Psychology and Aging
2015, Vol. 30, No. 1, 68–73

0882-7974/15/$12.00 http://dx.doi.org/10.1037/a0038669

BRIEF REPORT

Doubly Blessed: Older Adults Know More Vocabulary and Know Better What They Know

Gitit Kavé
The Open University

Vered Halamish
Tel Aviv University and Yezreel Valley College

This study examined age-related differences in the ability to judge one's vocabulary. Young, middle-age, and older adults completed a multiple-choice test of vocabulary, judged their confidence in each answer, and estimated their overall performance. Older adults performed better and were more confident in their knowledge than were the other 2 groups. Importantly, relative to young adults, older adults demonstrated better calibration both on item-by-item confidence judgments and on global estimates. Resolution, as defined by correlations between item-by-item performance and confidence judgments, was age-invariant. We suggest that age-related accumulation of vocabulary is accompanied by enhanced perception of mastery in one's knowledge.

Keywords: vocabulary, general knowledge, confidence, metacognition, calibration

Older persons often complain of difficulties in word retrieval (Burke, 2006; Condret-Santi et al., 2013; Goral, 2004; Ossher, Flegal, & Lustig, 2013; Wingfield & Stine-Morrow, 2000). In a recent study of the tip-of-the-tongue (TOT) experience, Salthouse and Mandell (2013) reported that neither vocabulary knowledge nor measures of episodic memory could account for the association between age and TOT incidence. Instead, they suggested that the age-related increase in TOT might reflect difficulties in metacognitive monitoring. TOTs involve prediction of the likelihood of overcoming a retrieval failure (Schwartz & Metcalfe, 2011), and overconfidence in one's vocabulary may lead to more TOTs. No previous research has investigated differences in confidence judgment of vocabulary knowledge across the adult life span, and this is the purpose of the current study.

Research on metacognitive monitoring in old age has focused primarily on episodic memory and judgment of learning (e.g., Halamish, McGillivray, & Castel, 2011; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Stine-Morrow, Shake, Miles, & Noh, 2006). According to Hertzog and Dunlosky (2011), aging affects episodic memory, but it does not influence monitoring accuracy on tasks of episodic memory. In these studies, participants are typically asked to study paired associates, lists of words, or short passages, and then to predict the likelihood of later recall. Both young and older adults are equally correct in predicting which items they will recall later, demonstrating similar resolution (e.g., the relative correspondence between performance and confidence level for each item). Participants are also asked after the recall task to judge how many items they remembered (Hertzog, Sinclair, & Dunlosky, 2010) or how many items they forgot (Halamish et al., 2011). No age-related differences were found in the ability to monitor performance that has already occurred (Baker, Dunlosky, & Hertzog, 2010).

Nevertheless, other studies found age-related differences in monitoring on tasks of episodic memory, and emphasized that if differences are found, they are more noticeable on tasks of episodic memory than on tasks of semantic memory. For example, Souchay, Moulin, Clarys, Taconnat, and Isingrini (2007) compared the correlation between performance and ratings on a paired-associate episodic memory task and on a general knowledge task. Participants were asked to judge whether they would be able to recognize the correct answer if it were presented among some likely but incorrect choices. Following this assessment, a multiple-choice recognition test was administered to examine prediction accuracy. Younger adults performed better than older adults on both the episodic and the semantic tasks, but age-differences in monitoring resolution emerged only on the episodic task. Dodson, Bawa, and Krueger (2007) examined item-by-item calibration (e.g., the deviation rather than the correlation between actual performance and confidence rating). They used a cued-recall episodic task as well as a task that tested semantic knowledge. On the semantic task, young (ages 17–22 years) and older adults (ages

This article was published Online First January 19, 2015.

Gitit Kavé, Department of Education and Psychology, The Open University; and Vered Halamish, School of Psychological Sciences, Tel Aviv University, and Department of Psychology, Yezreel Valley College.

Supported in part by a research fund (Grant 43018) from the Open University. We thank Dafna Arnheim, Erez Ben Natan, Irit Feldman, Amira Glicksberg, Zohar Kaufman, Talia Pertman, Neta Sagiv, and Adi Spector for their help in data collection and preparation. Special thanks are also extended to Yoseph Hockman. We declare no conflict of interests.

Correspondence concerning this article should be addressed to Gitit Kavé, Department of Education and Psychology, The Open University, The Dorothy de Rothschild Campus, 1 University Road, Ra'anana 43537, Israel. E-mail: gkave@012.net.il

60–77 years) were presented with 60 general knowledge questions in a two-forced-choice format, with confidence level assessed after each choice. Relative to the young participants, older adults performed worse and were worse at judging their performance on the episodic task, but they performed better and demonstrated better calibration on the semantic task.

It is possible that level of performance in response to general knowledge questions interacts with age-related changes in monitoring, so that older adults would show good monitoring only when they perform well. Marquié and Huet (2000) presented adults in three age groups (18–30, 40–52, and 61–77) with questions such as "Which planet of the solar system looks brightest from earth?" and then asked them to judge whether they would know the answer (Phase 1). On Phase 2, the same questions were presented and participants were asked to select the correct response from four alternatives as well as to judge how confident they were in their selection. Results showed that young adults provided higher assessments of their performance on Phase 1, but there were no age-related differences in confidence ratings on Phase 2. Nevertheless, resolution in both phases was age-invariant. Importantly, Marquié and Huet (2000) pretested their questions to equate the number of correct answers across groups, implying that older adults do not have to outperform younger adults to show comparable resolution. Dodson et al. (2007) examined age differences in calibration as a function of level of difficulty of general knowledge questions. On easy questions, for which there were no age-related differences in performance, calibration was comparable. On the difficult questions, for which older adults knew more answers, the older group demonstrated better calibration. Souchay et al. (2007) found that even when older adults performed worse than younger adults on the knowledge test (most likely due to a significantly lower level of education), groups did not differ in their assessment of knowledge. Dahl, Allwood, and Hagberg (2009) analyzed responses to 10 questions of general knowledge in 1,384 participants between ages 60 and 93. Participants were asked to choose the correct answer from two alternatives, and then judge how confident they were in their response. In this study, both performance and confidence levels declined with age. However, no significant decrease in either resolution or calibration was seen. These findings suggest that on tasks of semantic memory there are no significant differences in various monitoring abilities, whether or not participants know the correct answers.

Previous studies on judgment of semantic memory in old age have looked at general knowledge, whereas the current study is the first to focus specifically on word meanings. Vocabulary is no doubt a subset of general knowledge, but knowledge tests often require retrieval of very specific pieces of information (e.g., Venus in response to the Marquié & Huet's, 2000, question quoted above, or Mozart in response to Dodson et al.'s, 2007, question "What is the last name of the composer who wrote the opera Don Giovanni?"). Performance on vocabulary tests depends to some degree on retrieval abilities (Kavé & Yafé, 2014), but despite allegedly poor performance on tests of retrieval, older people still outperform younger adults on tests of vocabulary (Bowles, Grimm, & McArdle, 2005; Bowles & Salthouse, 2008; Kavé & Yafé, 2014; Kemper & Sumner, 2001; Verhaeghen, 2003). It is assumed that age is associated with an accumulation of vocabulary due to education, incremental reading, or life experience (Uttl, 2002). Better performance on vocabulary tests may also reflect a cohort effect favoring the earlier born, either because they used to read more or because the tests themselves include obsolete items (Verhaeghen, 2003). This greater mastery may result in an increased sense of confidence in one's knowledge. Alternatively, the decrease in cognitive abilities that accompanies aging could lead to a reduction in one's level of cognitive self-efficacy (e.g., Stine-Morrow et al., 2006), and thus to greater underconfidence among older relative to younger adults.

The purpose of the current study was to examine age-related differences in monitoring of vocabulary knowledge. Young, middle-age, and older adults completed a four-alternative multiple-choice test of vocabulary, assessed their confidence level in each response, and provided a global estimate of performance.

Method

Participants

One-hundred fifty participants were recruited for the study, 50 in each of three age groups: young, middle-age, and older adults (see Table 1). The younger and older adults participated in Kavé and Yafé (2014) and were included in the current analyses if they were also asked to provide confidence ratings. All participants reported having no history of learning disorders, psychiatric disturbances, neurological disease, or head trauma. There was a significant difference in education across groups, $F(2, 147) = 7.566$, $p < .01$, originating from a slightly higher level of education in the middle-age group. Post hoc tests revealed no significant difference in education level between young and older adults. Older individuals reported no subjective decline in cognitive abilities and scored within the normal range (27–30) on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975). Participant recruitment adhered to institutional research guidelines.

Vocabulary Task

We used a paper-and-pencil multiple-choice Hebrew vocabulary test developed by Kavé and Yafé (2014). The original test consisted of 40 target nouns, which were on average 4.63 letters ($SD = 1.08$) and 2.30 syllables ($SD = 0.56$) in length. Target nouns were relatively infrequent in the language, with a mean rated frequency of 2.29 ($SD = 0.76$) on a 1 (*infrequent*) to 5 (*frequent*) scale, as determined by 20 individuals (10 women; age range: 30–60; $M_{age} = 44.5$, $SD = 9.87$; $M_{education} = 17.0$ years, $SD = 2.58$) who did not participate in the current study. Each item had four alternative responses that included either one word or a short phrase, and the three distracters were closely related to the correct answer. Target nouns were emphasized in bold font within the phrase "What is the meaning of . . .?" Participants were asked to mark the correct interpretation from the given alternatives on an answering sheet. No time limitation was placed. Following Kavé and Yafé (2014), five items were deleted from the test due to errors in construction of distracters. Thus, although participants were tested on 40 items, the analyses reported below were conducted on 35 items.

Table 1
Sample Characteristics, Correct Responses, and Judgment, by Age Group

	Young adults	Middle age	Older adults
No. of participants (% women)	50 (50%)	50 (56%)	50 (56%)
Age (years)			
M	24.58	51.56	75.06
SD	2.43	8.21	3.64
Range	20 to 29	35 to 60	70 to 84
Education (years)			
M	14.06	15.86	14.66
SD	1.42	2.62	2.80
Range	12 to 17	12 to 21	8 to 22
Correct responses (%)[a]			
M	61.37	59.03	80.11
SD	16.58	10.99	10.33
Range	22.86 to 91.43	28.57 to 80.00	54.29 to 97.14
Item-by-item judgment (scale)[b]			
M	3.95	4.76	5.85
SD	1.36	1.13	0.89
Range	1.00 to 6.49	1.86 to 6.77	3.46 to 6.94
Item-by-item judgment (%)[c]			
M	49.12	62.70	80.88
SD	22.66	18.76	14.87
Range	0 to 91.43	14.29 to 96.19	40.95 to 99.05
Item-by-item calibration score[d]			
M	−12.40	3.64	.16
SD	15.87	14.65	11.18
Range	−35.71 to 20.48	−28.57 to 41.09	−31.83 to 20.48
Resolution (gamma)[e]			
M	.55	.54	.51
SD	.24	.37	.51
Range	0to1.00	−1.00 to 0.95	−1.00 to 1.00
Global estimate (%)[f]			
M	41.43	55.56	73.07
SD	25.43	24.08	20.08
Range	0 to 87.50	12.50 to 95.00	25 to 100
Global calibration score[g]			
M	−19.56	−4.79	−6.93
SD	17.44	19.45	16.61
Range	−53.21 to 15.00	−43.57 to 27.50	−49.29 to 27.14

[a] Correct responses is the percent correct out of 35. [b] Item-by-item judgment (scale) is the mean of one's confidence ratings for individual responses. [c] Item-by-item judgment (%) is the mean of one's ratings for individual responses converted to percentages. [d] Item-by-item calibration score is the percent correct minus the item-by-item judgment converted to percentages. [e] Resolution (gamma) is the gamma correlations between item-by-item scores and confidence judgment. [f] Global estimate (%) is the percent estimate out of 40. [g] Global calibration score is the percent correct minus the global estimate.

Confidence Ratings

Item-by-item confidence judgments. Next to each item appeared a scale from 1 (*not at all confident*) to 7 (*completely confident*). Participants were asked to report their judgment of confidence in each response immediately after providing the response.

Global judgments. After completing the test, participants were asked to place a number in a blank space that appeared at the end of the following sentence: "On the vocabulary test there were 40 questions. In your view, how many questions did you answer correctly?"

Results

Despite group differences in education, key variables did not correlate with the number of years of schooling and thus the analyses reported below did not control for this variable.

Performance Accuracy

Across groups, participants failed to provide 39 responses (.007% of 5,250 possible responses). Table 1 presents the percentage of correct responses out of 35 in each age group. A one-way analysis of variance showed a significant group difference in scores, $F(2, 147) = 39.88$, mean square error (*MSE*) = .003, $p < .01$, $\eta_p^2 = .35$. A post hoc Tukey's test ($p < .05$) revealed significantly better performance in the older group than in the young group or the middle-age group, with no difference between the young and middle-age groups.

Item-by-Item Confidence Judgments

Across groups, participants failed to provide 38 item-by-item confidence judgments (.007% of 5,250 possible responses). Item-by-item confidence judgments were averaged for every participant. Age groups differed in their item-by-item confidence judgments,

$F(2, 147) = 35.048$, $MSE = .005$, $p < .01$, $\eta_p^2 = .32$ (see Table 1). A post hoc Tukey's test ($p < .05$) showed that each group differed significantly in its judgment from the other two groups, with younger adults providing the lowest judgments, middle-age adults providing higher judgments than the younger adults and lower than the older adults, and older adults demonstrating the greatest confidence.

After analyzing the mean of the item-by-item confidence judgments, we examined its correspondence to actual test performance. First, we converted the average item-by-item judgment of each participant from the 1–7 scale to percentages by subtracting 1 from the mean rating, dividing the difference in 6, and multiplying it by 100. We then calculated an item-by-item calibration score for each participant by calculating the difference between performance accuracy (in terms of percentage of correct answers out of 35) and the converted mean confidence judgments. Group means of calibration scores are presented in Table 1. Note that the smaller the absolute value of the calibration score is, the better is the correspondence between test performance and confidence judgment.

Young participants were underconfident in their performance, and their item-by-item calibration score was statistically different from zero, $t(49) = 5.52$, $p < .01$, Cohen's $d = 1.58$. Middle-age adults were calibrated in their judgments, so that their item-by-item calibration score was not statistically different from zero, $t(49) = 1.76$, $p = .085$, Cohen's $d = .50$, and the same was true for older adults, $t(49) = .10$, $p = .920$, Cohen's $d = .03$. In addition, calibration scores differed significantly across groups, $F(2, 147) = 17.805$, $MSE = .02$, $p < .01$, $\eta_p^2 = .20$. A post hoc Tukey's test ($p < .05$) revealed that younger adults were significantly less calibrated than either of the other two groups, with no difference in item-by-item calibration between the middle-age and older groups. Figure 1 shows patterns of correct responses and judgments across groups.

Resolution

Next, we examined the extent to which item-by-item confidence judgments discriminate between correct and incorrect responses. We thus calculated resolution in terms of gamma correlations between item-by-item response accuracy and item-by-item confidence judgments (see Table 1). There was no significant group difference in gamma correlations, $F(2, 147) = .133$, $MSE = .138$, $p = .876$, $\eta_p^2 = .00$.

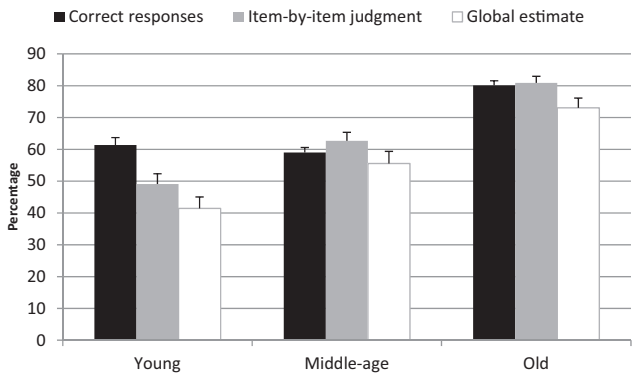

Figure 1. Percentages of correct responses and judgments, by age group.

Global Judgments

One young adult, 10 middle-age adults, and six older adults did not provide a global estimate of their performance (11% of 150 participants). To convert global estimates to percentages, we divided participants' responses in 40 (the original number of items to which participants referred when requested to provide the estimate). Calibration was calculated by subtracting actual proportion of correct responses from these estimated proportions. As seen in Table 1, all groups were underconfident. In the younger and older groups the underconfidence bias differed statistically from zero, young: $t(48) = 7.85$, $p < .01$, Cohen's $d = 2.27$, older: $t(43) = 2.77$, $p < .01$, Cohen's $d = .84$, whereas in the middle-age group this bias did not differ significantly from zero, $t(40) = 1.56$, $p = .127$, Cohen's $d = .48$. Analysis of the degree of underconfidence across groups revealed a significant effect, $F(2, 130) = 9.24$, $MSE = .032$, $p < .01$, $\eta^2 = .12$. A post hoc Tukey's test ($p < .05$) showed that the young adults were significantly more underconfident than either of the other two groups, with no significant difference between middle-age and older adults.

Discussion

In this study we investigated how young, middle-age, and older adults judge their vocabulary knowledge, which was tested on a four-alternative multiple-choice questionnaire. As expected, older adults demonstrated greater vocabulary knowledge than did the other two groups. In addition, their confidence in their responses was highest. These results differ from Dahl et al.'s (2009) findings of decreased performance on a test of general knowledge as well as decreased confidence levels with increased age. Yet, Dahl et al. measured age effects from ages 60 to 90, with no younger or middle-age adults. Our results also contrast with Marquié and Huet's (2000) findings, according to which age groups differ neither in level of performance nor in level of confidence. However, Marquié and Huet selected questions to match performance across age groups and assessed confidence with a markedly different procedure from the one that we used.

As seen in previous research of both episodic (Hertzog & Dunlosky, 2011) and semantic memory (Souchay et al., 2007), we found that resolution, or the ability to discriminate between correct and incorrect responses, was age-invariant. However, when we examined calibration, or the ability to predict the overall level of performance on the task, we found that older adults demonstrated a significant advantage. This pattern of results emerged both for the item-by-item confidence judgments and for the global judgments. These results resemble Dodson et al.'s (2007) report that older adults were better calibrated than were younger adults on difficult general knowledge questions. In fact, in the current study, older adults demonstrated perfect item-by-item calibration, so that their item-by-item calibration score did not differ significantly from zero. This finding shows that they were correctly confident in their word knowledge, unlike younger adults who were underconfident relative to older adults as well as relative to chance level.

In contrast to the perfect calibration found in the older group on judgments provided immediately after response selection, older adults were biased in their overall estimate. Nevertheless, they were less biased than were younger adults, who at the end of the test estimated that they provided almost 20% fewer correct responses than they had actually provided. Thus, although older

adults demonstrated an underconfidence bias (7%), which was statistically different from zero, they were still better calibrated than were young adults on global judgments. Note that our global estimates reflect judgments aggregated over multiple test items, and as such they might be less reliable than item-by-item judgments (Schraw, 2009). In principle, global estimates could reflect a general appraisal of one's baseline vocabulary level as well as knowledge updating on the basis of the experience on the actual task (e.g., Price, Hertzog, & Dunlosky, 2008).

According to Dunning, Johnson, Ehrlinger, and Kruger (2003), people base perceptions of their own performance on their preconceived notions of their skills. It is possible that older adults have gained a feeling of mastery of their vocabulary knowledge over a lifetime of word usage. To test this possibility, future research should assess perceptions of vocabulary self-efficacy, as well as the effect that these perceptions might have on both item-by-item and global calibration. These perceptions, together with one's calibration ability, might also affect age differences in the experience of TOT. With regard to TOT, Salthouse and Mandell (2013) have suggested that TOT incidence is explained by age-related difficulties in metacognitive monitoring, assuming that metacognitive skills decrease with age. Our results show that the relevant metacognitive skills increase rather than decrease with age. Future research will have to determine whether older adults' perceptions about the level of their knowledge of the elusive words affect the frequency of TOT occurrence.

Confidence judgment is determined not only by participants' perception of their lifelong mastery of vocabulary but also by their experience while performing the task at hand. This experience reflects various cues, such as familiarity with task material or the amount and ease of deliberation in response selection (Koriat, 2012). Our test included four alternatives, with both target items and alternatives consisting of relatively infrequent words. Older adults might have been more familiar not only with target items and correct responses but also with the distracters. Their perfect item-by-item calibration could thus reflect the fact that they eliminated the incorrect distracters more easily. Koriat (2012) has also suggested that when confidence judgments are high, there appears to be a greater degree of overconfidence. Although older adults gave high confidence judgments, the lack of certainty about the distracters in the current study explains why all participants, including older adults, experienced under- rather than overconfidence in their global estimates. Future research will have to measure level of familiarity with response alternatives, and examine whether it affects age differences in global calibration.

Due to the cross-sectional nature of our study, we cannot rule out cohort effects on word knowledge. The relatively infrequent words that appeared in our test could have been more frequent in Hebrew in previous generations and could thus have been more familiar to older adults than they were to younger adults. However, the stimuli were not foreign to all younger adults, and some scored above 90% correct. In addition, although knowledge is most likely affected by cohort, calibration abilities are less likely to be affected by cohort-related language changes.

Conclusion

We suggest that the accumulation of vocabulary that comes with age is accompanied by enhanced perception of mastery in knowledge and subsequently by better calibration in judgment of knowledge of word meanings. Dunning et al. (2003) described poor performers as doubly cursed: They do not know and do not know that they do not know. We found that older adults are doubly blessed: Not only do they know the meaning of more words, but they also know better what they know. This double blessing might be specific to vocabulary, which benefits from a lifetime of practice and appears to be less susceptible to the deleterious effects of cognitive aging.

References

Baker, J. M. C., Dunlosky, J., & Hertzog, C. (2010). How accurately can older adults evaluate the quality of their text recall? The effect of providing standards on judgment accuracy. *Applied Cognitive Psychology, 24,* 134–147. http://dx.doi.org/10.1002/acp.1553

Bowles, R. P., Grimm, K. J., & McArdle, J. J. (2005). A structural factor analysis of vocabulary knowledge and relations to age. *The Journals of Gerontology. Series B: Psychological Sciences, 60,* P234–P241. http://dx.doi.org/10.1093/geronb/60.5.P234

Bowles, R. P., & Salthouse, T. A. (2008). Vocabulary test format and differential relations to age. *Psychology and Aging, 23,* 366–376. http://dx.doi.org/10.1037/0882-7974.23.2.366

Burke, D. M. (2006). Representation and aging. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 193–206). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195169539.003.0013

Condret-Santi, V., Barbeau, E. J., Matharan, F., Le Goff, M., Dartigues, J.-F., & Amieva, H. (2013). Prevalence of word retrieval complaint and prediction of dementia in a population-based study of elderly subjects. *Dementia and Geriatric Cognitive Disorders, 35,* 313–324. http://dx.doi.org/10.1159/000342594

Dahl, M., Allwood, C. M., & Hagberg, B. (2009). The realism in older people's confidence judgments of answers to general knowledge questions. *Psychology and Aging, 24,* 234–238. http://dx.doi.org/10.1037/a0014048

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging, 22,* 122–133. http://dx.doi.org/10.1037/0882-7974.22.1.122

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12,* 83–87. http://dx.doi.org/10.1111/1467-8721.01235

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

Goral, M. (2004). First-language decline in healthy aging: Implications for attrition in bilingualism. *Journal of Neurolinguistics, 17,* 31–52. http://dx.doi.org/10.1016/S0911-6044(03)00052-6

Halamish, V., McGillivray, S., & Castel, A. D. (2011). Monitoring one's own forgetting in younger and older adults. *Psychology and Aging, 26,* 631–635. http://dx.doi.org/10.1037/a0022852

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological Science, 20,* 167–173. http://dx.doi.org/10.1177/0963721411409026

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17,* 209–225. http://dx.doi.org/10.1037/0882-7974.17.2.209

Hertzog, C., Sinclair, S. M., & Dunlosky, J. (2010). Age differences in the monitoring of learning: Cross-sectional evidence of spared resolution

across the adult life span. *Developmental Psychology, 46,* 939–948. http://dx.doi.org/10.1037/a0019812

Kavé, G., & Yafé, R. (2014). Performance of younger and older adults on tests of word knowledge and word retrieval: Independence or interdependence of skills? *American Journal of Speech-Language Pathology, 23,* 36–45. http://dx.doi.org/10.1044/1058-0360(2013/12-0136)

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging, 16,* 312–322. http://dx.doi.org/10.1037/0882-7974.16.2.312

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review, 119,* 80–113. http://dx.doi.org/10.1037/a0025648

Marquié, J. C., & Huet, N. (2000). Age differences in feeling-of-knowing and confidence judgments as a function of knowledge domain. *Psychology and Aging, 15,* 451–461. http://dx.doi.org/10.1037/0882-7974.15.3.451

Ossher, L., Flegal, K. E., & Lustig, C. (2013). Everyday memory errors in older adults. *Aging, Neuropsychology, and Cognition: A Journal on Normal and Dysfunctional Development, 20,* 220–242. http://dx.doi.org/10.1080/13825585.2012.690365

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. http://dx.doi.org/10.1080/13825580801956225

Salthouse, T. A., & Mandell, A. R. (2013). Do age-related increases in tip-of-the-tongue experiences signify episodic memory impairments? *Psychological Science, 24,* 2489–2497. http://dx.doi.org/10.1177/0956797613495881

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4,* 33–45. http://dx.doi.org/10.1007/s11409-008-9031-3

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition, 39,* 737–749. http://dx.doi.org/10.3758/s13421-010-0066-8

Souchay, C., Moulin, C. J. A., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and Cognition: An International Journal, 16,* 769–784. http://dx.doi.org/10.1016/j.concog.2006.11.002

Stine-Morrow, E. A. L., Shake, M. C., Miles, J. R., & Noh, S. R. (2006). Adult age differences in the effects of goals on self-regulated sentence processing. *Psychology and Aging, 21,* 790–803. http://dx.doi.org/10.1037/0882-7974.21.4.790

Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology, 24,* 1123–1137. http://dx.doi.org/10.1076/jcen.24.8.1123.8375

Verhaeghen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Aging, 18,* 332–339. http://dx.doi.org/10.1037/0882-7974.18.2.332

Wingfield, A., & Stine-Morrow, E. A. L. (2000). Language and speech. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 359–416). Mahwah, NJ: Erlbaum.

Received July 3, 2014
Revision received December 3, 2014
Accepted December 10, 2014 ■