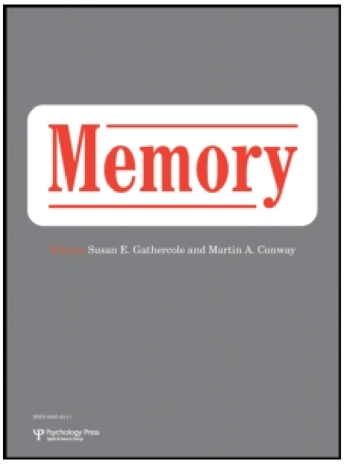

Memory
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683358

Age-related differences in guessing on free and forced recall tests

Mark J. Huff[a]; Michelle L. Meade[a]; Keith A. Hutchison[a]
[a] Montana State University, Bozeman, MT, USA

Online publication date: 15 June 2011

To cite this Article Huff, Mark J. , Meade, Michelle L. and Hutchison, Keith A.(2011) 'Age-related differences in guessing on free and forced recall tests', Memory, 19: 4, 317 — 330
To link to this Article: DOI: 10.1080/09658211.2011.568494
URL: http://dx.doi.org/10.1080/09658211.2011.568494

MEMORY, 2011, 19 (4), 317–330

Psychology Press
Taylor & Francis Group

Age-related differences in guessing on free and forced recall tests

Mark J. Huff, Michelle L. Meade, and Keith A. Hutchison

Montana State University, Bozeman, MT, USA

This study examined possible age-related differences in recall, guessing, and metacognition on free recall tests and forced recall tests. Participants studied categorised and unrelated word lists and were asked to recall the items under one of the following test conditions: standard free recall, free recall with a penalty for guessing, free recall with no penalty for guessing, or forced recall. The results demonstrated interesting age differences regarding the impact of liberal test instructions (i.e., forced recall and no penalty) relative to more conservative test instructions (i.e., standard free recall and penalty) on memory performance. Specifically, once guessing was controlled, younger adults' recall of categorised lists varied in accordance with test instructions while older adults' recall of categorised lists did not differ between conservative and liberal test instructions, presumably because older adults approach standard free recall tests of categorised lists with a greater propensity towards guessing than young adults.

Keywords: Ageing; Free recall; Forced recall; Confidence ratings.

Forced recall requires participants to produce a predetermined number of responses on a recall test, guessing if necessary, and has been used to inform the nature of responding on free recall tests (cf. Erdelyi & Becker, 1974; Koriat & Goldsmith, 1996). Free recall allows participants to determine their own criterion for reporting and withholding items, so by comparing responses on forced recall (where guessing is required) to responses on free recall (where guessing is not controlled), one can examine the influence of guessing on free recall tests. Guessing in the current context refers to producing items on a memory test by some means other than recollection. The goal of the current study is to examine possible age-related differences in guessing on free recall tests by comparing young and older adults' recall and metacognition for categorical and unrelated word lists across various instructional manipulations (free recall, penalty against guessing, encouragement to guess, and forced recall).

The use of forced recall to inform the nature of responding on free recall tests is not new. The original rationale for requiring participants to report guesses on a free recall test was based on the assumptions of generate-recognise models of recall (e.g., Anderson, 1983; Anderson & Bower, 1972) which hypothesised that participants complete a recall test by generating possible items and then selecting a subset of the generated items to report. Forced recall procedures were adopted to equate response criteria between participants and ultimately determine the role of response bias and guessing on recall performance (Cofer, 1967; Erdelyi, Finks, & Feigin-Pfau, 1989; Ritter & Buschke, 1974; Roediger & Payne, 1985). Much

Address correspondence to: Michelle Meade, Department of Psychology, Montana State University, PO Box 173440, Bozeman, MT 59717, USA. E-mail: mlmeade@montana.edu

This research was completed to fulfil the Master's Thesis requirements at Montana State University for the first author. This research was supported through a Research and Creativity Grant awarded to Michelle L. Meade and Keith A. Hutchison at Montana State University. We thank Sara Davis and Jim Barber for help in running participants. Additionally we thank Roddy Roediger for helpful comments regarding the methodology of this project, and Richard Block for comments on earlier drafts.

http://www.psypress.com/memory DOI:10.1080/09658211.2011.568494

research comparing free and forced recall tests suggests that, for unrelated materials, the number of correct items produced on free and forced recall tasks are generally equivalent; additional items reported on forced recall tests are intrusions rather than additional correct items (e.g., Roediger & Payne, 1985; for a review see Roediger, Wheeler, & Rajaram, 1993).

One important exception to the conclusion that recall criterion (i.e., free or forced recall) does not influence the number of items correctly produced on a recall test is that forced recall does show increases in the proportion of correct items produced relative to free recall when the test materials lend themselves to guessing (e.g., categorised pictures and concrete nouns; Erdelyi et al., 1989; Roediger, Srinivas, & Waddill, 1989). That is, when producing guesses on a forced recall test, participants studying categorical or other guessable materials may produce some studied items simply by guessing, or generating possible related items. To distinguish recall from guessing on forced recall tests, Roediger et al. (1993) asked participants to provide a confidence rating for each item produced on free or forced recall tests of a story depicted in pictures. As reported in Roediger et al., 1989, more items were produced under forced recall instructions, but when items produced were conditionalised on confidence (i.e., only items that participants were confident had appeared in the study episode were counted as correct items), the difference between free and forced recall disappeared. The role of guessing, then, in informing the relationship between free and forced recall is quite important.

Considered together, prior research on forced recall with young adults suggests that, when uncorrected for guessing, forced recall increases recall (or production) of related or guessable items relative to free recall, but has no effect on recall of unrelated items. Once forced recall of guessable materials is corrected for guessing, however, any advantage in forced recall over free recall disappears. This pattern suggests that young adults' recall of unrelated materials is not heavily influenced by guessing, and even with related materials the influence of guessing can be corrected (cf. Roediger et al., 1989). Importantly, this general conclusion has not yet been demonstrated with older adults. Are older adults also generally unlikely to guess on free recall tests, such that the effect of forced recall (corrected and uncorrected for guessing) is equivalent between young and older adults for related and unrelated materials? And, importantly, what role do age-related changes in guessing and metacognition play in informing the relationship between free and forced recall?

There are remarkably few studies that have directly compared young and older adults' individual memory performance on forced recall tests (see Meade & Roediger, 2009, for a related study of age differences in collaborative forced recall). Most relevant to the current experiment is a study by Meade and Roediger (2006), which utilised forced recall as a means to induce self-generated false memories among young and older adults. The focus of their paper was on the effect of prior forced recall on subsequent memory performance. Nonetheless, data reported on the initial test are directly relevant to the research questions addressed in the current study, as they demonstrated that older adults produced more items on a forced recall test of categorised lists relative to cued recall, although older adults reported fewer correct items on an initial forced recall test than younger adults. Young and older adults did not differ on initial cued recall. Although Meade and Roediger did not report this directly, the data imply that when comparing across conditions, older adults derived a relatively smaller benefit from forced recall relative to cued recall than did younger adults for categorised lists (although see Henkel, 2007, for research in a different paradigm demonstrating that young and older adults derived no benefit from forced recall relative to free recall for conceptually similar, physically similar, or unrelated pairs of pictures and/or words across repeated tests). The two previous studies on forced recall among young and older adults suggest that older adults may be similarly affected by forced recall as young adults. Importantly, however, neither study included guessing controls across related and unrelated word lists and so cannot answer questions posed in the current study regarding possible age differences in the items produced under free and forced recall of related and unrelated materials, and possible age differences in guessing.

POSSIBLE AGE DIFFERENCES IN ITEMS PRODUCED ACROSS RECALL CRITERIA

Research on young adults suggests that, when recall is uncorrected for guessing, forced recall has no impact on unrelated materials, but increases recall (or production) of related materials, presumably because some correct items produced under forced recall are the result of pure guessing (Roediger et al., 1989). Given recent research demonstrating age-related differences in guessing, or generating items, there is reason to question whether the relationship between free and forced recall across related and unrelated materials will apply to older adults. Specifically, studies using cued recall and recognition tests have demonstrated that older adults rely disproportionately on guesses and/or plausibility (Kelley & Sahakyan, 2003), they volunteer a greater number of responses than do younger adults (Pansky, Goldsmith, Koriat, & Pearlman-Avnion, 2009), and they are less able to improve memory accuracy by withholding items (Jacoby, Bishara, Hessels, & Toth, 2005). Given such findings, we hypothesise that older adults should also report more guesses on free recall tests, thus resulting in a smaller difference between free and forced recall conditions in older adults' recall relative to younger adults' recall. Of course, guessing should only aid recall for categorised or related materials (cf. Erdelyi et al., 1989), since categorical exemplars may be easily generated, and so may be more confusable due to increased fluency (cf. Jacoby & Hollingshead, 1990; for a discussion of memory errors in a categorised list paradigm relative to an associative list paradigm, see Smith, Gerkens, Pierce, & Choi, 2002). The inclusion of both related and unrelated word lists in the current study, then, allows an examination of the relationship between free and forced recall across materials so as to inform the generality of any increased propensity towards guessing that older adults might demonstrate on recall tests.

To further inform possible age-related differences in guessing on recall tests, the current study compares recall across standard free recall instructions, instructions warning participants there is a penalty for guessing, instructions telling participants there is no penalty for guessing, and forced recall instructions that require guessing. The use of various instructional manipulations allows an examination of possible age-related differences in guessing across two relatively conservative recall criteria (free recall and penalty instructions) and two relatively liberal recall criteria (no penalty and forced recall). If older adults approach free recall tests with a greater propensity towards guessing—as they approach cued recall (Kelley & Sahakyan, 2003) and recognition tests (Pansky et al., 2009)—are they able to suppress this tendency when instructed to do so? Research in other paradigms suggests that older adults' responses across varying instructional manipulations are less flexible than those of younger adults (Kelley & Sahakyan, 2003), and that older adults are less able to withhold inaccurate items (e.g., Jacoby et al., 2005). Assuming findings from cued recall of word pairs generalise to free recall of word lists, it is hypothesised that older adults should be less able to moderate guessing across conditions, especially for categorised lists.

POSSIBLE AGE DIFFERENCES AFTER CORRECTING FOR GUESSING

Roediger et al. (1993) utilised confidence ratings to effectively separate younger adults' guesses from memory on forced recall. However, older adults may be less successful at distinguishing memory from guesses because they demonstrate a lower correspondence between accuracy and confidence (e.g., Kelley & Sahakyan, 2003) and they often report false memories with high confidence (e.g., Dodson, Bawa, & Krueger, 2007; Mitchell, Johnson, & Mather, 2003). As such, the interesting findings by Roediger et al. (1989) reported above, demonstrating no difference in young adults' free and forced recall corrected for guessing, may not apply to older adults. Consistent with this idea, Meade and Roediger (2006) demonstrated that, under forced recall of categorised lists, young and older adults were equally confident in accurate items produced, but that older adults demonstrated higher confidence than young adults in the errors produced. Of interest to the current study is how such metacognitive errors might vary across age groups as a function of test instructions and list relatedness. It is hypothesised that older adults should be especially susceptible to metacognitive errors under forced recall and no penalty conditions for related word lists.

To further assess possible age-related differences in guessing, the current study also employs a corrected recall procedure that allows a measure of how likely participants were to recall a given item when it was studied relative to how likely they were to guess that same item simply because it was a typical exemplar of the presented category (see Meade & Roediger, 2009; Roediger, 1973). Correcting for guesses is especially important on related word lists because the lists contain the most typical exemplars of a given category, and so participants may produce them by generating guesses rather than by retrieving them from memory. To our knowledge there is not a well-established procedure to control for guessing on recall, although several acceptable methods have been established for recognition (e.g., *d'*, hits minus false alarms). Given the scarcity of guess control procedures for recall tests available in the literature, we chose to utilise the corrected recall procedure reported by Meade and Roediger (2009) because, like the current study, that study examined corrected recall for categorised lists among young and older adults.

The current study extends prior research by examining young and older adults' veridical and erroneous recall of categorised and unrelated lists across various instructional conditions (standard free recall, penalties against guessing, encouragement to guess, and forced recall). The current study also assessed participants' confidence that a reported memory item was correct. Of interest is whether potential age-related differences in recall, guessing, and metacognition, vary in relation to list type and instructional manipulations.

METHOD

Participants

A total of 80 undergraduates at Montana State University (age range 18–39 years) and 80 community-dwelling older adults (age range 63–89 years) were recruited for participation in the study. The younger adults participated for course credit as part of an introductory psychology class, while older adults were given a $10 compensation for their participation in the study. Older adults were recruited from the MSU older adult participant pool and the Bozeman community. Older adults were responsible for transportation to and from the testing site at the university.

Participants' mean age and education level are presented in Table 1 along with neuropsychological data including the Mini Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975), the Shipley Institute of Living Scale (multiple-choice vocabulary test; Shipley, 1940), and a Memory Anxiety Questionnaire (adapted from Davidson, Dixon, & Hultsch, 1991). Older adults attained a greater number of years of education ($M = 16.06$) than younger adults ($M = 13.16$), $t(158) = 9.23$, $SEM = 0.13$. This pattern was not surprising given that younger adult participants were enrolled in a freshman-level course. Older adults ($M = 28.70$) scored significantly lower on the MMSE than younger adults ($M = 29.20$), $t(158) = 2.88$, $SEM = 0.10$. This difference was not a concern because all participants scored at least 26 on the MMSE, and so were within the healthy range for cognitive performance. Shipley vocabulary scores were also compared between age groups, with older adults ($M = 34.73$) significantly outperforming younger adults ($M = 28.07$), $t(152) = 11.84$, $SEM = 0.42$.

TABLE 1
Demographic characteristics as a function of participant group

	Younger (N = 80)	*Older (N = 80)*
Age	19.90 (3.01)	73.29 (6.15)
Education	13.16 (1.17)	16.06 (2.57)
MMSE	29.20 (0.92)	28.07 (3.74)
Shipley	28.70 (1.21)	34.73 (3.34)
Memory Anxiety	2.95 (0.58)	3.04 (0.68)

Standard deviations are listed in parentheses.

The Memory Anxiety Questionnaire (adopted from Davidson et al., 1991), assessed participants' anxiety about their memory performance on a scale from 1 to 5. Higher ratings on the questionnaire indicated a greater level of memory anxiety. A 2 (age group: old vs young) $\times$ 4 (test instructions: standard vs penalty vs no penalty vs forced) between- participants analysis of variance (ANOVA) was completed on anxiety scores. No main effects were found for age group or test instructions (both Fs < 1.86, ps $> .14$) and the interaction between age groups and test instructions failed to reach significance, $F < 1.0$, $p > .05$. Anxiety scores from three younger adults were missing due to experimenter error.

Design

The design was a 2 (Age: young vs older adult) $\times$ 4 (Test Instructions: standard vs penalty vs no penalty vs forced) $\times$ 2 (List Type: unrelated vs related) mixed design. Age and instructional conditions were between- participants manipulations and all participants were tested on related and unrelated word lists.

Materials

The Battig and Montague (1969) categorical word association norms were used to construct four word lists of 20 items that were all related by one overlying categorical theme. The related lists were constructed by using the top 25 exemplars from four different categorical lists (for related paradigms see Meade & Roediger, 2006, 2009; Smith, Ward, Tindell, Sifonis, & Wilkenfeld, 2000). Each word in the categorised lists was matched to an unrelated word selected from the English Lexicon Project (ELP) database (Balota et al., 2007) to create four unrelated lists. The unrelated words were matched to the related words on the basis of frequency and word length using the ELP and also matched on concreteness, familiarity, and imageability using The University of Western Australia MRC Psycholinguistic Database (Coltheart, 1981). The matched materials yielded a total of eight 25-item word lists. Each of these word lists was then divided into two versions in an effort to control for guessing of the most common exemplars. Specifically, the top 10 exemplars in each list were divided into two five-word groupings. For each participant, one of the five-word groupings was presented along with the last 15 items in each list and the other five-word grouping was not presented. In this way we could compare how likely participants were to *recall* the most typical exemplars of a given category above and beyond how likely they were to *guess* the most typical exemplars of a given category (see Meade & Roediger, 2009; Roediger 1973).

Additional materials included a locally developed demographics questionnaire, the MMSE, the Shipley Vocabulary Test, and the Memory Anxiety Questionnaire. The memory anxiety questionnaire was taken from Davidson et al. (1991) and modified to include 11 statements (question 12 was locally developed). A 5-point Likert scale was used to assess varying levels of agreement with each statement. The modified Memory Anxiety Questionnaire used in the current study is presented in the Appendix.

Procedure

Participants completed the experiment alone or with one other participant of the same age group. Participants were presented with eight study–test trials. During study, one categorised or unrelated list was presented via computer. Each list item was presented for 1500 ms with a 500 ms inter-stimulus interval (ISI). Participants were instructed to pay attention to each of the items presented in preparation for a memory test. Following the list presentation, participants completed a 90 second math filler task.

Following study, participants were asked to complete a recall test under one of four possible sets of instructions. Under Standard Free Recall instructions participants were asked to write down only words they remembered as being presented on the computer screen (with no mention of guessing or penalty). Under Penalty instructions participants were informed that their score on the test would be based on the correct responses with the subtraction of any incorrect responses, and therefore it would be advantageous to write down only those items they were absolutely positive occurred on the study list. Under No Penalty instructions participants were told that no deduction was associated with an incorrect response so they were encouraged to guess because it could only benefit their overall performance. Under Forced Recall instructions participants were required to provide a response for every blank space on their recall sheet, providing a total of 20 responses, even if that required guessing. All participants reported a total of 20 items in the forced recall condition. Any repeated items were counted only once. Young adults ($M = 19.93$) and older adults ($M = 19.84$) did not differ in the number of non-repeated items reported on forced recall tests $t(38) = 1.23$, $SEM = .02$, $p = .23$. Participants in all conditions were informed that points would be awarded for correct responses and those who performed well on the recall tasks would be entered into a raffle drawing for a prize. These final instructions were given to equate motivation across all conditions.

Additionally, participants in every condition were required to provide a confidence rating for each response according to a 4-point scale (4 = very confident the item was on the list, 3 = somewhat confident the item was on the list, 2 = somewhat confident the item was not on the list, 1 = very confident the item was not on the list). Participants were given 3 minutes to complete the recall task, with more time provided if needed. They rarely used more than 3 minutes.

Participants completed the above procedure for each of the eight word lists, which were presented in a random order. Retrieval instructions were repeated prior to each recall phase to increase instructional salience to the participant. Then all participants completed several questionnaires, which included demographics information, MMSE, Shipley Vocabulary, and the Memory Anxiety Questionnaire. Finally, participants were fully debriefed, and older adults were provided with compensation. A typical experimental session lasted slightly over an hour.

RESULTS

For all results reported, statistical significance is set at $p < .05$ unless otherwise noted.

List recall

The proportion of correctly produced items along with the proportion of items participants assigned the highest confidence rating of 4 are reported in Table 2. A 2 (Age: young vs old) × 2 (List Type: related vs unrelated) × 4 (Test Instructions: standard vs penalty vs no penalty vs forced) mixed factorial ANOVA, with List Type as a within-participants variable, was conducted on the proportion of correctly recalled items. The ANOVA revealed significant main effects of Age, $F(1, 152) = 56.37$, $MSE = 0.01$, and List Type, $F(1, 152) = 2918.11$, $MSE = 0.003$, and a significant Age × List Type interaction, $F(1, 152) = 26.41$, $MSE = 0.003$. Follow-up tests confirmed that although older adults produced a smaller proportion of list items than younger adults on both unrelated word lists ($M = 0.18$ for older adults, $M = 0.30$ for younger adults), $t(158) = 9.33$, $SEM = 0.01$, as well as related word lists ($M = 0.54$ for older adults, $M = 0.60$ for younger adults), $t(158) = 4.10$, $SEM = 0.01$, the age difference was larger for unrelated lists than for related lists. Presumably, this is because related lists offered categorical organisation as retrieval support (cf. Rabinowitz, Craik, & Ackerman, 1982).

The main effect of Test Instructions was only marginally significant, $F(3, 152) = 2.37$, $MSE = 0.03$, $p = .07$, but the interaction between List Type and Test Instructions was significant, $F(3, 152) = 6.64$, $MSE = 0.02$. Follow-up comparisons indicated the proportion of correctly recalled items varied in accordance with test instructions, but only for related word lists (recall on the unrelated lists did not vary as a function of test instructions, ts < 0.76, ps $> .05$; a finding consistent with past research demonstrating that forced recall offers no benefit over free recall for unrelated lists, e.g., Roediger & Payne, 1985). Specifically, for related lists, participants in the forced recall condition ($M = 0.60$) produced more correct list items than participants in the penalty ($M = 0.53$), $t(78) = 3.51$, $SEM = 0.01$, and standard conditions ($M = 0.55$), $t(78) = 2.13$, $SEM = 0.01$, a finding consistent with previous work suggesting that forced recall may sometimes improve production of categorised list items (e.g., Meade & Roediger, 2006). Participants' veridical recall was also higher under no penalty instructions ($M = 0.59$) relative to penalty instructions ($M = 0.53$), $t(78) = 3.15$, $SEM = 0.01$. All

TABLE 2
Mean proportion of items correctly recalled as a function of age, list type, and test instruction; proportion of responses assigned confidence rating of 4 (highest confidence in correct response)

	Younger (N = 80)		*Older (N = 80)*	
	Related	*Unrelated*	*Related*	*Unrelated*
Recall				
Standard	0.59 (0.08)	0.31 (0.08)	0.52 (0.10)	0.18 (0.09)
Penalty	0.55 (0.10)	0.28 (0.09)	0.50 (0.08)	0.18 (0.07)
No penalty	0.63 (0.06)	0.30 (0.08)	0.55 (0.07)	0.17 (0.05)
Forced	0.61 (0.06)	0.30 (0.10)	0.59 (0.09)	0.18 (0.11)
Confidence rating 4				
Standard	0.54 (0.08)	0.29 (0.08)	0.45 (0.13)	0.17 (0.08)
Penalty	0.50 (0.11)	0.27 (0.08)	0.44 (0.10)	0.17 (0.07)
No penalty	0.57 (0.07)	0.28 (0.08)	0.45 (0.09)	0.14 (0.05)
Forced	0.53 (0.08)	0.29 (0.10)	0.50 (0.11)	0.16 (0.11)

Standard deviations are listed in parentheses.

other differences in test instructions were not significant (ts < 1.67, ps $> .10$). Finally, the three-way interaction between Age, List Type, and Test Instructions failed to reach significance, $F < 1.0$. Considered together, the list recall data suggest that both young and older adults varied the proportion of correct items produced on related lists in accordance with test instruction, but that test instructions had little effect on unrelated list recall. Such findings suggest that, when not corrected for guessing, forced recall increases production of related list items equally for young and older adults.

Confidence ratings (list recall)

Confidence ratings were collected to assess participants' belief that a reported response was correct. Confidence ratings may be especially important to age differences in recall across test instructions because, relative to young adults, older adults demonstrate a lower correspondence between confidence ratings and accuracy (e.g., Kelley & Sahakyan, 2003). In other words, even though young and older adults were equally influenced by test instructions for the proportion of items produced (reported above), possible age-related differences in high-confidence responses for items may provide insight into the possible age-related differences in guessing on recall tests. The proportion of correct responses made with a confidence rating of 4 (highest confidence the response was correct) was analysed using a separate ANOVA. Confidence ratings of 4 were selected because they reflect the most conservative measure of high-confidence responses provided by young and older adults. However, because relatively few correct items produced were awarded a confidence rating of 3, the pattern of results reported below does not change when both 3 and 4 ratings were included. As is evident in Table 2, the pattern of results for highest confidence ratings closely resemble the results reported for the proportion of correct responses. Statistically, all main effects and interactions were replicated with the exception that for confidence ratings, there was only a marginal List Type × Test Instructions interaction, $F(3, 152) = 2.23$, $MSE = 0.003$, $p = .09$. Follow-up pairwise comparisons on the marginal interaction revealed no significant differences, except that on the related lists young adults recalled significantly more high confidence items in the no penalty condition ($M = 0.57$) than in the penalty condition ($M = 0.50$), $t(38) = 2.27$, $SEM = .02$. Interestingly, for both young and older adults, the majority of correct responses produced were given the highest confidence rating.

Corrected recall

Guessing is critically important to our research question as participants may be producing additional correct items on related lists due to guessing rather than memory (cf. Roediger et al., 1993), and further because the propensity towards guessing may vary with age. Specifically, older adults may be more likely to respond on a memory test with plausible items (Reder, Wible, & Martin, 1986) and older adults also generally show greater confidence in memory errors than do young adults (e.g., Mitchell et al., 2003). Of interest to the current project is whether young and older adults demonstrate differences in guesses produced across test conditions and also confidence in those guesses. Further, once recall is corrected for guessing, do young and older adults still benefit equally from forced recall? To address these questions, the current study employed a corrected recall measure developed by Meade and Roediger (2009; see also Roediger, 1973). Although other methodologies may exist, we chose to utilise the corrected recall procedure reported by Meade and Roediger (2009) because, like the current study, that study examined corrected recall for categorised lists among young and older adults.

Five critical presented items and five critical non-presented items. The top 10 most common exemplars from each list were divided such that 5 of the most common exemplars were presented to participants (5 critical presented items) and the remaining 5 were not presented to participants (5 critical non-presented items). Table 3 displays the proportion of five critical presented items and five critical non-presented items recalled by participants. The five critical presented items are a subset of the items included in the list recall analyses and were separated out so as to provide a baseline memory measure from which to subtract out guessing. A 2 (Age: old vs young) × 2 (List Type: related vs unrelated) × 4 (Test Instructions: standard vs penalty vs no penalty vs forced) mixed factorial ANOVA computed on the mean proportion of the five presented items revealed results equivalent to the list recall results discussed above. A separate ANOVA computed on the five critical non-presented items revealed results that closely resemble the results of total number of extra list intrusions (discussed later). The statistical results are not discussed further to avoid redundancy; recall proportions are provided in Table 3 so as to illustrate how corrected recall was computed.

Corrected recall. Corrected recall (displayed at the bottom of Table 3) was calculated by subtracting the proportion of recall of the five critical non-presented items from the proportion of recall from the five critical presented items. Corrected recall was subjected to a 2 (Age: old vs young) × 2 (List Type: related vs unrelated) × 4 (Test Instructions: standard vs penalty vs no penalty vs forced) mixed factorial ANOVA with List Type as a within-participants variable. Main effects of List Type, $F(1, 152) = 12.75$, $MSE = 0.02$, Test Instruction, $F(3, 152) = 9.19$, $MSE = 0.03$, and Age, $F(1, 152) = 33.96$, $MSE = 0.03$, were qualified by a significant List Type × Test Instructions × Age three-way interaction, $F(3, 152) = 5.46$, $MSE = 0.02$.

Follow-up comparisons on the three-way interaction indicated that, consistent with early work on forced recall of unrelated items (cf. Roediger & Payne, 1985), test instructions did not influence corrected recall for unrelated word lists (all ts < 1.68, ps $> .09$). However, for related lists, younger adults' corrected recall under standard instructions ($M = 0.54$) was equal to penalty instructions ($M = 0.60$, $t < 1.1$, $p > .05$), but greater than no penalty instructions ($M = 0.28$), $t(38) = 4.06$, $SEM = .04$, and forced instructions ($M = 0.34$), $t(38) = 3.32$, $SEM = .04$. There was no difference between the no penalty and forced recall conditions, $t(38) = 1.03$, $SEM = .05$, $p = .31$. Similar to younger adults, the proportion of corrected recall for older adults under standard instructions ($M = 0.42$) was equal to penalty instructions ($M = 0.46$, $t < 1.0$, $p > .05$). Critically, older adults' corrected recall under standard instructions was also equal to no penalty instructions ($M = 0.36$, $t < 1.4$, $p > .05$), as well as to forced recall instructions ($M = 0.35$, $t < 1.5$, $p > .05$). In other words, younger adults' corrected recall mimicked instructional conditions, with higher levels of corrected recall in conservative conditions where they were encouraged to withhold intrusions and lower corrected recall in conditions in which they were encouraged to guess. In contrast, older adults' corrected recall under standard conditions was equivalent to their corrected recall under liberal response conditions, presumably because older adults were likely to report plausible critical intrusions even without instructions to do so. Another way to conceptualise this pattern of data is that older adults are not gaining as much from instructions to minimise guessing as are young adults because they are less able to reduce their

TABLE 3

Mean proportion of first 10 items (5 critical presented, 5 critical non-presented, and corrected items) recalled as a function of age, list type, and test instruction

	Younger (N = 80)		*Older (N = 80)*	
	Related	*Unrelated*	*Related*	*Unrelated*
5 Critical presented items				
Standard	0.70 (0.12)	0.49 (0.14)	0.64 (0.16)	0.32 (0.15)
Penalty	0.68 (0.15)	0.45 (0.15)	0.63 (0.12)	0.29 (0.12)
No penalty	0.65 (0.12)	0.43 (0.14)	0.66 (0.11)	0.24 (0.10)
Forced	0.80 (0.10)	0.48 (0.15)	0.76 (0.10)	0.24 (0.13)
5 Critical non-presented items				
Standard	0.16 (0.18)	0.00 (0.00)	0.23 (0.13)	0.00 (0.01)
Penalty	0.08 (0.09)	0.00 (0.00)	0.19 (0.15)	0.00 (0.01)
No penalty	0.38 (0.15)	0.01 (0.02)	0.31 (0.15)	0.00 (0.00)
Forced	0.46 (0.13)	0.03 (0.01)	0.41 (0.14)	0.01 (0.01)
5 Corrected recall (presented – non-presented)				
Standard	0.54 (0.20)	0.49 (0.14)	0.42 (0.15)	0.32 (0.15)
Penalty	0.60 (0.15)	0.45 (0.15)	0.46 (0.16)	0.29 (0.12)
No Penalty	0.28 (0.21)	0.42 (0.14)	0.36 (0.15)	0.24 (0.10)
Forced	0.34 (0.18)	0.46 (0.15)	0.35 (0.17)	0.23 (0.14)

Standard deviations are listed in parentheses.

errors with instruction. It is quite noteworthy that, once guessing is controlled, older adults' recall of related lists does not differ between standard free recall instructions and conditions that encourage or require them to guess.

Older adults' relatively reduced modulation of errors across instructional manipulations is consistent with previous research in other paradigms (e.g., Kelley & Sahakyan, 2003), and may reflect age differences in monitoring processes, and/or report criterion (cf. Pansky et al., 2009). Specifically, because accuracy and confidence are relatively less calibrated among older adults, they may be producing errors under conservative test instructions that they perceive as accurate. That is, without good accuracy/confidence calibration, older adults have no means by which to reduce errors on related word lists and to distinguish between plausible exemplars that were studied and plausible exemplars that they generated. In contrast young adults, who typically demonstrate better calibration between accuracy and confidence, may use metacognitive indicators such as confidence as a means to reduce errors under conservative test instructions. Alternatively, older adults may be approaching retrieval from categorised lists from a more generative-heavy strategy such that they simply generate categorical exemplars rather than engage in recollective processes. In either case, the data suggest that once a categorically related item has been generated/retrieved, older adults are less able to monitor the accuracy of the item (cf. Pansky et al., 2009). Older adults' reduced flexibility in error production across conditions, then, is problematic because it illustrates that they are less able than young adults to modulate their own memory accuracy, and are prone to produce greater numbers of plausible erroneous items even under conservative test instructions.

Total number of intrusions

Evidence for our conclusions that older adults report additional plausible items under conservative test instructions, and so demonstrate less difference between free and forced recall tests than young adults, comes from the total number of intrusions reported. The total numbers of intrusions reported per list are presented in Table 4 and include the non-presented critical items displayed in Table 3 plus any additional extralist intrusions. Note that intrusions are quite low in the standard and penalty conditions, and that the relatively higher intrusion rates in the forced recall condition provide a nice manipulation check that participants were in fact producing a greater number of items in accordance with task demands. The average total number of intrusions per list was submitted to a 2 (Age Group: young vs old) $\times$ 2 (List Type: related vs unrelated) $\times$ 4 (Test Instructions: standard vs penalty vs no penalty vs forced) mixed factorial ANOVA. Significant main effects were found for Test Instructions, $F(3, 152) = 22.17$, $MSE = 1755.30$, and List Type, $F(1, 152) = 194.97$, $MSE = 414.62$. These main effects were qualified by a significant List Type $\times$ Test Instruction interaction, $F(3, 152) = 107.70$, $MSE = 2.13$, and further qualified by a significant three-way interaction with Age, $F(3, 152) = 3.46$, $MSE = 2.13$.

Post hoc analyses were carried out to examine the three-way interaction. Specifically, when recalling from related lists younger adults reported significantly fewer intrusions than older adults under standard instructions (for younger adults $M = 1.24$; for older adults $M = 2.30$); $t(38) = 2.28$, $SEM = .31$, and penalty instructions (for younger adults $M = 0.65$; for older adults $M = 1.51$); $t(38) = 2.83$ $SEM = .14$. However, younger and older adults did not differ in the number of intrusions reported under no penalty instructions (for younger adults $M = 3.95$; for older adults $M = 3.46$; $t < 1.0$, $p > .05$), or forced recall instructions (for younger adults $M = 7.70$; for older adults $M = 7.88$; $t < 1.0$, $p > .05$). In fact, the data in Table 4 are numerically quite striking: for related lists, young and older adults produced nearly identical numbers of intrusions per list under no penalty and forced recall instructions, while older adults produced nearly twice as many intrusions per list as young adults under standard and penalty instructions. These results suggest that, under conservative test instructions, older adults were more likely to guess than were young adults, a finding consistent with older adults' increased likelihood to volunteer responses on recognition tests (cf. Pansky et al., 2009).

Two additional results are noteworthy from the total number of intrusions: first, the ANOVA revealed a significant main effect of List Type, $F(1, 152) = 23.61$, $MSE = 0.38$, that was qualified by a List Type $\times$ Test Instruction interaction, $F(1, 152) = 2.85$, $MSE = 0.38$, and a marginal List Type $\times$ Age Group interaction, $F(1, 152) = 3.30$, $MSE = 0.38$, $p = .07$. The only significant difference obtained between age groups across

TABLE 4
Mean total number of intrusions incorrectly recalled per list as a function of age, list type, and test instruction; number of responses assigned confidence rating of 3 or 4 (highest confidence in response)

	Younger (N = 80)		*Older (N = 80)*	
	Related	*Unrelated*	*Related*	*Unrelated*
Recall				
Standard	1.24 (1.38)	1.51 (2.49)	2.30 (1.56)	2.28 (2.53)
Penalty	0.65 (0.65)	1.14 (0.92)	1.51 (1.20)	1.39 (1.56)
No penalty	3.95 (1.78)	5.51 (4.36)	3.46 (2.27)	4.99 (4.12)
Forced	7.70 (1.12)	13.93 (1.98)	7.88 (2.15)	16.11 (2.06)
Confidence rating 3/4				
Standard	0.68 (0.67)	0.54 (0.56)	1.69 (1.31)	0.95 (1.05)
Penalty	0.49 (0.52)	0.73 (0.77)	1.19 (1.06)	0.75 (0.53)
No penalty	1.16 (0.96)	0.89 (0.96)	1.23 (1.26)	0.83 (0.57)
Forced	1.46 (1.20)	0.80 (1.04)	2.05 (1.94)	1.30 (1.43)

Standard deviations are listed in parentheses.

relatedness and test instruction was that, for the unrelated lists, under forced recall instructions, older adults ($M = 16.11$) reported more intrusions than younger adults ($M = 13.93$), $t(38) = 3.50$, $SEM = 0.44$; (all other ts < 1.0, ps $> .05$). This pattern of results replicates Meade and Roediger (2006) and is not surprising since older adults were not able to recall as many correct items after studying an unrelated word list and were therefore required to report additional intrusions to meet task demands. Second, note that older adults' increased extralist intrusions cannot be fully explained by their increased susceptibility to proactive interference (cf. Kane & Hasher, 1995), since younger adults ($M = 0.24$) and older adults ($M = 0.19$) produced an equivalent number of extralist intrusions that had been studied on a previously presented list, $t < 1.2$, $p > .05$.

Confidence ratings (total number of intrusions)

The mean number extralist intrusions per list awarded a high confidence rating (CR 3 or 4) are presented in Table 4. Note that CR 3 ratings were included in the current analyses (only CR 4 were analysed in the List Recall section) because CR 4 ratings alone resulted in missing cells across conditions. Confidence ratings on the intrusions are especially important in identifying whether older adults are aware of their increased intrusions or whether they are producing intrusions with high confidence that the items had been presented. An Age × Test Instruction × List Type ANOVA computed on high confidence intrusions revealed significant main effects of Test Instruction, $F(3, 152) = 2.74$, $MSE = 1.59$, and List Type, $F(1, 152) = 23.63$, $MSE = 0.38$, and a Test Instruction × List Type interaction, $F(3, 152) = 2.85$, $MSE = 0.38$. Follow-up tests confirmed that for related lists, older adults ($M = 1.40$) were more likely to produce high confidence intrusions than were young adults ($M = 0.95$), $F(1, 152) = 5.41$, $MSE = 1.59$, although age did not interact with other variables suggesting that across test instructions and list type, older adults were consistently more likely than young adults to report high confidence intrusions. When considered in conjunction with the number of errors produced (discussed above), the data suggest that not only are older adults more likely than younger adults to produce errors across conditions (even when instructed to be conservative), they are also more confident across conditions that the errors produced actually occurred in the lists. Interestingly, recalculating the confidence ratings to examine confidence ratings of 1 and 2 eliminates the age difference in intrusions (3.61 vs 3.82 for young and old, respectively). Thus older adults are not simply guessing more. Instead, the age difference is due solely to high confidence intrusions. Perhaps older adults are more likely to use plausibility as evidence for confidence than young adults (cf. Kelley & Sahakyan, 2003).

Changes in error production across trials

One potentially interesting age difference in error production may result from possible differences in error production across trials. That is, older and younger adults may differentially increase or decrease the number of reported intrusions across study–test trials. To examine this possibility with critical intrusions on related lists, a 4 (List Order: 1–4) × 2 (Age: young vs old) × 4 (Test Instructions: standard vs penalty vs no penalty vs forced) mixed ANOVA was used. Consistent with the results reported above, the proportion of critical intrusions produced varied in accordance with Test Instruction, $F(3, 152) = 38.99$, $MSE = 0.08$. Critically, however, the proportion of critical intrusions produced did not change across study test trials, $F(3, 456) = 1.21$, $MSE = 0.04$, $p = .31$, nor was there a main effect of age or any interaction between age and other variables, $F\text{s} < 1.0$, demonstrating that, across trials, younger and older adults did not differentially change the proportion of critical intrusions produced. This same pattern held when other intrusions (non critical intrusions) were examined: young and older adults did not vary the proportion of other intrusions produced across test trials. Considered together, these analyses suggest that young and older adults were not differentially increasing or decreasing error production across trials, but rather that the proportion of errors produced by young and older adults remained relatively constant across trials.

DISCUSSION

The present experiment provides an examination of possible age-related differences in guessing on free and forced recall tests for categorical and unrelated lists. By comparing the pattern of responses between young and older adults' correct and erroneous recall across list type and test instructions, the current experiment revealed novel findings regarding age differences in recall, guessing, and metacognition on recall tests.

When correct recall was measured as the proportion of list items produced across test conditions, uncorrected for guessing, participants were able to adjust their correct recall levels in accordance with test instructions, but only when recalling from categorised word lists. This finding is consistent with previous research demonstrating that free recall tasks are only influenced by guessing when they involve related materials (Erdelyi et al., 1989; Roediger et al., 1989). Interestingly, the current study found that older and younger adults adjust correct recall to the same magnitude such that, for related lists, young and older adults both derived a benefit from forced recall relative to free recall.

When list recall was corrected for guessing, the current experiment demonstrated very interesting age differences across test conditions. Namely, younger adults' corrected recall still varied in accordance with test instructions, but older adults' recall did not. Specifically, older adults' recall (above and beyond guessing), was equivalent regardless of whether they were instructed to minimise guessing or they were encouraged or forced to produce guesses. The pattern of data obtained among young adults is seemingly contradictory to previous research findings demonstrating that, for young adults, corrections for guessing eliminate any differences between free and forced recall for related materials. However, if data from the current study are re-examined in relation to high confidence ratings (as was used for the guessing control by Roediger et al., 1993), the data are entirely consistent with previous research as the proportion of high confidence items produced by young adults under free recall of related lists ($M = 0.54$) was identical to the proportion of high confidence items produced by young adults under forced recall of related lists ($M = 0.53$; data in Table 2). The corrected recall methodology used in the current study (rotating sets of items through presented and non-presented conditions) may offer a more age-sensitive guessing control in light of older adults' reduced correspondence between accuracy and confidence (cf. Kelley & Sahakyan, 2003), and in fact this method resulted in the interesting finding that young adults' corrected recall varied in accordance with recall criterion, but that older adults' corrected recall did not vary as a function of recall criterion.

Evidence that older adults' reduced flexibility across test instructions was the result of increased guessing under more conservative test instructions comes from the intrusion data. The total number of intrusions suggested that older adults were more likely than young adults to produce intrusions under conservative instructions, but that young and older adults did not differ in the proportion of intrusions produced under liberal instructions. In other words, recall from young adults could be made to look like that from older adults when encouraged or required to guess. However, simple discouragement from guessing was insufficient to make older adults' recall mimic that of young adults. By comparing intrusion rates across test conditions, then, one conclusion from the current experiment is that older adults derive a relatively smaller benefit from forced recall relative to free recall of related materials because they are already more likely to guess the related materials on the free recall test. One interesting question regarding this conclusion is whether older adults would still demonstrate a greater propensity towards guessing than younger adults if they were better matched on veridical recall (see Craik & McDowd, 1987, for a discussion of scaling issues in comparing young and older adults' memory across various retrieval criteria). In the current study veridical recall was not matched perfectly between age groups. However, age differences in veridical recall were relatively small (i.e., no greater than .08 differences), and so we anticipate older adults' increased guessing is not due entirely to base-rate differences.

Can older adults' increased intrusion rate be classified as guessing? As noted by Roediger et al. (1993), young adults' confidence ratings on forced recall tests may be useful in distinguishing between items produced from memory and items produced by guessing. However, given older adults' reduced metacognitive abilities (cf. Kelley & Sahakyan, 2003), confidence ratings may not effectively distinguish memory from guessing among older adults. In fact, in the current study older adults were more confident in their intrusions than were young adults. In other words, older adults reported errors with high confidence ratings, suggesting they did not believe they were guesses, but rather that they were highly confident that the item had appeared in the list.

Several theoretical models may provide insight into the nature of older adults' increased production of high confidence errors across recall tests of related lists. Most relevant to the age-related response differences obtained in the current study is the metacognitive model proposed by Koriat and Goldsmith (1996). Briefly, Koriat and Goldsmith define quantity as an input bound measure in which the assessment is based on all of the memory items that are presented at study. Accuracy is defined as an output-bound measure where an individual has control over the items reported and the willingness to report an item from memory is contingent on one's subjective confidence that the item was studied earlier. Accuracy and quantity are described in terms of a trade-off and are regulated by strategic ability. The quantity/accuracy trade-off can only be assessed under conditions of free report where individuals have the opportunity to volunteer or abstain from reporting information.

Past research has adapted the metacognitive model for use with older adults and has demonstrated that older adults suffer reduced calibration between accuracy and confidence (Kelley & Sahakyan, 2003; Pansky et al., 2009) and a lower report criterion (Pansky et al., 2009) such that older adults are less able to monitor the accuracy of items generated on memory tests. Consistent with these findings, the confidence ratings obtained in the current study suggest that older adults are not aware that the errors produced on free recall tests are errors, but instead report them with high confidence. More generally, the age-related response differences obtained in cued recall and recognition may also generalise to free recall tests of related, or categorised, materials.

What implications do the current data have for our understanding of age-related changes in guessing on free recall? The findings of the current study suggest that age-related differences in guessing on free recall vary in accordance with list materials. Specifically, consistent with earlier work on young adults (cf. Erdelyi et al., 1989; Roediger et al., 1989), older adults' free recall of unrelated list materials is not heavily influenced by guessing. Regarding free recall of related materials, however, older adults' free recall might be more heavily influenced by guessing than young adults' free recall. This finding is quite important because it suggests that older adults may approach standard free recall tests of categorised lists with a greater propensity towards guessing than young adults.

Considered more broadly, this finding is important because much work on forced recall with young adults concluded that free recall was an unbiased measure of memory (i.e., free recall was not conservatively biased such that participants could generate or recall more items than they reported on free recall; see Roediger et al., 1993, for a review). Given the current findings, however, there may be age-related biases on free recall tests of categorised lists such that older adults are relatively more likely than young adults to produce plausible, high-confidence errors on such tests. Forced report has proven a powerful tool to examine age-related differences in cued recall (e.g., Kelley & Sahakyan, 2003) and recognition (e.g., Pansky et al., 2009), and the current study suggests that forced recall may also be useful in informing age-related differences in guessing on free recall. Finally, given the focus on forced recall to inform responding on free recall, it may be important to highlight some everyday examples of forced recall. As noted by Meade and Roediger (2006), people may be encouraged to produce additional information in a variety of settings, including pressure from a police officer to remember everything from a crime scene, and encouragement from a therapist or friend to reconstruct a past experience.

The current study was designed to explore possible age-related differences in recall, guessing, and metacognition on free recall tests. Towards that end, the results of the current study can be summarised as follows: First, replicating past studies with young adults (e.g., Erdelyi et al., 1989; Roediger et al., 1989), forced recall improves memory for categorically related lists, but does not influence memory for unrelated items. When list recall is not corrected for guessing, this conclusion applies to both young and older adults. Second, once list recall was corrected for guessing, young adults' corrected recall varied in accordance with instructional conditions, while older adults' corrected recall under conservative instructions (standard free recall and penalty against guessing) was equivalent to their corrected recall under liberal response conditions (no penalty and force recall), again only for categorised word lists. This suggests that older adults were reporting greater errors on free recall tests, even when explicitly instructed not to. Finally, older adults reported intrusions with greater confidence than young adults. This demonstrates that older adults may not necessarily be aware that an item they reported was incorrect. Thus evidence that older adults differ from young adults in their propensity towards guessing on free recall tests of categorised lists is marked by the increased number of intrusions provided, the reduced flexibility in error production across test conditions, and the increased level of confidence that reported intrusions were studied.

Manuscript received 10 August 2010
Manuscript accepted 6 February 2011

REFERENCES

Anderson, J. R. (1983). Retrieval of information from long-term memory. *Science*, *220*, 25–30. doi: 10.1126/science.6828877

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*, 97–123. doi: 10.1037/h0033773

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory in healthy older adults. In F. I. M. Craik & E. Tulving (Eds.), *The Oxford handbook of memory* (pp. 411–425). New York: Oxford University Press.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, *80*, 1–45. doi: 10.1037/h0027577

Cofer, C. N. (1967). Does conceptual organisation influence the amount retained in immediate free recall? In B. Kleinmuntz (Ed.), *Concepts and the structure of memory*. New York: Wiley.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*, 474–479. doi: 10.1037/0278-7379.13.3.474

Davidson, H. A., Dixon, R. A., & Hultsch, D. F. (1991). Memory anxiety and memory performance in adulthood. *Applied Cognitive Psychology*, *5*, 423–433. doi: 10.1002/acp.2350050504

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Ageing, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Ageing*, *22*, 122–133.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, *6*, 159–171.

Erdelyi, M. H., Finks, J. & Feigin-Pfau, M. B. (1989). The effect of response bias on recall performance, with some observations on processing bias. *Journal of Experimental Psychology: General*, *118*, 245–254. doi: 10.1037/0096-3445.118.3.245

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. doi: 10.1016/0022-3956(75)90026-6

Henkel, L. A. (2007). The benefits and costs of repeated memory tests for young and older adults. *Psychology and Ageing*, *22*, 580–595. doi: 10.1037/0882-7974.22.3.580

Jacoby, L. L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005). Ageing, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General*, *34*, 131–148. doi: 10.1037/0096-3445.134.2.131

Jacoby, L. L., & Hollingshead, A. (1990). Reading student essays may be hazardous to your spelling: Effects of reading incorrectly and correctly spelled words. *Canadian Journal of Psychology*, *44*, 345–358.

Kane, M. J., & Hasher, L. (1995). Interference. In G. L. Maddox (Ed.), *Encyclopedia of ageing* (2nd ed., pp. 514–516). New York: Springer-Verlag.

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language*, *48*, 704–721. doi: 10.1016/S0749-596X(02)00504-1

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517. doi: 10.1037/0033-295X.103.3.490

Meade, M. L., & Roediger, H. L. III. (2006). The effect of forced recall on illusory recollection in younger and older adults. *The American Journal of Psychology*, *119*, 433–462.

Meade, M. L., & Roediger, H. L. III. (2009). Age differences in collaborative memory: The role of retrieval manipulations. *Memory & Cognition*, *37*, 962–975. doi: 10.3758/MC.37.7.962

Mitchell, K. J., Johnson, M. K., & Mather, M. (2003). Source monitoring and suggestibility to misinformation: Adult age-related differences. *Applied Cognitive Psychology*, *17*, 107–119.

Pansky, A., Goldsmith, M., Koriat, A., Pearlman-Avnion, S. (2009). Memory accuracy in old age: Cognitive, metacognitive, and neurocognitive determinants. *European Journal of Cognitive Psychology*, *21*, 303–329. doi: 10.1080/09541440802281183

Rabinowitz, J. C., Craik, F. I. M., & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology*, *39*, 325–344. doi: 10.1037/h0080643

Reder, L. M., Wible, C., & Martin, J. (1986). Differential memory changes with age: Exact retrieval vs. plausible inference. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*, 72–81.

Ritter, W., & Buschke, H. (1974). Free, forced and restricted recall in verbal learning. *Journal of Experimental Psychology*, *6*, 1204–1207. doi: 10.1037/h0037411

Roediger, H. L. III. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning & Verbal Behavior*, *12*, 644–657.

Roediger, H. L. III, & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesias: A puzzle for generate/recognise theories. *Memory & Cognition*, *13*, 1–7.

Roediger, H. L. III., Srinivas, K., & Waddill, P. (1989). How much does guessing influence recall? Comment on Erdelyi, Finks, and Feigin-Pfau. *Journal of Experimental Psychology: General*, *118*, 255–257. doi: 10.1037/0096-3445.118.3.255

Roediger, H. L. III., Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing, and reconstructing the past. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 97–134). San Diego, CA: Academic Press.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology*, *9*, 371–377.

Smith, S. M., Gerkens, D. R., Pierce, B. H., & Choi, H. (2002). The roles of associative responses at study and semantically guided recollection at test in false memory: The Kirkpatrick and Deese hypotheses. *Journal of Memory and Language*, *47*, 436–447.

Smith, S. M., Ward, T. B., Tindell, D. R., Sifonis, C. M., & Wilkenfeld, M. J. (2000). Category structure and created memories. *Memory & Cognition*, *28*, 386–395.

APPENDIX

Memory Anxiety Questionnaire (Questions 1–11 adapted from Davidson et al., 1991; Question 12 locally developed).

Respond: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree.

1. I would feel on edge right now if I had to take a memory test.
2. When someone I don't know very well asks me to remember something I get nervous.
3. I am usually uneasy when I attempt a problem that requires me to use my memory.
4. I get tense and anxious when I feel my memory is not as good as other people's.
5. I get anxious when I am asked to remember.
6. I do not get flustered when I am put on the spot to remember new things.
7. I feel jittery if I have to introduce someone I just met.
8. I get anxious when I have to do something that I haven't done in a long time.
9. If I am put on the spot to remember names, I know I will have difficulty doing it.
10. I would feel very anxious if I visited a new place and had to remember how to find my way back.
11. I get upset when I cannot remember something.
12. When taking a memory test, I feel it is a more serious memory error to leave something out than it is to write down something extra.