Aging and memory improvement from semantic clustering: The role of list-presentation format

By: Beatrice G. Kulmann* and Dayna R. Touron

Kuhlmann, B. G. & Touron, D. R. (2016). Aging and memory improvement from semantic clustering: The role of list presentation format. *Psychology and Aging, 31*(7), 771-785.

Abstract:

The present study examined how the presentation format of the study list influences younger and older adults' semantic clustering. Spontaneous clustering did not differ between age groups or between an individual-words (presentation of individual study words in consecution) and a whole-list (presentation of the whole study list at once for the same total duration) presentation format in 132 younger (18–30 years, M = 19.7) and 120 older (60–84 years, M = 69.5) adults. However, after instructions to use semantic clustering (second list) age-related differences in recall magnified, indicating a utilization deficiency, and both age groups achieved higher recall in the whole-list than in the individual-words format. While this whole-list benefit was comparable across age groups, it is notable that older adults were only able to improve their average recall performance after clustering instructions in the whole-list but not in the individual-words format. In both formats, instructed clustering was correlated with processing resources (processing speed and, especially, working memory capacity), particularly in older adults. Spontaneous clustering, however, was not related to processing resources but to metacognitive beliefs about the efficacy and difficulty of semantic clustering, neither of which indicated awareness of the benefits of the whole-list presentation format in either age group. Taken together, the findings demonstrate that presentation format has a nontrivial influence on the utilization of semantic clustering in adults. The analyses further highlight important differences between output-based and list-based clustering measures.

Keywords: recall | cognitive aging | semantic clustering | encoding strategy | utilization deficiency

Article:

* Correspondence concerning this article should be addressed to Beatrice G. Kuhlmann, who is now at the Department of Psychology, School of Social Sciences, University of Mannheim, Schloss EO 240, D-68131 Mannheim, Germany. E-mail: kuhlmann@psychologie.uni-mannheim.de

Given the pronounced decline in episodic long-term memory with even healthy aging (Park et al., 2002), an exciting finding is that older adults (~60+) can substantially improve their memory simply through using effective encoding strategies like mental imagery and semantic organization (see Gross et al., 2012; Verhaeghen, Marcoen, & Goossens, 1992, for meta-analyses). However, several studies suggest that older adults benefit (in terms of memory improvement) less from using encoding-strategies than younger adults (e.g., Kliegl, Smith, & Baltes, 1990; Verhaeghen & Marcoen, 1996) and sometimes even not at all (e.g., Mason & Smith, 1977; Taconnat et al., 2009). Encoding strategy deficiencies may thereby contribute to observed age-related differences in episodic memory. The present study thus examined whether older adults' use of semantic clustering (i.e., the strategy of organizing study words into semantic categories, such as "flowers" and "vehicles") can be supported by the design of the memory task, specifically the presentation format of the study list.

AGE-RELATED ENCODING STRATEGY DEFICIENCIES

There is good evidence that strategy deficiencies contribute to age-related differences in cognitive performance in general (Lemaire, 2010; Touron, 2015) and in episodic memory in particular (e.g., Dunlosky & Hertzog, 1998, 2001; Dunlosky, Hertzog, & Powell-Moman, 2005; Froger, Taconnat, Landré, Beigneux, & Isingrini, 2009; Hulicka & Grossman, 1967; Hulicka, Sterns, & Grossman, 1967; Kausler, 1994; Naveh-Benjamin, Brav, & Levy, 2007). Broadly, such deficiencies can occur at the production and/or utilization stage of strategy implementation (Dunlosky et al., 2005; Dunlosky & Hertzog, 1998; Kausler, 1994): In the case of a *production deficiency*, older adults fail to spontaneously use an effective strategy, which for episodic memory are encoding strategies fostering deep (semantic) elaboration and interconnection of the to-be-studied material (Craik & Lockhart, 1972; Hunt & Einstein, 1981; Richardson, 1998). Several studies have found that older adults spontaneously use effective encoding strategies less frequently than younger adults, both on laboratory memory tasks (e.g., Taconnat et al., 2009; Verhaeghen & Marcoen, 1994) and in daily life (e.g., Bouazzaoui et al., 2010; Hultsch, Hertzog, & Dixon, 1987; Loewen, Shaw, & Craik, 1990). Further, some studies have provided evidence for the contribution of this production deficiency to age-group differences in episodic memory by showing that age-group differences are smaller or even eliminated when encoding strategy use is instructed or otherwise controlled for (e.g., Dunlosky & Hertzog, 2001; Hulicka & Grossman, 1967; Naveh-Benjamin et al., 2007). In the case of a *utilization deficiency*, older adults benefit less from using a strategy than younger adults. In contrast to a pure production deficiency, a utilization deficiency implies that age-group differences in episodic memory actually become larger after instruction to use an encoding strategy. Indeed, several studies document such a magnification of age-group differences in episodic memory after instructions to use an encoding strategy (e.g., Dunlosky & Hertzog, 1998; Kliegl, Smith, & Baltes, 1989; Verhaeghen & Marcoen, 1996).

What causes strategy deficiencies in older adulthood? Notably, older adults generally have spared metamemory knowledge about memory tasks and effective encoding strategies (Hertzog & Hultsch, 2000; Hultsch et al., 1987; but see Price, Hertzog, & Dunlosky, 2008), rendering the observed production deficiencies somewhat surprising. Some researchers have thus attributed older adults' strategy deficiencies to the well-documented age-related declines in general cognitive processing resources, which are assumed to underlie age-related deficits in various higher cognitive functions, including episodic memory (e.g., Park, 2000; Salthouse, 1996). Indeed, spontaneous use of effective encoding strategies in older adults is correlated with processing speed (Hertzog & Dunlosky, 2004; Verhaeghen & Marcoen, 1994), working memory capacity (Wegesin, Jacobs, Zubin, Ventura, & Stern, 2000), and executive functioning (Bouazzaoui et al., 2010; Glisky, Rubin, & Davidson, 2001; Taconnat et al., 2009). Likewise, utilization of an encoding strategy causes greater interference with a secondary task in older than younger adults (Naveh-Benjamin, Craik, Guez, & Kreuger, 2005) and the amount of memory improvement through encoding strategies observed in older adults is correlated with processing speed (Kliegl et al., 1990; Singer, Lindenberger, & Baltes, 2003; Verhaeghen & Marcoen, 1996), working memory capacity (Bryan, Luszcz, & Pointer, 1999; Drevenstedt & Bellezza, 1993), and executive functioning (Bryan et al., 1999).

Despite the good evidence for age-related strategy deficiencies and their relation to age-related declines in general cognitive processing resources, findings are mixed with several studies reporting comparably frequent use of encoding strategies in younger and (healthy) older adults (e.g., Bryan et al., 1999; Kuhlmann & Touron, 2012; Lachman & Andreoletti, 2006), as well as at least comparable, if not larger, memory improvements through encoding strategies in older adults (Dunlosky & Hertzog, 2001; Dunlosky et al., 2005; Froger et al., 2009; Hulicka & Grossman, 1967; Kuhlmann & Touron, 2012; Naveh-Benjamin et al., 2007). Of course, some encoding strategies (e.g., multistep mnemonic techniques like the loci method) may be more cognitively demanding than others. However, there have been inconsistent findings regarding older adults' ability to produce and utilize the same encoding strategy. Such inconsistencies in research findings have rightfully led to the rejection of the hypothesis that strategy deficiencies fully account for age-group differences in episodic memory (e.g., Dunlosky et al., 2005; Light, 1991). However, rather than viewing these inconsistencies as a sign of poor replicability, we believe that they point to the existence of moderating factors in the task environment which either foster or hinder older adults' encoding-strategy use and thereby partially contribute to age-group differences in episodic memory.

TASK AFFORDANCE OF OLDER ADULTS' ENCODING STRATEGY USE

Whereas much research has been devoted to identifying whether there are age-related encoding-strategy deficiencies or not, surprisingly little research has examined which factors influence older (and younger) adults' ability to implement effective encoding strategies. The fact that some studies find older adults to be fully able to spontaneously produce and utilize an encoding strategy whereas others find age-related deficiencies for the *same* strategy suggests that task factors influence older adults' strategy abilities. This implies that it may be possible to systematically manipulate a task's *strategy affordability* and to influence older (and younger) adults' cognitive performance through such a manipulation (Bottiroli, Dunlosky, Guerini, Cavallini, & Hertzog, 2010; Touron & Hertzog, 2004).

Of course, studies differ in many methodological aspects other than the task design. Consequently, to identify strategy-affording task factors, systematic manipulations must be conducted within an experiment using random assignment. Unfortunately, to date few studies have done this with regard to aging and encoding strategies. Perhaps the most attention has been paid to the role of study material characteristics. In particular, older adults' encoding-strategy use is influenced by word concreteness and interrelatedness (Dunlosky & Hertzog, 1998; Howard, McAndrews, & Lasaga, 1981; Hulicka & Grossman, 1967; Rowe & Schnore, 1971; Tournier & Postal, 2011; Witte, Freund, & Sebby, 1990) and effects of encoding-strategy training on older adults' memory performance are more pronounced on tasks with more concrete, everyday materials (Cavallini, Pagnin, & Vecchi, 2003). Recently, Bottiroli et al. (2010) examined the influence of presentation format on older adults' use of self-testing when studying words. They found no age-group differences in spontaneous use of this beneficial study strategy when the words were presented on flashcards, conducive to self-testing, but a production deficiency in older adults when the words were presented affixed to a board. One explanation of how task factors affect older adults' encoding strategy production and utilization is that they may reduce the cognitive demands of that strategy (cf., Bottiroli et al., 2010; see also Touron & Hertzog, 2004). Indeed, a few studies found longer study times to increase older adults' encoding strategy production (Verhaeghen & Marcoen, 1994) and utilization (Kliegl et al., 1990; Verhaeghen & Marcoen, 1996). Relatedly, Derwinger and colleagues found older adults to only improve their memory for numbers through a complex, multistep mnemonic strategy when they were allowed to write down individual steps of the mnemonic, but not when they had to perform the steps in working memory (Derwinger, Stigsdotter Neely, & Bäckman, 2005).

THE PRESENT STUDY: LIST-PRESENTATION FORMAT AND SEMANTIC CLUSTERING

The goals of the present study were to examine the influence of list-presentation format on the (a) spontaneous production as well as (b) successful utilization of semantic clustering in younger and, especially, older adults. We focused on semantic clustering because it is a very effective strategy for episodic memory (particularly recall; e.g., Hunt & Einstein, 1981) for which research regarding age-related differences in its production and utilization has yielded mixed results. When words from various semantic categories are randomly intermixed on the study list, as is the case in typical test inventories using categorizable word material (e.g., the *California Verbal Learning Test*; Delis et al., 1991), the mental organization of study words into categories should be demanding of cognitive processing resources. Indeed, older adults' clustering production is correlated with measures of their working memory capacity (Wegesin et al., 2000), verbal fluency (Jacobs et al., 2001), as well as processing speed and executive functioning (Taconnat et al., 2009). Therefore, it is worthwhile to search for task factors that may facilitate the semantic organization of study words, particularly for older adults.

List-Presentation Format and Semantic Clustering Production

Most aging studies on semantic clustering have focused on spontaneous clustering production. Given the resource demands of clustering it is not surprising that many studies report reduced spontaneous clustering production in older compared to younger participants (Amrhein, Bond, & Hamilton, 1999; Jacobs et al., 2001; Mungas, Ehlers, & Blunden, 1991; Taconnat et al., 2009; Wegesin et al., 2000; Witte, Freund, & Brown-Whistler, 1993), especially under dual task (Fernandes & Grady, 2008; Park, Smith, Dudley, & Lafronza, 1989). Surprisingly, however, these studies are opposed by an approximately equal number of studies, reporting no age-group differences in clustering production (Blatt-Eisengart & Lachman, 2004; Hertzog, Dixon, & Hultsch, 1990; Hess, Auman, Colcombe, & Rahhal, 2003; Kliegl et al., 1990; Lachman &

Andreoletti, 2006; West, Dark-Freudeman, & Bagwell, 2009; West & Thorn, 2001). Upon closer examination, we noted that, although all these studies used randomly intermixed word material, almost all the studies reporting age-group differences (with the exception of Witte et al., 1993) presented the words individually (i.e., one after the other, each for a short time only) for study, whereas those not finding age-group differences (with the exception of Schneider & Uhl, 1990) presented all words at once for study (i.e., by presenting the whole study list on a computer screen or by providing a stack of flashcards). In the former *individual-words format*, semantic clustering should pose great demands on cognitive processing resources because earlier words have to be maintained in (or retrieved into) working memory to be grouped with later study words. In contrast, the *whole-list format*, allows revisiting the earlier parts of the study list, which should greatly decrease cognitive demands of clustering. Thereby, the whole-list presentation format might facilitate the *mental*organization of the words. Jacobs et al. (2001) reported that preblocking the study words by categories (as opposed to randomly intermixing them) increases younger and older adults' clustering use. However, in this case the blocked presentation format explicitly organized the categories for the participant, rather than facilitating use of this strategy as the whole-list presentation format may do.

In the present study, we systematically manipulated list-presentation format, keeping other potentially relevant factors (e.g., word material, total study time), that also varied between the previous studies, constant. We expected both younger and older adults to more frequently use semantic clustering spontaneously in the whole-list than in the individual-words format and we specifically expected this whole-list format benefit to involve a reduction of semantic clusterings' demands on general cognitive processing resources. Therefore, we expected weaker (or no) correlations between semantic clustering and resource measures (processing speed, working memory capacity) in the whole-list format but strong correlations in the individual-words format. Given older adults' reduced cognitive processing resources, we expected them to particularly benefit from the whole-list format such that there would be an age-related deficit in semantic clustering production in the individual-words but not in the whole-list format. Importantly, because metamemory beliefs may also influence spontaneous strategy use (see Hertzog & Dunlosky, 2004, for an overview), we also assessed general metamemory beliefs and daily strategy use as well as specific beliefs about semantic clustering efficacy and difficulty as potential predictors of spontaneous clustering use. With the latter, we were also interested in participants' awareness of whole-list format benefits to semantic clustering.

List-Presentation Format and Memory Improvement Through Semantic Clustering

The second main goal of the present study was to examine age-related differences in semantic clustering utilization and whether the whole-list format would enable more effective semantic clustering in younger and, particularly, older adults. Only two of the prior studies, both using the individual-words format, have reported correlations between semantic clustering use and recall performance separately for younger and older participants: Witte et al. (1993) report strong positive clustering—recall correlation for both younger and older participants (numerically even larger in the older participants). Taconnat et al. (2009) report null or even reliable *negative* correlations for their older participants as opposed to strong positive correlations for their younger participants, pointing to a potential utilization deficiency in the individual-words format.

However, clustering—recall correlations may be confounded with characteristics of the persons spontaneously choosing to use semantic clustering (cf., Siegler & Lemaire, 1997). In the present study, we directly instructed participants to use semantic clustering on a second study-list to more directly measure the memory improvement through semantic clustering. We are not aware of a prior study directly instructing participants to use semantic clustering; some studies have instructed participants to sort study words printed on flashcards but typically did not compare recall performance to a nonsorting control group (Basden, Basden, & Bartlett, 1993; Guttentag, 1984; Worden & Meggison, 1984). A notable exception is the study by Hultsch (1971) in which older adults' recall improved greatly (more than younger adults') with sorting instructions compared with a nonsorting control group. However, it remains unclear to what extent older adults are able to improve their memory performance through semantic clustering when the sorting must be completed *mentally* and whether this is moderated by the presentation-format. We predicted that, given the expected cognitive demands of mental organization, older adults would benefit less from semantic clustering instructions than younger adults (i.e., a utilization deficiency), especially in the individual-words format (cf. Taconnat et al., 2009). We further predicted that the whole-list format would result in higher semantic clustering utilization success (i.e., larger memory improvement) in both younger and older adults but that this benefit might be particularly pronounced in the older adults.

Comparison of Semantic Clustering Measures

A crucial issue here is the measurement of semantic clustering, which has been widely debated (e.g., Roenker, Thompson, & Brown, 1971). Traditional measures of semantic clustering, such as the *Ratio of Repetition* (RR; Bousfield, 1953), assess what proportion of the recall output are category repetitions (i.e., successive recall of words from the same semantic category). However, while these measures certainly index use of semantic clustering, they ignore important indicators of semantic clustering *utilization success*, specifically the size and total number of the clusters. This is particularly evident in the study by Taconnat et al. (2009), where paradoxically an increase in relative output clustering was accompanied by a decrease in recall performance in older adults. Therefore, a *list-based semantic clustering score* (LBC; Stricker, Brown, Wixted, Baldo, & Delis, 2002) has been developed, which takes the semantic structure of the study list into account, not just what was recalled. Thereby, the LBC measure does not only consider the number of semantic category repetitions in the recall output but also the number and size of the clusters relative to the number and size of categories on the original list, resulting in a measure of *absolute clustering success*. To illustrate the difference between output-based and list-based clustering measures, imagine two participants studied a list including five animal words randomly intermixed with words from other categories. Later, both participants only remember the animal words but Participant A recalls only two of them whereas Participant B recalls all five. Output-based clustering measures would not differentiate between these participants but instead indicate a perfectly clustered recall output for both. In contrast, LBC would be higher (but far from perfect given that other category words were also on the list) for Participant B than A, thus better capturing the difference in absolute clustering success. Thus, an additional goal of the present study was to compare both indices. We expected that LBC would more closely reflect the recall improvements from clustering instructions (and age differences therein) than an output-based measure.

Summary of Hypotheses

In summary, we expected the whole-list presentation format to increase spontaneous semantic clustering productions (evident in either clustering index) and utilization (i.e., memory improvements from semantic clustering instructions; LBC). We expected this whole-list benefit to involve reduced resource demands of semantic clustering (i.e., weaker correlations with processing speed and working memory capacity). Therefore, given older adults' typical declines in processing resources, we expected the whole-list presentation format to more strongly affect older than younger adults.

METHOD

Participants

The sample consisted of 132 undergraduates (18–26 years old) and 120 community-dwelling older adults (60–81 years old). Younger adults took about 1.5 hr and received course credit; older adults took about 2.5 hr and received a $25 gift card to a local department store. Inclusion criteria were age 18 to 30 for younger adults and 60+ for older adults, fluency in English, no neurological disorder or prior history of stroke or heart attack, no current use of medication affecting cognition, and no prior participation in an experiment using the same material. Older adults were prescreened for these criteria on the phone; undergraduates who did not meet these criteria were allowed to participate for course credit but their data were discarded. Because of experimenter error in the data collection procedure, data from three additional (two younger and one older) participants were excluded; further, data from one additional older participant who produced almost exclusively intrusions on the recall tests were excluded. Participant characteristics are provided in Table 1 and are typical for studies in the cognitive aging literature. The older participants had on average received more years of formal education than the younger participants and outperformed them on a vocabulary measure. Nonetheless, the older participants showed typical worse performance than the younger participants on measures of processing speed and working memory capacity.[1]

Design

The design was a 2 (age group) × 2 (list-presentation format) × 2 (clustering instructions) mixed factorial. Younger and older adults were randomly assigned to either the individual-words or the whole-list presentation format (66 younger and 60 older adults each). Clustering instructions were manipulated within participants: A first list was studied without explicit instructions to use a specific strategy. Later in the session, a second list was studied with instructions to use semantic clustering at encoding. Specifically, participants were informed that the words on the next list could be grouped into different categories and that they should try to do so because such grouping improves memory. After recall of the first list, participants had been queried about the category structure of that list and were then given feedback—hence they had examples of categories. They were not informed how many or what categories would apply to the second list and they did not practice clustering. Participants later also studied a third list (again instructed to use semantic clustering) under a dual-task manipulation; results from this third list were not informative and are not reported for brevity.[2]

Table 1
Sample Characteristics

Measure	Individual words: Younger adults	Individual words: Older adults	Whole list: Younger adults	Whole list: Older adults
Age (years)	19.61 (.02)	68.72 (.57)	19.71 (.25)	70.28 (.62)
Formal education (years)*	13.48 (.17)	16.33 (.29)	13.52 (.19)	15.05 (.38)
Subjective health rating	4.41 (.08)	4.43 (.08)	4.30 (.08)	4.25 (.10)
Number medications*	0.67 (.13)	2.50 (.24)	0.50 (.11)	2.87 (.24)
Vocabulary*	14.80 (.42)	22.72 (.91)	13.70 (.45)	21.87 (.95)
Processing speed tasks				
Pattern comparison*	39.09 (1.02)	28.81 (0.69)	40.12 (0.84)	28.77 (0.65)
Digit-symbol substitution*	66.17 (1.36)	52.37 (1.22)	67.22 (1.26)	52.42 (1.56)
Working memory tasks				
Reading span*	49.25 (0.91)[a]	45.02 (1.20)[b]	45.74 (1.16)	40.92 (1.48)
Symmetry span*	28.49 (0.84)[a]	17.02 (0.94)[c]	26.39 (0.99)[d]	14.45 (1.00)[e]
Metamemory in adulthood questionnaire scales				
Achievement*	3.93 (0.05)	4.03 (0.05)	3.85 (0.05)	4.00 (0.04)[b]
Locus	3.63 (0.06)	3.60 (0.08)	3.55 (0.05)	3.65 (0.06)[b]
Strategy–internal*	3.83 (0.06)	3.57 (0.07)	3.80 (0.05)	3.51 (0.07)[b]
Strategy–external*	3.60 (0.07)	4.03 (0.08)	3.67 (0.07)	3.78 (0.08)[b]

Note. *SEM* in parentheses. Unless otherwise indicated, the number of subjects in each presentation-format condition was 66 younger and 60 older adults. Subjective health rating was on a 5-point Likert scale (1 = *very poor* to 5 = *very good*). Maximum scores are 36 for vocabulary (Ekstrom et al., 1979), 60 for pattern comparison (Salthouse, 1996), 100 for digit-symbol substitution (Wechsler, 1981), 60 for reading span and, 42 for reading/symmetry span, respectively (Unsworth et al., 2005; 70% accuracy was required on processing component [see Footnote 1] hence the reduced numbers of subjects as indicated). Metamemory in Adulthood Questionnaire (Dixon & Hultsch, 1983) scale scores are average scores from 5-point Likert response scales indicating average agreement with statements reflecting high memory-achievement aspiration or belief in internal controllability (locus) of memory or average frequency of using internal and external memory strategies.
[a] $n = 65$. [b] $n = 59$. [c] $n = 50$. [d] $n = 61$. [e] $n = 51$.
* Age-group difference significant, all $ps \leq .006$.

Measures, Material, and Procedure

Word list study and recall. Three categorizable 20-word lists containing five words each from four distinct semantic categories were constructed based on category norms (Van Overschelde, Rawson, & Dunlosky, 2004). For each category, exemplars named by at least 20% but no more than 80% of the norming sample in response to the category label were selected. Lists were matched in mean exemplar typicality (*M* = 48.43%) and word length (*M* = 5.23 [range 3–10] letters and 1.67 [range 1–3] syllables). Each participant studied and recalled all three lists during the session, with the order of lists counterbalanced across participants. To minimize interference between lists, different semantic categories were used on each list. Presentation format varied between conditions but was constant within participants. In the individual-words format, each of the 20 words was presented individually, centered on the computer screen for 3 s with a 500 ms fixation cross between words. For the whole-list conditions, all 20 words were presented at once on the screen—in two columns of 10 words each—for 69.5 s. Thereby, the total study time (including ISIs) was held constant across presentation-format conditions. In both formats, the order of words was randomized for each participant with the restriction that at least two words intervened between words from the same semantic category (rule applied within the columns for the whole-list condition). After studying, participants were asked to continuously deduct three from a provided three-digit starting number for 30 s, entering each intermediate result. Then participants were instructed to type any words they remembered from the study list in the order they came to mind. Participants saw all words they had already typed on the screen. Three minutes were given for recall, timed by the computer with no option to end early. For each list, participants were next asked about their strategy use during studying (described next). They also had to list the semantic categories on the just-studied list (which participants of both ages did with very high accuracy) and completed a category-cued recall (yielding the same pattern of results as the free recall); these two measures assured that the semantic categories of each list were highly meaningful to all participants.

Metacognitive beliefs about clustering. After study and test of each list, participants were asked (a) how effective they believe a semantic-clustering strategy is for improving their own and a peer's memory (two separate questions) for a list like the one they just studied on a 5-point Likert scale (1 = *not at all*, 2 = *slightly*, 3 = *somewhat*, 4 = *quite*, 5 = *very much*) and (b) to rate how difficult they thought semantic clustering of the words was during study and during test (two separate questions) on a 5-point Likert scale (1 = *not at all*, 2 = *slightly difficult*, 3 = *somewhat difficult*, 4 = *quite difficult*, 5 = *very difficult*).

Processing speed and working memory tasks. Participants completed two paper-based processing speed measures, the *Digit-Symbol Substitution Task* (Wechsler, 1981) and the *Pattern Comparison Task* (Salthouse, 1996). For each, the experimenter timed with a stopwatch, giving simultaneous "start" and "stop" signals to all participants in a session. Two computerized, automated complex-span tasks were used to measure working memory capacity (Unsworth, Heitz, Schrock, & Engle, 2005): the verbal reading-span task and the spatial symmetry-span task. In the reading-span task, participants read a sentence, judged whether it was sensical, and then saw a letter for 1 s. After two to six sentence-letter pairs, participants had to recall the letters in order. There were three trials of each set size, resulting in a total of 48 trials. In the symmetry-span task, participants judged whether patterns were symmetric along the vertical axis and then memorized the position of a red square in a 4 × 4 matrix presented for 650 ms. After sets of two to five symmetry judgments and square presentations, participants had to select the previously presented red squares in the 4 × 4 matrix in the shown order. With three trials for each set size, there were a total of 42 trials. Importantly, during 15 practice trials for the processing component (sentence sensibility or symmetry judgments) the participants' mean response time was measured and the main trials timed out if participants did not respond within 2 *SD*s of their mean time to limit opportunity for letter/square position rehearsal. Participants were given accuracy feedback after each trial for both the processing and the memorization component and were asked to maintain 85% accuracy on the processing task (see Footnote 1).

General procedure and task order. Participants were tested in age-homogenous groups of up to six in individual cubicles with an experimenter present throughout the entire session. Participants first signed a consent form and took a near (corrected) visual acuity test (all had at least 20/50 acuity). Tasks were then completed in the following order with breaks offered between tasks as needed: (a) Digit-symbol substitution test, (b) pattern comparison test, (c) study and recall of first, spontaneous word list (no strategy instructions), (d) vocabulary test, (e) reading span, (f) second, clustering-instructed word list, (g) dual-task practice followed by study and recall of a third word list under dual task, (h) completion of the achievement, locus, and strategy subscales from the *Metamemory in Adulthood (MIA) Questionnaire* (Dixon & Hultsch, 1983), and (i) symmetry span. Upon completion of all tasks, participants completed a computerized demographic questionnaire and were then debriefed and dismissed.

RESULTS

For all analyses, we set $\alpha = .05$.

Recall Performance

Figure 1 shows mean recall performance by age group and list-presentation format before (first list) and after (second list) clustering instructions. A 2 (age group) × 2 (presentation format) × 2 (strategy instructions) mixed analysis of variance (ANOVA) yielded an expected main effect of age group, $F(1, 248) = 53.76$, $p < .001$, $\eta p^2 = .18$, with the older adults consistently recalling fewer words than the younger adults. There was further a main effect of strategy instructions, $F(1, 248) = 53.18$, $p < .001$, $\eta p^2 = .18$, that interacted with age group, $F(1, 248) = 9.22$, $p = .003$, $\eta p^2 = .04$, and with list-presentation format, $F(1, 248) = 11.96$, $p < .001$, $\eta p^2 = .05$. No other effects were significant, all $F < 1$. Follow-up planned comparisons showed that recall improved on the second, clustering-instructed list in both age groups but this strategy-instructions effect was much stronger in the younger, $F(1, 248) = 56.01$, $p < .001$, $\eta p^2 = .18$, than in the older participants, $F(1, 248) = 8.65$, $p = .004$, $\eta p^2 = .03$, evidencing a utilization deficiency. Overall, recall performance significantly improved on the second, clustering-instructed list in both list-presentation formats, but this improvement was much stronger in the whole-list format, $F(1, 248) = 57.79$, $p < .001$, $\eta p^2 = .19$, than in the individual-words format, $F(1, 248) = 7.35$, $p = .007$, $\eta p^2 = .03$. Thus, as expected, the whole-list format led to greater success in implementing the clustering instructions. Notably, this benefit of the whole-list format was comparable for younger and older adults (i.e., no three-way interaction), contrary to our initial hypotheses. Given our a priori hypotheses, however, we examined the format and instruction effects separately for each age group. Qualitatively, it is notable that older adults did not at all improve their average recall performance on the second, clustering-instructed list in the individual-words format, $t(59) = 0.65$, $p = .518$, $dz = 0.08$, whereas a significant medium-sized improvement was possible in the whole-list format, $t(59) = 3.25$, $p = .002$, $dz = 0.42$. For younger adults, however, already a medium-sized improvement was possible in the individual-words format, $t(65) = 3.77$, $p < .001$, $dz = 0.46$, which rose to a large-sized improvement in the whole-list format, $t(65) = 7.00$, $p < .001$, $dz = 0.86$. Thus, for older adults, list-presentation format determined whether *any* average memory improvement was achieved on the second, clustering-instructed list, whereas younger adults were able to significantly improve their recall performance in either format (but more so in the whole-list format).

Semantic Clustering Performance

We will first present analyses of RR, a more traditional relative output clustering measure, followed by analyses of LBC, a measure of absolute clustering success. Note that correlations between both indices were quite high in all conditions (younger adults: .928 ≤ r ≤ .969; older adults: .549 ≤ r ≤ .893), all $p < .001$.

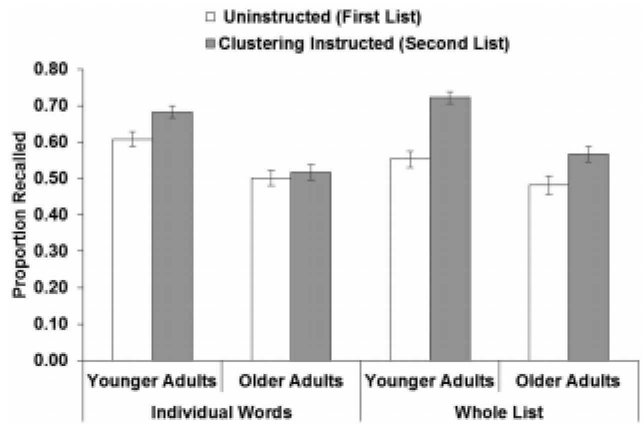

Figure 1. Mean proportion of words recalled from the first, uninstructed (white bars) and the second, clustering-instructed (gray bars) list by age group and presentation format. Error bars display the *SEM*.

Relative output clustering (RR). The RR (Bousfield, 1953) score is computed as the number of category repetitions divided by the total number of words recalled −1 and has been particularly recommended for developmental research (Frender & Doubilet, 1974).[3] Mean RR scores (see Table 2) were submitted to a 2 (age group) × 2 (list-presentation format) × 2 (strategy instructions) mixed ANOVA. Expectedly, there was a large main effect of strategy instructions, $F(1, 248) = 239.52$, $p < .001$, $\eta p^2 = .49$, with higher RR scores on the clustering-instructed (second) list, that further interacted with list-presentation format, $F(1, 248) = 4.89$, $p = .028$, $\eta p^2 = .02$. Follow-up planned comparisons revealed that RR scores increased with clustering instructions in both presentation-format conditions but this increase was larger in the whole-list, $F(1, 248) = 156.42$, $p < .001$, $\eta p^2 = .39$, than in the individual-words format, $F(1, 248) = 87.99$, $p < .001$, $\eta p^2 = .26$. That is, participants implemented the clustering instructions more successfully in the whole-list format. The main effect of age group, $F(1, 248) = 6.47$, $p = .012$, $\eta p^2 = .03$, also interacted with list-presentation format, $F(1, 248) = 4.96$, $p = .027$, $\eta p^2 = .02$, but not with any other factors, all Fs < 1. Follow-up planned comparisons revealed that presentation format had no main effect in younger adults, $F < 1$, but older adults' recall outputs were significantly more clustered in the whole-list than in the individual-words format, $F(1, 248) = 4.36$, $p = .017$, $\eta p^2 = .02$. Given this selective effect of presentation format on older adults' output clustering, age-related differences in RR varied across presentation format: Average output clustering did not differ between age groups in the individual-words format, $F < 1$, but older adults' recall outputs were, on average, significantly *more* clustered than those of younger adults' in the whole-list format, $F(1, 248) = 10.50$, $p = .001$, $\eta p^2 = .04$. Note that *more*clustering in older adults is also the direction of the significant main effect of age group. In sum, there was no evidence for a production or utilization deficiency in older adults' output clustering. Quite contrary, older adults' achieved better output clustering than younger adults. Both age groups' output clustering was supported by the whole-list presentation format, with some evidence for particular benefits of the whole-list format in the older adults.

Table 2
Semantic Clustering Performance by Age Group and Presentation Format

Measure	Individual words		Whole list	
	Younger adults	Older adults	Younger adults	Older adults
RR	L1: 0.37 (.02)	L1: 0.38 (.03)	L1: 0.31 (.03)	L1: 0.41 (.03)
	L2: 0.60 (.02)	L2: 0.59 (.02)	L2: 0.61 (.03)	L2: 0.70 (.02)
LBC	L1: 2.12 (.32)	L1: 1.90 (.28)	L1: 1.46 (.36)	L1: 2.22 (.35)
	L2: 5.05 (.35)	L2: 3.80 (.29)	L2: 5.58 (.40)	L2: 5.01 (.25)
Clustering difficulty	L1: 2.12 (.10)	L1: 2.39 (.10)	L1: 2.16 (.11)	L1: 2.33 (.13)
	L2: 1.90 (.11)	L2: 2.59 (.15)	L2: 1.90 (.11)	L2: 2.21 (.11)
Clustering efficacy	L1: 3.81 (.11)	L1: 3.87 (.11)	L1: 3.71 (.10)	L1: 3.72 (.10)
	L2: 4.06 (.11)	L2: 3.93 (.11)	L2: 4.22 (.11)	L2: 3.91 (.11)

Note. *SEM* in parentheses. The number of subjects in each presentation-format condition was 66 younger and 60 older adults. L1 = first, spontaneous (no clustering instruction) list. L2 = second, clustering-instructed list. RR = ratio of repetition (0 = recall output contains no category repetitions; 1 = recall output contains only category repetitions; Bousfield, 1953; Frender & Doubilet, 1974). One older adult in the whole-list condition who recalled only one word on L2 was assigned a RR of 0 (see Footnote 3). LBC = list-based semantic clustering index (0 = chance-level semantic clustering; 12 = maximum semantic clustering [recalling all 20 words ordered by category]; Stricker et al., 2002).

Absolute clustering success (LBC). LBC was computed according to the formula provided by Stricker et al. (2002), taking the original list structure (20 words totals; 4 per each of 5 different semantic categories) into account. Mean LBC scores are provided in Table 2 and were also analyzed with a 2 (age group) × 2 (list-presentation format) × 2 (strategy instructions) mixed ANOVA. Like with RR scores, there was a large main effect of strategy instructions, $F(1, 248) = 237.02$, $p < .001$, $\eta p^2 = .49$, with higher LBC scores on the second, clustering-instructed list. This strategy-effect again interacted with list-presentation format, $F(1, 248) = 7.47$, $p = .007$, $\eta p^2 = .03$. Follow-up planned comparisons revealed that, just like RR scores, LBC scores increased with clustering instructions in both presentation formats but this increase was larger in the whole-list, $F(1, 248) = 164.33$, $p < .001$, $\eta p^2 = .40$, than in the individual-words format, $F(1, 248) = 80.16$, $p < .001$, $\eta p^2 = .24$. The main effect of age group was not significant, $F(1, 248) = 1.37$, $p = .243$, and age group also did not interact with presentation format for this clustering measure, $F(1, 248) = 2.31$, $p = .130$, but with strategy instruction, $F(1, 248) = 9.53$, $p = .002$, $\eta p^2 = .04$. The three-way interaction was not significant, $F < 1$. Follow-up planned comparisons on the significant Age Group × Strategy Instruction interaction revealed that although both age groups clustered more on the second, clustering-instructed list, this increase in LBC with strategy instructions was more strongly pronounced in the younger, $F(1, 248) = 179.33$, $p < .001$, $\eta p^2 = .42$, than in the older adults, $F(1, 248) = 72.31$, $p < .001$, $\eta p^2 = .23$. Consequently, LBC scores did not differ between age groups for the first spontaneous list, $F < 1$, but were significantly higher in younger adults on the second, clustering-instructed list, $F(1, 248) = 7.40$, $p = .007$, $\eta p^2 = .03$, mirroring the magnification of age-group differences after clustering instructions observed in recall. In summary, in line with the RR analyses, the LBC measure suggests that spontaneous clustering did not differ between the two presentation-formats but once clustering was instructed (on the second list) participants were more successful in implementing this strategy in the whole-list format. The measures diverge regarding age-group differences in semantic clustering: Although the (much smaller) recall outputs of older adults were at least as clustered (RR measure) as those of younger adults (even more in the whole-list format), the LBC measure suggests that younger adults clustered more successfully than older adults (i.e., more and larger clusters) on the second, clustering-instructed list.

Clustering—Recall relationship. As evident in Table 3 (first rows), both clustering indices were highly correlated with recall performance on both lists in all conditions with one exception (i.e., null correlation between RR and recall on the instructed list in the older adult whole-list condition). Numerically, LBC's correlation with recall was consistently higher than that of RR. We used the software by Lee and Preacher (2013) to compare these dependent correlation coefficients within each condition (separately for each list). LBC was indeed significantly more correlated with recall than RR in all but one of the eight comparisons (not for the uninstructed, first list in the younger adult whole-list condition, $z = 0.12$, $p = .904$), all other $z \geq 2.78$, $p \leq .005$. This supports our interpretation of the LBC measure as an index of absolute clustering success. Finally, to more directly test whether increases in recall from the first to the second list were indeed related to increased clustering we computed increase scores (i.e., List 2–List 1) for recall and clustering (RR or LBC) for each participant and correlated these. In younger adults, both the increases in RR (individual words: $r(64) = .48$; whole list: $r(64) = .54$) and LBC (individual words: $r(64) = .67$; whole list: $r(64) = .61$) were significantly correlated with the observed recall increases, all $p < .001$. For older adults, increases in RR were not significantly correlated with the (on average null) recall increases in the individual words format, $r(58) = .197$, $p = .132$, but the correlation was significant in the whole-list format, $r(58) = .320$, $p = .013$. Increases in LBC were significantly correlated with older participants' recall increases in both the individual-words, $r(58) = .564$, $p < .001$, and the whole-list format, $r(58) = .717$, $p < .001$.

Metacognitive Beliefs About Semantic Clustering

Participants' estimates of the efficacy of semantic clustering for improving their own or peers' memory were very similar and thus averaged. Likewise, the very similar difficulty ratings for using semantic-clustering during study and during test were averaged. Table 2 displays mean values for both measures, which were each analyzed with a 2 (age group) × 2 (list-presentation format) × 2 (strategy instructions) mixed ANOVA.

Clustering efficacy. Regarding clustering-efficacy beliefs, there was a main effect of strategy instruction, $F(1, 248) = 23.97$, $p < .001$, $\eta p^2 = .09$, with participants ascribing higher efficacy to the strategy after having been instructed to use it on the second list. There was no main effect of presentation format, $F < 1$, but a marginally significant interaction between presentation format and strategy instruction, $F(1, 248) = 3.76$, $p = .054$, $\eta p^2 = .02$. Tentatively, the increase in clustering-efficacy ratings from the first to the second list was somewhat more pronounced in the whole-list, $F(1, 248) = 23.35$, $p < .001$, $\eta p^2 = .09$, than in the individual-words format, $F(1, 248) = 4.37$, $p = .038$, $\eta p^2 = .02$, mirroring the actual format differences in recall benefits. There was no main effect of age group, $F < 1$, and no interaction of age group and presentation format, $F < 1$, but an interaction of age group and strategy instruction, $F(1, 248) = 6.14$, $p = .014$, $\eta p^2 = .02$. Follow-up planned comparisons revealed that the increase in efficacy ratings from the first to the second list was large and significant in younger adults, $F(1, 248) = 28.55$, $p < .001$, $\eta p^2 = .10$, but smaller and only trending in older adults, $F(1, 248) = 2.79$, $p = .096$, $\eta p^2 = .01$. This mirrors the observed smaller recall benefits of the semantic-clustering instruction in older compared with younger adults. The three-way interaction was not significant, $F < 1$.

Table 3

Correlations Between Semantic-Clustering Indices and Proportion Recalled, Processing Speed, Working Memory, and Vocabulary

	Individual words				Whole list			
	Younger adults		Older adults		Younger adults		Older adults	
Measure	RR	LBC	RR	LBC	RR	LBC	RR	LBC
Uninstructed clustering (List 1)								
Prop recalled	**.543***	**.625***	**.491***	**.682***	**.607***	**.610***	**.484***	**.713***
Processing speed	.061	.061	−.032	.054	.151	.174	−.095	.136
Working memory capacity	.151	.160	−.173	−.051	.045	.089	−.121	.000
Years of formal education	.092	.034	.010	−.012	.089	.033	−.060	.116
Vocabulary	.062	.028	.245†	.248†	−.186	−.136	−.048	.046
Clustering efficacy	.233†	.231†	.013	.125	**.389***	**.422***	.224†	.217†
Clustering difficulty	**−.236***	−.247†	−.199	**−.292***	**−.373***	**−.377***	**−.315***	**−.451***
MIA–Achievement	−.125	−.130	−.050	.000	.094	.119	−.031	−.038
MIA–Locus internal	.078	.066	−.037	.003	−.004	.011	.154	.129
MIA–Strategy internal	−.185	−.189	.075	.103	.017	.028	.026	.066
MIA–Strategy external	.021	−.006	.127	.138	.017	.020	−.141	−.175
Instructed clustering (List 2)								
Prop recalled	**.373***	**.662***	**.419***	**.851***	**.426***	**.649***	−.043	**.868***
Processing speed	.015	.079	.245†	**.267***	.018	.042	.181	**.293***
Working memory capacity	.212†	.211†	**.375***	**.414***	.165	.125	.155	**.385***
Years of formal education	−.088	−.077	−.031	.163	.163	.198	−.048	**.112**
Vocabulary	.044	.106	.030	.238†	−.015	.019	.070	**.264***
Clustering dfficacy	.047	.070	.218†	**.303***	.226†	.230†	.093	**.421***
Clustering difficulty	−.025	−.049	−.128	−.200	−.211	**−.404***	.058	−.252†
MIA–Achievement	.109	.156	.055	−.006	.129	.182	.011	.146
MIA–Locus internal	.042	.114	.093	−.107	−.021	.019	.039	.151
MIA–Strategy internal	.096	.078	.096	−.033	−.043	−.032	.096	.199
MIA–Strategy external	−.014	.018	.147	.170	.036	.009	.061	.067

Note. There were 66 younger and 60 older adults in each presentation-format condition. For clustering efficacy and difficulty beliefs and MIA scales, which were measured on 5-point Likert scales, Spearman *rho* is presented. All other values are Pearson *r* coefficients. Significant correlations (two-tailed test) are bolded. Processing speed is the average of the *z*-standardized performance on the digit-symbol substitution and letter comparison tasks. Working memory is the average of the *z*-standardized performance on the reading and symmetry span tasks (see Footnote 1). MIA = scores from the Metamemory in Adulthood Questionnaire (Dixon & Hultsch, 1983) scales, which are average scores from 5-point Likert response scales indicating average agreement with statements reflecting high memory-achievement aspiration or belief in internal controllability (locus) of memory or average frequency of using internal and external memory strategies. RR = ratio of repetition (Bousfield, 1953; Frender & Doubilet, 1974). LBC = list-based (semantic) clustering index (Stricker et al., 2002).
† *p* (one-tailed) < .05 (in predicted direction). * *p*(two-tailed) < .05.

Clustering difficulty. Regarding ratings of clustering difficulty, there were no main effects of strategy instructions, $F(1, 248) = 2.10$, $p = .149$, or presentation format, $F < 1$. However, older adults rated clustering to be more difficult, $F(1, 248) = 15.50$, $p < .001$, $\eta p^2 = .06$, and this effect of age group interacted with strategy instructions, $F(1, 248) = 3.96$, $p = .048$, $\eta p^2 = .02$. Follow-up planned comparisons revealed that younger adults decreased their ratings of clustering difficulty from the first to the second list, $F(1, 248) = 6.20$, $p = .013$, $\eta p^2 = .02$, whereas older adults' difficulty ratings did not change between lists, $F < 1$. Consequently, there was a larger age-group difference in clustering difficulty ratings on the second, clustering-instructed list, $F(1, 248) = 17.60$, $p < .001$, $\eta p^2 = .07$, than on the first, uninstructed list, $F(1, 248) = 4.03$, $p = .046$, $\eta p^2 = .02$. No other interactions were significant, all $F \leq 1.72$. That is, although participants achieved higher recall performance and higher clustering in the whole-list format, difficulty beliefs did not differ between presentation formats.

Correlates of Clustering

Finally, we examined our hypothesized correlations between semantic clustering and general processing resources (processing speed and working memory capacity) and metacognitive measures (general metamemory beliefs, daily strategy use, and specific beliefs about semantic clustering). For processing speed and working memory capacity, average performance scores across the two tasks measuring each of these constructs were computed by first *z*-standardizing the scores on each individual task (with the full sample [both age groups and both presentation formats combined] mean and *SD*) and then averaging these standardized scores for the two processing-speed/working memory tasks, respectively. To be as inclusive as possible, if a participant only met the processing criterion for one of the span tasks (see Footnote 1) the *z*-score on that span task was used. All zero order correlations are displayed in Table 3. For completeness, correlations with vocabulary are also included in Table 3; these were nonsignificant with one exception, presumably because of the easiness of the used semantic categories.

Correlations with general processing resources. We expected positive correlations between clustering indices and general processing resources for both spontaneous and instructed clustering, particularly in the individual-words format. Contrary to our expectations, there was no evidence for any correlation between these general processing resource measures and spontaneous clustering, neither in younger nor in older adults. However, in line with the idea that clustering is resource-demanding, *instructed* clustering positively correlated with processing speed (with the LBC measure only) and working memory capacity (with LBC in both formats; with RR only in the individual-words condition) in older adults (only trending in younger adults). Notably, this was true for both presentation formats with no evidence for the expected stronger correlations in the individual-words format, all p (one-tailed) $\geq .102$.

Table 4 displays results from regression models with both general processing resource measures as predictors of RR and LBC variance. General processing resources did not predict significant variance in spontaneous clustering (List 1) in either age group. On List 2 (instructed clustering) general processing resources explained significant variance in both older adults' instructed (List 2) RR (16%) and LBC (19%) in the individual-words conditions. In the whole-list conditions, general processing resources explained significant variance in older adults' instructed LBC (15%) but not RR (4%). In all significant models, only working memory capacity predicted unique variance in RR/LBC whereas the coefficients for processing speed became nonsignificant when working memory was simultaneously included (see Table 4). For younger adults, none of the correlations between clustering indices and either processing resource measure were significant nor were any regression models with both entered as predictors (see Table 4). However, it seems premature to conclude that clustering was fully independent of working memory capacity in the younger adults. At least in the individual-words condition, the younger participants' correlations between working memory capacity and both RR and LBC were significant when tested *one-tailed* (in the expected positive direction). Finally, when formally comparing the multiple *R*s, the correlation between processing resources and clustering was not significantly larger in older than younger adults (although trending for LBC) in either format, RR: $z = 1.07$ (individual words)/0.17 (whole list), both p (one-tailed) = .142/.440, LBC: z = 1.31/1.46, p (one-tailed) = .095/.072.

Metacognitive measures. General metamemory beliefs and daily strategy use, as measured in the MIA, did not correlate with spontaneous or instructed clustering in any condition. However, some of the correlations between specific metacognitive beliefs about semantic clustering were significant (in the expected directions, that is positive for efficacy beliefs and negative for difficulty beliefs) for both spontaneous and instructed clustering in both age groups.

Table 4

Prediction of Spontaneous and Instructed Clustering by General Processing Resources or Specific Metacognitive Beliefs

		Individual words						Whole list					
		RR			LBC			RR			LBC		
Predictor group	Age group		β	R^2		β	R^2		β	R^2		β	R^2
Uninstructed clustering (List 1)													
General processing resources	Younger adults	PS:	.03	.02	PS:	.07	.03	PS:	.15	.02	PS:	.16	.03
		WMC:	.14		WMC:	.14		WMC:	.01		WMC:	.05	
	Older adults	PS:	.02	.03	PS:	.08	.01	PS:	.15	.12	PS:	.22	.03
		WMC:	−.18		WMC:	−.08		WMC:	.01		WMC:	−.13	
Specific metacognitive beliefs	Younger adults	Eff:	.19	.07†	Eff:	.12	.06	**Eff:**	**.33***	**.27***	**Eff:**	**.36***	**.31***
		Diff:	−.13		Diff:	−.12		**Diff:**	**−.32***		**Diff:**	**−.34***	
	Older adults	Eff:	−.01	.02	Eff:	.03	.05	Eff:	.12	.09†	Eff:	.12	**.25***
		Diff:	−.15		Diff:	−.23†		Diff:	−.25†		**Diff:**	**−.46***	
Instructed clustering (List 2)													
General processing resources	Younger adults	PS:	−.04	.05	PS:	.03	.05	PS:	−.02	.03	PS:	.02	.02
		WMC:	.22†		WMC:	.20		WMC:	.17		WMC:	.12	
	Older adults	PS:	.14	**.16***	PS:	.15	**.19***	PS:	.14	.04	PS:	.09	**.15***
		WMC:	**.33***		**WMC:**	**.37***		WMC:	.07		**WMC:**	**.33***	
Specific metacognitive beliefs	Younger adults	Eff:	.04	.01	Eff:	.03	.01	**Eff:**	**.42***	**.15***	**Eff:**	**.37***	**.15***
		Diff:	−.06		Diff:	−.09		Diff:	.09		Diff:	−.04	
	Older adults	**Eff:**	**.26***	.09†	Eff:	.23†	**.20***	Eff:	.16	.03	**Eff:**	**.35***	**.17***
		Diff:	−.14		Diff:	**−.35***		Diff:	.09		Diff:	−.16	

Note. Causality cannot be inferred from the present data. Significant R^2 values and predictors and their regression coefficient are bolded. RR = ratio of repetition (Bousfield, 1953; Frender & Doubilet, 1974). LBC = list-based (semantic) clustering index (Stricker et al., 2002). PC = processing speed (average of the z-standardized performance on the digit-symbol substitution and letter comparison tasks). WMC = working memory capacity (averaged of the z-standardized performance on the reading and symmetry span tasks; see Footnote 1). Eff = clustering efficacy rating. Diff = clustering difficulty rating.

† p (one-tailed) < .05 (in predicted direction). * p(two-tailed) < .05.

Like for the analysis of processing resources, we again entered both specific belief measures simultaneously into a regression model predicting clustering. Note that, of course, no causal relationship can be inferred from these data, particularly for these measures that were assessed after each study-test cycle. Table 4 shows that both specific belief measures predicted significant variance in spontaneous LBC (25–31%) in the whole-list but not in the individual-words (5–6%) format in both age groups. For RR, only the model in the younger-adults whole-list condition was significant (27%). For instructed LBC, both measures continued to predict significant variance in the whole-list condition for both age groups (15–17%) and also in the individual-words format for older (20%) but not younger adults (1%). For instructed RR, the model again was only significant in the younger adult whole-list condition (15%). No clear pattern emerged regarding a differential predictive power of effectiveness versus difficulty beliefs (see Table 4).

Given that both general processing resource measures as well as specific metacognitive beliefs predicted significant variance in instructed LBC for the older participants, we also ran a final regression model including both predictor groups. This model explained 33% of older adults' instructed LBC variance in the individual-words format, $F(4, 55) = 6.74$, $p < .001$, and 23% in the whole-list format, $F(4, 55) = 4.10$, $p = .006$.

DISCUSSION

The results confirm that presentation of the whole study-list during encoding supports semantic clustering in both younger and older adults. However, whole-list presentation only increased clustering and, consequently, recall performance compared with the individual-words presentation when participants were instructed to use semantic clustering. On the uninstructed, first list neither younger nor older adults primarily used semantic clustering in either presentation format and, consequently, their recall performance did not vary between formats. Although clustering and recall benefits of the whole-list format were generally comparable for younger and older adults (with the notable exception of increased whole-list benefits on RR in older adults), there was an important qualitative difference: For older adults presentation format made the difference between not improving at all (in the individual-words format) and substantially improving memory performance (in the whole-list format) on the second, clustering-instructed list. Younger adults, on the other hand, were able to improve their recall performance with clustering instructions in either format (but more so in the whole-list format).

The present results suggest equivalent spontaneous use of semantic clustering by younger and older adults, in line with a host of other studies (Blatt-Eisengart & Lachman, 2004; Hertzog et al., 1990; Hess et al., 2003; Lachman & Andreoletti, 2006; Rankin, Karol, & Tuten, 1984; West et al., 2009; Witte et al., 1993; Zivian & Darjes, 1983) but contradicting some prior studies reporting less frequent spontaneous semantic clustering in older adults (Amrhein et al., 1999; Howard et al., 1981; Hultsch, 1971; Jacobs et al., 2001; Mungas et al., 1991; Schneider & Uhl, 1990; Taconnat et al., 2009). Most of the prior studies documenting an age-related production deficiency for semantic clustering used an individual-words format and this motivated our systematic manipulation of presentation format, but we ultimately found no evidence for effects of presentation format on spontaneous clustering use in either age group. Metacognitive judgments, assessed postrecall of the first list, suggest that participants did not recognize the benefits of the whole-list format in increasing clustering efficacy and decreasing clustering difficulty. This is in line with Koriat's metacognitive cue-utilization framework and the finding that initial metacognitive judgments primarily reflect item-intrinsic but not external task factors (Koriat, 1997). An interesting find was that both younger and older participants seem to have updated their strategy-efficacy beliefs after performing the semantic-clustering strategy on the second, clustering-instructed list, after which beliefs were somewhat sensitive to presentation format. Our regression analyses suggest that if these metacognitive beliefs are treated as predictors of clustering (and they may just as well be consequences of the clustering experience given their assessment after each study-test cycle), only a small (not always significant) portion of variance in clustering can be explained. Although at first glance, our older participants' high education and vocabulary knowledge may explain the lack of differences in spontaneous clustering, neither of these variables correlated with spontaneous clustering and in several of the studies reporting a clustering production deficiency the older participants also had higher verbal skills (Jacobs et al., 2001; Mungas et al., 1991; Taconnat et al., 2009). Overall, we thus could not identify any clear predictors of spontaneous semantic clustering in either age group.

In terms of memory-improvement through semantic clustering, the present study clearly demonstrates impairments in older adults' semantic clustering utilization. This deficit was equally present in both presentation formats but it is particularly notable that older adults' average recall performance did not improve at all despite increases in clustering on the second, instructed list in the individual-words format. This confirms Taconnat et al.'s (2009) conclusions regarding older adults’ utilization deficiency, but with the important extension that in a whole-list presentation format older adults did significantly benefit from semantic clustering, albeit less so than younger adults. Thereby, this study adds to a growing body of research suggesting that task format affects older (and younger) adults’ ability to benefit from cognitive strategies (Bottiroli et al., 2010; Touron & Hertzog, 2004). More important, this study was one of the first to test the impact of semantic clustering on recall performance through instruction effects rather than through correlations with spontaneous use (but see also Hultsch, 1971). Assessing clustering instructions allows examination of the causal influence of clustering on memory performance whereas correlations between spontaneous clustering and recall performance cannot be clearly attributed to strategy effects because spontaneous use is potentially confounded with other participant characteristics (cf., Siegler & Lemaire, 1997). The within-subject manipulation used in this study, however, came at the cost of the instructed list always being the second list studied. Hence, any of the improvements seen on the second list may also reflect practice effects. However, such practice effects should be comparable in both presentation formats; therefore, the differences between formats clearly point to the role of semantic clustering in the observed recall-improvements. This interpretation is also supported by the substantial correlations between recall and clustering increases from List 1 to List 2. Further, even though the clustering indices do not clarify whether semantic clustering occurred at encoding and/or retrieval, the presentation-format effects suggest a role of encoding, in line with other studies that documented the influence of manipulations at the encoding stage on semantic clustering observed in recall outputs (Jacobs et al., 2001; Park et al., 1989).

Further, our results also show that the choice of clustering measure is not trivial: Older adults’ difficulty to successfully implement the clustering strategy (as evident in their smaller recall improvements) was only evident in the list-based clustering measure LBC but not in output-based clustering measures like RR. Both younger and older adults were able to produce highly clustered recall outputs—however, younger adults did so for many more words than older adults, indicating a superior absolute clustering performance that was not captured in relative output clustering. Similarly, Delis et al. (2010) report that LBC better discriminates between older adults with and without dementia than a relative output-clustering measure. A potential criticism of the LBC is that it is not completely independent from recall performance. Note, however, that LBC is not always high when recall is high; it is only high when recall is high *and* clustered. Indeed, results from our first, uninstructed list show that LBC does not necessarily differ between two age groups just because their recall performance differs. Nonetheless, we believe that additionally considering the output-based clustering indices is useful—in our study, the RR measure confirmed that older adults were able to follow the semantic-clustering instructions, producing recall outputs at least as clustered as those of younger adults. In a way, this measure can serve as a manipulation check with regards to measuring clustering *attempts* as opposed to measuring clustering *success*, which we believe to be better reflected in the LBC index. Considering different clustering indices alongside each other thus offers clear advantages.

Clustering (both RR and LBC) on the instructed list was related to working memory capacity, particularly in the older adults. Thereby, this research adds to a small body of literature evidencing the common assumption that semantic clustering is a resource-demanding strategy (Jacobs et al., 2001; Park et al., 1989; Taconnat et al., 2006; Wegesin et al., 2000) and also fits into a broader literature documenting that age-related deficiencies with encoding strategies correlate with age-related declines in basic cognitive processing resources (Kliegl et al., 1990; Verhaeghen & Marcoen, 1996). Notably, the resource-demands of clustering only became evident when clustering use was instructed. Nonetheless some prior studies have found (weak) correlations between spontaneous clustering use and general processing resource measures (Taconnat et al., 2009; Wegesin et al., 2000) and divided attention reduces spontaneous clustering (Park et al., 1989), in line with formal models which assume that a strategy's resource demands influence strategy choices (Siegler & Shipley, 1995Shrager & Siegler, 1998). It is possible that in the studies reporting age-related differences in spontaneous clustering there were subtle demands to use clustering in the instructions. Either way, our results suggest that it is worthwhile to separately examine instructed clustering when trying to establish the resource-demands of the strategy (cf., Siegler & Lemaire, 1997).

We originally hypothesized that the whole-list presentation format would decrease the processing-resource demands of clustering, but the positive correlations between working memory capacity and clustering success were generally comparable in both presentation formats. Interestingly, whole-list participants also did not rate clustering to be less difficult than individual-word participants. Thus, mental organization of words into categories is still effortful even when the words do not have to be maintained in (or retrieved into) working memory. Nonetheless, despite clustering's continuing resource-demands in the whole-list format, both younger and older participants achieved greater clustering (and recall) performance in this format. Future research could combine Jacobs et al.'s (2001) blocking manipulation (i.e., organization of the words by the experimenter) with the whole-list format to potentially eliminate clustering's working-memory demands and maybe even mitigate age-group differences in clustering success and recall performance. Finally, it must be considered that we only used brief strategy instructions and did not allow for practice of the clustering strategy. It is possible that the resource demands of clustering would reduce with training, especially in the whole-list format. We are not aware of any prior study training semantic clustering in older adults. Organization was trained alongside other strategies in a *strategic training* condition in Cavallini et al. (2003) but there was no specific assessment of clustering performance or of performance at different points in the training nor was a younger-adult comparison group included. Note, however, that for the demanding (i.e., memory improvements correlated with processing speed; Kliegl et al., 1990) loci method strategy, Baltes and Kliegl (1992) have documented that, despite improvements with training, the magnification of age-group differences after strategy instructions persists over as many as 38 1-hr training sessions (distributed across 1 year and 4 months). Thus, for strategies demanding of general processing resources, it may not be possible to overcome older adults' utilization deficiency, even with prolonged training.

In summary, the present study aimed to reconcile inconsistent findings on aging and semantic clustering by systematically evaluating the role of the list-presentation format (individual words vs. whole list). Although spontaneous clustering unexpectedly did not vary with presentation format, the present results clearly demonstrate that the presentation format substantially influences both younger and older adults' ability to use the semantic clustering strategy when instructed to. The choice of presentation format thus is not trivial. Depending on the diagnostic goal, one format may be preferred over the other: The individual-words format may be more sensitive to age-group differences whereas the whole-list format seems to better capture older adults' potential. Future research should determine in as much the whole-list format may also support semantic clustering in older adults with dementia, who have profound deficits in this strategy compared with healthy older adults (e.g., Delis et al., 2010), and whether one presentation format is better at discriminating between older adults with versus without dementia.

Acknowledgement

This research was supported by an APA Dissertation Research Award presented to Beatrice G. Kuhlmann and was completed in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of North Carolina at Greensboro. We thank the other members of the dissertation committee—Stuart Marcovitch, Michael J. Kane, Lili Sahakyan, and Robert Guttentag—for helpful comments on this project. Parts of this research were presented at the 2013 *Tagung experimentell arbeitender Psychologen* in Vienna (Austria) and at the 2014 Cognitive Aging Conference in Atlanta, Georgia. We thank Christopher Dillon, Lauren Harden, and Ashley Jennings for help with participant recruitment and data collection.

REFERENCES

Amrhein, P. C., Bond, J. K., & Hamilton, D. A. (1999). Locus of control and the age difference in free recall from episodic memory. *Journal of General Psychology*, *126*, 149–164. 10.1080/00221309909595358

Baltes, P. B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology*, *28*, 121–125. 10.1037/0012-1649.28.1.121

Basden, B. H., Basden, D. R., & Bartlett, K. (1993). Memory and organization in elderly subjects. *Experimental Aging Research, 19*, 29–38. 10.1080/03610739308253921

Blatt-Eisengart, I., & Lachman, M. E. (2004). Attributions for memory performance in adulthood: Age differences and mediation Effects. *Neuropsychology, Development, and Cognition*, *11*, 68–79. 10.1076/anec.11.1.68.29364

Bottiroli, S., Dunlosky, J., Guerini, K., Cavallini, E., & Hertzog, C. (2010). Does task affordance moderate age-related deficits in strategy production?*Neuropsychology, Development, and Cognition*, *17*, 591–602. 10.1080/13825585.2010.481356

Bouazzaoui, B., Isingrini, M., Fay, S., Angel, L., Vanneste, S., Clarys, D., & Taconnat, L. (2010). Aging and self-reported internal and external memory strategy uses: The role of executive functioning. *Acta Psychologica*, *135*, 59–66. 10.1016/j.actpsy.2010.05.007

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*, 229–240. 10.1080/00221309.1953.9710088

Bryan, J., Luszcz, M. A., & Pointer, S. (1999). Executive function and processing resources as predictors of adult age differences in the implementation of encoding strategies. *Neuropsychology, Development, and Cognition*, *6*, 273–287. 10.1076/1382-5585(199912)06:04;1-B;FT273

Cavallini, E., Pagnin, A., & Vecchi, T. (2003). Aging and everyday memory: The beneficial effect of memory training. *Archives of Gerontology and Geriatrics, 37*, 241–257. 10.1016/S0167-4943(03)00063-3

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. 10.1016/S0022-5371(72)80001-X

Delis, D. C., Fine, E. M., Stricker, J. L., Houston, W. S., Wetter, S. R., Cobell, K., . . .Bondi, M. W. (2010). Comparison of the traditional recall-based versus a new list-based method for computing semantic clustering on the California Verbal Learning Test: Evidence from Alzheimer's disease. *The Clinical Neuropsychologist*, *24*, 70–79. 10.1080/13854040903002232

Delis, D. C., Massman, P. J., Kaplan, E., Mckee, R., Kramer, J. H., & Gettman, D. (1991). Alternate form of the California Verbal Learning Test: Development and reliability. *Clinical Neuropsychologist*, *5*, 154–162. 10.1080/13854049108403299

Derwinger, A., Stigsdotter Neely, A., & Bäckman, L. (2005). Design your own memory strategies! Self-generated strategy training versus mnemonic training in old age: An 8-month follow-up. *Neuropsychological Rehabilitation*, *15*, 37–54. 10.1080/09602010343000336

Dixon, R. A., & Hultsch, D. F. (1983). Structure and development of metamemory in adulthood. *Journal of Gerontology*, *38*, 682–688. 10.1093/geronj/38.6.682

Drevenstedt, J., & Bellezza, F. S. (1993). Memory for self-generated narration in the elderly. *Psychology and Aging, 8,* 187–196. 10.1037/0882-7974.8.2.187

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production?*Psychology and Aging*, *13*, 597–607. 10.1037/0882-7974.13.4.597

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition*, *29*, 247–253. 10.3758/BF03194918

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology, 41*, 389–400.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*, *79*, 3–84.

Fernandes, M. A., & Grady, C. (2008). Age differences in susceptibility to memory interference during recall of categorizable but not unrelated word lists. *Experimental Aging Research*, *34*, 297–322. 10.1080/03610730802273860

Frender, R., & Doubilet, P. (1974). More on measures of category clustering in free recall-although probably not the last word. *Psychological Bulletin*, *81*, 64–66. 10.1037/h0035493

Froger, C., Taconnat, L., Landré, L., Beigneux, K., & Isingrini, M. (2009). Effects of level of processing at encoding and types of retrieval task in mild cognitive impairment and normal aging. *Journal of Clinical and Experimental Neuropsychology*, *31*, 312–321. 10.1080/13803390802112554

Glisky, E. L., Rubin, S. R., & Davidson, P. S. R. (2001). Source memory in older adults: An encoding or retrieval problem?*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1131–1146. 10.1037/0278-7393.27.5.1131

Gross, A. L., Parisi, J. M., Spira, A. P., Kueider, A. M., Ko, J. Y., Saczynski, J. S., . . .Rebok, G. W. (2012). Memory training interventions for older adults: A meta-analysis. *Aging & Mental Health*, *16*, 722–734. 10.1080/13607863.2012.667783

Guttentag, R. E. (1984). The mental effort requirement of cumulative rehearsal: A developmental study. *Journal of Experimental Child Psychology*, *37*, 92–106. 10.1016/0022-0965(84)90060-2

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, *5*, 215–227. 10.1037/0882-7974.5.2.215

Hertzog, C., & Dunlosky, J. (2004). Aging, metacognition, and cognitive control. In B. H.Ross (Ed.), *Psychology of learning and motivation: Advances in research and theory*(Vol. *45*, pp. 215–251). San Diego, CA: Elsevier.

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M.Craik & T. A.Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 417–466). Mahwah, NJ: Erlbaum Publishers.

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *58*, 3–11. 10.1093/geronb/58.1.P3

Howard, D. V., McAndrews, M. P., & Lasaga, M. I. (1981). Semantic priming of lexical decisions in young and old adults. *Journal of Gerontology*, *36*, 707–714. 10.1093/geronj/36.6.707

Hulicka, I. M., & Grossman, J. L. (1967). Age changes and age differences in memory functionings: Proposals for research. *Journal of Gerontology*, *22*, 46–51. 10.1093/geront/7.2_Part_2.46

Hulicka, I. M., Sterns, H., & Grossman, J. (1967). Age-group comparisons of paired-associate learning as a function of paced and self-paced association and response times. *Journal of Gerontology*, *22*, 274–280. 10.1093/geronj/22.3.274

Hultsch, D. F. (1971). Adult age differences in free classification and free recall. *Developmental Psychology*, *4*, 338–342. 10.1037/h0030990

Hultsch, D. F., Hertzog, C., & Dixon, R. A. (1987). Age differences in metamemory: Resolving the inconsistencies. *Canadian Journal of Psychology*, *41*, 193–208. 10.1037/h0084153

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497–514. 10.1016/S0022-5371(81)90138-9

Jacobs, D. M., Rakitin, B. C., Zubin, N. R., Ventura, P. R., & Stern, Y. (2001). Cognitive correlates of mnemonics usage and verbal recall memory in old age. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *14*, 15–22.

Kausler, D. H. (1994). *Learning and memory in normal aging*. San Diego, CA: Academic Press.

Kliegl, R., Smith, J., & Baltes, P. B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology*, *25*, 247–256. 10.1037/0012-1649.25.2.247

Kliegl, R., Smith, J., & Baltes, P. B. (1990). On the locus and process of magnification of age differences during mnemonic training. *Developmental Psychology*, *26,* 894–904. 10.1037/0012-1649.26.6.894

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126,* 349–370. 10.1037/0096-3445.126.4.349

Kuhlmann, B. G., & Touron, D. R. (2012). Mediator-based encoding strategies in source monitoring in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1352–1364. 10.1037/a0027863

Lachman, M. E., & Andreoletti, C. (2006). Strategy use mediates the relationship between control beliefs and memory performance for middle-aged and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *61*, 88–94. 10.1093/geronb/61.2.P88

Lee, I. A., & Preacher, K. J. (2013, September). *Calculation for the test of the difference between two dependent correlations with one variable in common* [Computer software]. Retrieved from http://quantpsy.org

Lemaire, P. (2010). Cognitive strategy variations during aging. *Current Directions in Psychological Science*, *19*, 363–369. 10.1177/0963721410390354

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology*, *42,* 333–376. 10.1146/annurev.ps.42.020191.002001

Loewen, E. R., Shaw, R. J., & Craik, F. I. M. (1990). Age differences in components of metamemory. *Experimental Aging Research*, *16*, 43–48. 10.1080/03610739008253873

Mason, S. E., & Smith, A. D. (1977). Imagery in the aged. *Experimental Aging Research, 3*, 17–32. 10.1080/03610737708257085

Mulligan, N. W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 54–69. 10.1037/0278-7393.25.1.54

Mungas, D., Ehlers, C. L., & Blunden, D. (1991). Age differences in recall and information processing verbal and spatial learning. *Canadian Journal on Aging*, *10*, 320–332. 10.1017/S0714980800011351

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging*, *22*, 202–208. 10.1037/0882-7974.22.1.202

Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Kreuger, S. (2005). Divided attention in younger and older adults: Effects of strategy and relatedness on memory performance and secondary task costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 520–537. 10.1037/0278-7393.31.3.520

Park, D. C. (2000). The basic mechanisms accounting for age-related decline in cognitive function. In D. C.Park & N.Schwarz (Eds.), *Cognitive aging: A primer* (Vol. *11*, pp. 3–19). Philadelphia, PA: Psychology Press.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*, 299–320. 10.1037/0882-7974.17.2.299

Park, D. C., Smith, A. D., Dudley, W. N., & Lafronza, V. N. (1989). Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1185–1191. 10.1037/0278-7393.15.6.1185

Price, J., Hertzog, C., & Dunlosky, J. (2008). Age-related differences in strategy knowledge updating: Blocked testing produces greater improvements in metacognitive accuracy for younger than older adults. *Aging, Neuropsychology, and Cognition: A Journal on Normal and Dysfunctional Development*, *15*, 601–626. 10.1080/13825580801956225

Rankin, J. L., Karol, R., & Tuten, C. (1984). Strategy use, recall, and recall organization in young, middle-aged, and elderly adults. *Experimental Aging Research*, *10*, 193–196. 10.1080/03610738408258463

Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, *5*, 597–614. 10.3758/BF03208837

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*, 45–48. 10.1037/h0031355

Rowe, E. J., & Schnore, M. M. (1971). Item concreteness and reported strategies in paired-associate learning as a function of age. *Journal of Gerontology*, *26*, 470–475. 10.1093/geronj/26.4.470

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428. 10.1037/0033-295X.103.3.403

Schneider, W., & Uhl, C. (1990). Metagedächtnis, Strategienutzung und Gedächtnisleistung: Vergleichende Analysen bei Kindern, jüngeren Erwachsenen und alten Menschen [Metamemory, strategies, and memory performance: Comparative analysis among children, young adults, and older adults]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *22*, 22–41.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*, 405–410. 10.1111/1467-9280.00076

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, *126*, 71–92. 10.1037/0096-3445.126.1.71

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G.Halford & T.Simon (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.

Singer, T., Lindenberger, U., & Baltes, P. B. (2003). Plasticity of memory for new learning in very old age: A story of major loss?*Psychology and Aging*, *18*, 306–317. 10.1037/0882-7974.18.2.306

Stricker, J. L., Brown, G. G., Wixted, J., Baldo, J. V., & Delis, D. C. (2002). New semantic and serial clustering indices for the California Verbal Learning Test-Second Edition: Background, rationale, and formulae. *Journal of the International Neuropsychological Society*, *8*, 425–435. 10.1017/S1355617702813224

Taconnat, L., Baudouin, A., Fay, S., Clarys, D., Vanneste, S., Tournelle, L., & Isingrini, M. (2006). Aging and implementation of encoding strategies in the generation of rhymes: The role of executive functions. *Neuropsychology*, *20*, 658–665. 10.1037/0894-4105.20.6.658

Taconnat, L., Raz, N., Toczé, C., Bouazzaoui, B., Sauzéon, H., Fay, S., & Isingrini, M. (2009). Ageing and organisation strategies in free recall: The role of cognitive flexibility. *European Journal of Cognitive Psychology*, *21*, 347–365. 10.1080/09541440802296413

Tournier, I., & Postal, V. (2011). Strategy selection and aging: Impact of item concreteness in paired-associate task. *Aging, Neuropsychology, and Cognition: A Journal on Normal and Dysfunctional Development*, *18*, 195–213. 10.1080/13825585.2010.525623

Touron, D. R. (2015). Memory avoidance by older adults: When 'old dogs' won't perform their 'new tricks'. *Current Directions in Psychological Science*, *24*, 170–176. 10.1177/0963721414563730

Touron, D. R., & Hertzog, C. (2004). Strategy shift affordance and strategy choice in young and older adults. *Memory & Cognition*, *32*, 298–310. 10.3758/BF03196860

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. 10.3758/BF03192720

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*, 289–335. 10.1016/j.jml.2003.10.003

Verhaeghen, P., & Marcoen, A. (1994). Production deficiency hypothesis revisited: Adult age differences in strategy use as a function of processing resources. *Aging, Neuropsychology, and Cognition: A Journal on Normal and Dysfunctional Development*, *1*, 323–338. 10.1080/13825589408256585

Verhaeghen, P., & Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and Aging*, *11*, 164–178. 10.1037/0882-7974.11.1.164

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging*, *7*, 242–251. 10.1037/0882-7974.7.2.242

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. New York, NY: Psychological Corporation.

Wegesin, D. J., Jacobs, D. M., Zubin, N. R., Ventura, P. R., & Stern, Y. (2000). Source memory and encoding strategy in normal aging. *Journal of Clinical and Experimental Neuropsychology*, *22*, 455–464. 10.1076/1380-3395(200008)22:4;1-0;FT455

West, R. L., Dark-Freudeman, A., & Bagwell, D. K. (2009). Goals-feedback conditions and episodic memory: Mechanisms for memory gains in older and younger adults. *Memory*, *17*, 233–244. 10.1080/09658210802236407

West, R. L., & Thorn, R. M. (2001). Goal-setting, self-efficacy, and memory performance in older and younger adults. *Experimental Aging Research*, *27*, 41–65. 10.1080/036107301750046133

Witte, K. L., Freund, J. S., & Brown-Whistler, S. (1993). Adult age differences in free recall and category clustering. *Experimental Aging Research*, *19*, 15–28. 10.1080/03610739308253920

Witte, K. L., Freund, J. S., & Sebby, R. A. (1990). Age differences in free recall and subjective organization. *Psychology and Aging*, *5*, 307–309. 10.1037/0882-7974.5.2.307

Worden, P. E., & Meggison, D. L. (1984). Aging and the category-recall relationship. *Journal of Gerontology*, *39*, 322–324. 10.1093/geronj/39.3.322

Zivian, M., & Darjes, R. W. (1983). Free recall by in-school and out-of-school adults: Performance and metamemory. *Developmental Psychology*, *19*, 513–520. 10.1037/0012-1649.19.4.513

---

[1] Note that a somewhat more lenient accuracy criterion of 70% was set for the processing-task component of the span tasks instead of the typical 85% criterion (Unsworth et al., 2005) because 16.7% of the younger and 30.5% of the older adults performed below 85% accuracy on the symmetry judgments. Importantly, processing-task accuracy and span scores were positively correlated on both span tasks in both age groups, suggesting no speed–accuracy trade-off. One younger adult in the individual-words condition did not reach 70% accuracy on the processing tasks

of either span task and was not included in analyses involving working memory capacity. Next to the expected main effect of age group, there were unexpected significant main effects of presentation format on both reading-span, $F(1, 246) = 10.14$, $p = .002$, $\eta p^2 = .04$, and symmetry-span scores, $F(1, 223) = 6.04$, $p = .015$, $\eta p^2 = .03$. Although the presentation format of the working memory tasks was equal for both format conditions (brief individual letter/matrix presentations), participants studying the word lists in the individual-words format outperformed those in the whole-list format. This was equally true for younger and older participants, both interaction *F*s < 1. We do not believe that these span-score differences reflect actual differences in cognitive ability of these (randomly assigned) conditions, which did not differ in processing speed or vocabulary, all *F*s ≤ 1.96 for the presentation-format main effect or interaction. Instead, the higher similarity of the individual-words format to the presentation format of the working memory tasks may have facilitated performance for participants in these conditions. Note that these differences, if real, would work against the observed benefits of the whole-list format. These differences do not influence any of the correlational analyses, which were conducted separately by presentation format.

[2] During study of this third list (clustering again instructed), participants had to classify tones played over headphones as high- or low-pitched by pressing corresponding keys. We added this dual-task manipulation to measure cognitive costs of semantic clustering and to examine if performance in the individual-words format is more sensitive to dual-task interference. Performance results (recall and clustering indices) mirrored those of the second clustering-instructed list in both age groups (of course lower performance). That is, the whole-list conditions outperformed the individual-words conditions under dual-task. However, in both age groups the whole-list conditions responded slower on the tone task suggesting that they devoted more resources to encoding than the individual-words condition (no condition differences in accuracy [at ceiling]). Given this differential trade-off between conditions, the improved performance in the whole-list conditions cannot be unambiguously attributed to the format support of clustering. We thus refrain from further discussing and interpreting these results.

[3] The *Adjusted Ratio of Clustering* (ARC; Roenker et al., 1971) is another popular measure of relative output clustering and may be preferred because it corrects for chance clustering. However, the ARC score is incomputable in many cases (i.e., when only words from one category are recalled) and, further, its distribution strongly deviated from normality, a common problem with this measure (Mulligan, 1999) that prevents the use of powerful parametric tests. In the present data, RR outperformed ARC in terms of computability and distributional properties (skew and kurtosis) and was preferred for analysis. Importantly, mean patterns in ARC closely mirrored those of RR and (parametric) analyses of ARC would lead to similar conclusions. Note that one older adult in the whole-list condition only recalled one word on the second, clustering-instructed list, rendering RR incomputable (and LBC = 0). To include all participants in the repeated-measures analysis of RR, we set RR = 0 for this missing value. Not including the participant in the within-subjects ANOVA on RR scores yielded the same pattern of results.