and their development curve with aging, prompting the hypothesis of a residual cognitive flexibility that can be enhanced by the training activities considered here.

At the same time, given the greater cognitive decline to which old-old adults are liable (especially as concerns WM and brain regions associated with WM functioning), favoring more generalized transfer gains by means of a more intensive schedule of training sessions and the inclusion of booster sessions (Ball et al., 2002) might be more appropriate. Indeed, the short training procedure used here may suffice for the young-old to show large and persistent transfer gains, but be too “light” for the old-old, who experience a more accentuated decline in their cognitive resources. To induce cognitive plasticity, the old-old adult cognitive system may require a longer period of training. As the young-old have more malleable cognitive skills than the old-old (e.g., Schmiedek et al., 2010), and cognitive plasticity declines progressively with aging (e.g., Lövdén,
It is worth adding, however, that the duration of such training sessions is a matter of debate. WM training studies involving more sessions failed to identify any benefits of training (see Buschkuehl et al., 2008; Li et al., 2008), whereas our brief training schedule not only prompted specific gains, but also hinted at some near and far transfer effects in our old-old participants, and the effects of training persisted after 8 months. Indeed, unlike previous WM training studies, the present findings provide some evidence of both near (short-term memory) and far (inhibition processes) transfer effects and also of the significant benefits obtained being maintained in old-old participants. Although previous studies on old-old people (e.g., Buschkuehl et al., 2008) showed some near transfer effects, as in the present study, they identified no maintenance of any specific, near, and far transfer effects. Li and colleagues (2008) likewise found no signs of far transfer effects being maintained. The minimum amount of training needed to obtain and maintain transfer effects thus remains a crucial issue that should be addressed in future studies.

In future research, it would be interesting to focus on how to improve the present training program effectively, in terms of the number of sessions and the tasks involved, to take the advantage of the old-old adults’ learning potential.

Taken together, and compared with other WM training studies on old-old adults, the present findings suggest that training in which an adaptive procedure is combined with constant variations in the tasks’ requirements has the potential to enhance even old-old adults’ WM performance. With a flexible and variable training model such as the one considered here (in which the amount of information to recall and the processing and maintenance demands varied across training sessions), transfer effects also emerged, which is unusual in WM training studies. The presence of the experimenter (unlike the case of other WM training studies) may also have favored these training gains by keeping the participants motivated and ensuring their adherence to the program (e.g., Carretti, Borella, Zavagnin, & De Beni, 2011).

Some limitations of the present study have to be acknowledged, including the use of a single test to examine transfer effects on each of the mechanisms of interest (e.g., processing speed, fluid intelligence) and a small sample size. Concerning the first point, it would be of interest to better focus on transfer effects to shed light on the hierarchical structure of the cognitive tests, adopting multivariate design with multiple measures for each of the mechanisms examined with the transfer tasks, in order to further analyze the “source” of the training benefits identified (e.g., Schmiedek et al., 2010), and to ascertain whether the training could prompt the development of abilities unrelated to a specific task. As for the sample size, our results are based on 36 participants so the effects identified need to be interpreted with caution. The effect sizes were nonetheless consistent with those found in previous training studies on old-old adults. Larger samples would certainly be better and would allow for more sophisticated analyses, also enabling researchers to analyze training gains in terms of individual differences. Our sample size was consistent with those used in other training studies, however, and it is usually hard to conduct training studies on larger samples.

In conclusion, this study demonstrates that the old-old can benefit from verbal WM training: The trained group showed specific gains, and even some signs of transfer effects (not easy to find in the literature; see Table 1) on short-term memory and inhibition, which may well have been limited, but they persisted at the 8-month follow-up. The comparison with gains obtained in younger adults showed that training gains in the old-old are more limited, especially as concerns any transfer and maintenance effects, suggesting that the old-old acquire only a narrow set of skills for completing the verbal WM task in which they received training, with no improvement in their WM capacity per se. Our findings may indicate that the potential gains of training vary as a function of the trainees’ age and confirm that cognitive interventions are more effective in earlier than in late adulthood. Although the potential benefits are more evident in young-old than in old-old, our results show that there is still room for improvement even in the old-old—as clearly shown by the 8-month follow-up maintenance effect on the trained task—and that their plasticity is not completely lost (e.g., Buschkuehl et al., 2008).

Future research should be more systematic and should continue to concentrate on the crucial aspects that can promote training benefits in older adults. It should be noted that WM training studies in aging have been very limited and have used different programs and different tasks to assess training gains, and only two studies have focused on old-old adults.

From a practical point of view, the WM training considered here may be of particular interest because it proved capable of improving our old-old adults’ cognitive performance in such a short time (after three sessions), making it potentially valuable in both clinical and educational settings. Future studies could replicate this approach to see whether this promising type of WM training can benefit older adults with cognitive impairments, promote brain plasticity, and thereby attenuate the risk of dementia.

**Funding**

This work was partially supported by two grants from the University of Padova (CPDA087750/08 and CPDA103703/10) to Barbara Carretti and Rossana De Beni, respectively.
Conflict of Interest

None declared.

References


