Psychology and Aging

Taking the Testing Effect Beyond the College Freshman: Benefits for Lifelong Learning

Ashley N. D. Meyer and Jessica M. Logan

Online First Publication, February 25, 2013. doi: 10.1037/a0030890

Psychology and Aging

0882-7974/13/$12.00 DOI: 10.1037/a0030890

Taking the Testing Effect Beyond the College Freshman: Benefits for Lifelong Learning

Ashley N. D. Meyer and Jessica M. Logan
Rice University

Formal learning is a lifelong pursuit that does not occur exclusively within universities. Accordingly, methods for improving long-term learning, including the well-established use of testing, should be examined for various ages of learners outside typical university settings to properly assess their usefulness. This study examined testing effects in 60 younger university students aged 18–25, 60 younger community adults aged 18–25, and 60 middle-aged to older community adults aged 55–65 at immediate and longer delays (2-day). All groups similarly benefited from testing at both delays, implying that testing can be a beneficial lifelong learning tool for a diversity of learners.

Keywords: lifelong learning, repeated testing, aging, episodic memory, retrieval

Supplemental materials: http://dx.doi.org/10.1037/a0030890.supp

Formal learning is a lifelong pursuit. In addition to the multitude of degree-seeking adults, many younger, middle-aged, and older adults alike partake in lifelong learning, or education outside of traditional university settings for job training, career advancement, skill learning, or personal enrichment. For example, data from the *Digest of Education Statistics 2011* demonstrate that 53% of employed younger adults (aged 17–24) and 42 and 38% of employed middle-aged to older adults aged 55–59 and 60–64, respectively, participated in lifelong learning of some type in 2005 (U.S. Department of Education, 2012, p. 644).

Considering the prevalence of lifelong learning, it is noteworthy that relatively little is known about improving learning/training in populations other than those examined in most experiments (i.e., young university students). As one example, the recent interest in using testing as a learning technique (e.g., Roediger & Karpicke, 2006b) is concentrated on 18- to 25-year-olds in traditional university settings. This is likely due to convenience, but leaves many questions unanswered regarding the generalizability of testing effects to younger, middle-aged, and older adults participating in lifelong learning outside university settings. As such, learning and training techniques, including the use of testing, should be examined in those populations in order to assess their effectiveness in a more meaningful and generalized way.

The Benefits of Testing

Research shows that testing widely improves long-term learning (a major goal in education) in young populations. Specifically, it improves students' long-term memory for tested material, meaning they will do better on a delayed test if they were previously tested on material than if they were not or if they merely restudied it. Restudying, conversely, often only results in better short-term retention. This "testing effect," generalizes to different test types (from true/false to full essay tests), learning materials (from trigram number pairs to prose), and settings (from laboratories to classrooms); (for a review, see Roediger, Agarwal, Kang & Marsh, 2010). The benefits of testing using educationally relevant materials, however, have mostly been examined in traditional college-aged students and, to a lesser extent, in middle school, high school, and even medical school students (Logan, Thompson, & Marshak, 2011; Roediger, Agarwal, McDaniel, & McDermott, 2011). Whether testing helps middle-aged and older adult learners outside of traditional academia, however, remains an unanswered question.

The Benefits of Testing in Non-University Populations

Testing effects may not generalize to learners beyond traditional academic settings. People not tested regularly in school may react adversely to the use of tests as learning events, as they are likely to be unaccustomed to taking tests, may be more anxious taking tests, or may have difficulty accessing relevant knowledge when tests are introduced apart from initial learning. If they underperform on these tests, they may benefit less from them due to a lack of processing that occurs with successful retrieval (Carpenter & DeLosh, 2006). The use of testing as an educational aid in these learners, however, can still be valuable, so it is useful to assess whether learners outside typical university settings can benefit as much or at all from testing as a learning technique.

Ashley N. D. Meyer and Jessica M. Logan, Department of Psychology, Rice University.

Ashley N. D. Meyer is now at Houston VA Health Services Research and Development Center of Excellence, Michael E. DeBakey VA Medical Center, Houston, Texas.

We thank Dr. Margaret E. Beier, Dr. James L. Dannemiller, Dr. Nicoletta Orlandi, Dr. Gunes Avci, Dr. Sara Haber, and Rebecca A. Lundwall for their invaluable insights during the conceptualization and completion of this research. Additionally, we would also like to thank Elaine Mahoney for her help with data collection.

Correspondence concerning this article should be addressed to Ashley N. D. Meyer, Michael E. DeBakey Veterans Affairs Medical Center (MEDVAMC), HSR&D Center of Excellence (152), 2002 Holcombe Boulevard, Houston, TX 77030. E-mail: ashleyndmeyer@gmail.com

The Benefits of Testing in Middle-Aged to Older Student Populations

Testing benefits in lifelong learning are uncertain for relatively older adults because of comparatively poor long-term episodic memory (Balota, Dolan, & Duchek, 2000). Poor initial memory may adversely impact testing benefits by limiting the elaborative processing that occurs with successful retrieval (Carpenter & DeLosh, 2006). Indeed, previous researchers found a proportional relationship between initial, successful retrieval and benefits accrued from testing (Marsh, Agarwal, & Roediger, 2009). Thus, middle-aged to older adults may benefit less from educational testing by starting with lower performance, thus engaging in less elaborative processing for tested items.

A few studies have looked at the effects of age on the benefits of testing, but present mixed results and either did not include a non-tested control condition or educationally relevant material. For example, some studies showed older adults (63–90 years old) benefited from testing when memorizing seen and imagined images, but benefited less compared to younger adults (18–22 years old); other studies showed older adults (61–75 years old) benefited as much from testing as younger adults did when memorizing words (Henkel, 2007, 2008; Rabinowitz & Craik, 1986). Non-tested control conditions, however, were not used for comparison, making it difficult to draw strong conclusions regarding testing benefits for older learners. When comparisons of untested to tested items were included, other studies showed older adults benefited from testing just like younger adults did when learning low-associate word pairs or face-name pairs (Logan & Balota, 2008; Tse, Balota, & Roediger, 2010). These latter findings are promising, yet an examination of testing with educationally relevant material is needed to determine the benefits of testing for learning for middle-aged and older adult learners participating in lifelong learning.

Summary of Key Issues

Testing has been shown to benefit long-term learning in younger students in typical educational settings. The present study makes the novel contribution of generalizing these benefits to the longer-term learning of educationally relevant materials in populations beyond those typically studied. This was done by including working-age younger (18–25) and middle-aged to older learners (55–65) who might be likely to participate in lifelong learning.

Method

The experiment followed this general procedure: a study phase, a distractor phase, a learning phase where restudied and tested material was manipulated within subjects, another distractor phase, a between-subjects retention interval, a final test, and an intelligence test.

Participants

Participants included 60 younger university students (aged 18–25), 60 younger adults from the community of a large city (aged 18–25), and 60 middle-aged to older adults from the same community (aged 55–65) (see Table 1 for age means and standard errors). Younger community participants serve as a direct comparison to college students of the age typically used in testing effect studies, but from outside academia. Older participants represent a population showing memory decline (Park, Polk, Mikels, Taylor, & Marshuetz, 2001) but likely to be working (unlike most retired individuals over 65) and partaking in continuing education (U.S. Department of Education, 2012).

University students were recruited through a web-based experiment scheduling program and were compensated with partial course credit. Community participants were recruited through craigslist (craigslist, 2010) and/or community flyers, represent a heterogeneous sample of adults not currently seeking degrees, and were compensated $20. Inclusion criteria included: normal/corrected-to-normal vision, college experience (see Table 1 for group education levels), and English fluency (the latter two criteria ensured participants could understand the materials).

Additionally, to detect group intelligence differences, intelligence was measured using the vocabulary and matrix reasoning subtests of the Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999). This was administered orally to participants according to the scale's instructions. Raw WASI vocabulary scores (traditional IQ scores calculated by age would minimize/alter age differences) showed that the younger university students had better verbal abilities than both community groups did. The middle-aged to older community group, however, had higher verbal abilities than the younger community group did. This

Table 1
Characteristics of Participants From Each Age Group, Including: Age, Education Levels Achieved, and Wechsler Abbreviated Scale of Intelligence (WASI) Scores

	Young university students (n = 60)	Young adults from community (n = 60)	Middle-aged to older adults from community (n = 60)
Age: *M* (*SEM*)	19.3 (0.16)	22.4 (0.25)	59.8 (0.43)
Education level: *n* (%)			
Some college	60 (100.00)	35 (58.3)	14 (23.3)
Completed college	0 (0.00)	18 (30.0)	29 (48.3)
Graduate school	0 (0.00)	5 (8.33)	9 (15.0)
PhD/MD/JD/DDS	0 (0.00)	2 (3.33)	8 (13.3)
WASI Scores: *M* (*SEM*)			
Vocabulary[a]	52.67 (0.64)	41.58 (1.10)	47.78 (1.15)
Matrix Reasoning[b]	22.75 (0.30)	21.95 (0.37)	20.43 (0.52)

[a] Out of a possible score of 62. [b] Out of a possible score of 26.

was verified with a between-subjects ANOVA and post hoc pairwise comparisons (*LSD*) ($F(2, 177) = 31.5$, $p < .001$, $MSE = 58.7$, $\eta_p^2 = .26$; younger university adults > middle-aged to older community adults > younger community adults ($p < .001$ and $p = .001$, respectively)) (see Table 1 for raw group scores).

Scrutinizing raw WASI matrix reasoning scores, it was apparent that the younger groups outperformed the relatively older group on abstract reasoning. A between-subjects ANOVA and post hoc pairwise comparisons (*LSD*) verified this ($F(2, 177) = 8.32$, $p < .001$, $MSE = 9.98$, $\eta_p^2 = .09$; younger university adults and younger community adults > middle-aged to older community adults ($p < .001$ and $p = .001$, respectively) and younger university adults = younger community adults ($p = .17$)). Scores from both WASI sections were used in ensuing covariate analyses (see Table 1 for group scores).

Materials

Study and restudy phases. Study materials were four articles on armadillos, black holes, human hearts, and tsunamis—topics learners knew little about (verified in pilot testing) (National Geographic, 2012a, 2012b, 2012c, 2012d; also see online supplemental Appendix). Length ranged from 324 to 577 words.

Distractor phases. Distractor materials included multiplication problems with two factors between 10 and 50 (e.g., $14 \times 49 =$ _____). Use of multiplication ensured no overlap between distractor and other experiment material.

Learning phase recognition test. Initial multiple-choice tests included one correct and three incorrect answer choices per question. As an example, participants saw the following:

1. In the last stage before a black hole is formed, a detonation occurs, known as a(n)
   a. starburst b. explosion c. blastula d. supernovae

There were 10 questions for each passage, resulting in 40 questions total (each participant saw 20 randomly ordered questions in the learning phase: 10 from each of 2 topics).

Final cued-recall test. Final test questions were created from the initial test questions by deleting the answer choices and providing blanks to fill in answers (i.e., the root of the questions were the same in both phases). For example, participants saw the following:

1. In the last stage before a black hole is formed, a detonation occurs, known as ______.

There were 40 cued-recall questions (10 questions for each of the 4 passages). Multiple-choice questions were utilized for the learning phase of the experiment, because of their ease of use in classroom settings, emulating a common practice of periodically giving short quizzes over material in a class before a later exam. Cued-recall questions were used for the final test to emulate what is expected of students on a more comprehensive midterm or final.

Procedure

Procedures followed previously used testing effects study procedures (e.g., see Roediger & Marsh, 2005). After consenting, participants were randomly assigned to one of two retention interval conditions. All phases were on paper to minimize technology-related age differences.

Study phase. Participants were then instructed to read the passages. They were informed they would be tested on the information, but were not told specifics about the upcoming test. Each participant read the passages in the same order and was given a total of 15 minutes to read all of them (all participants finished reading within the allotted time frame).

Distractor phase 1. Participants then worked on 50 multiplication problems for 5 minutes to clear their working memory.

Learning phase: Initial recognition test and restudying. Participants then took a recognition test over two topics and restudied the other two topics (testing vs. restudying were manipulated within subjects and topics in each condition and condition order were counterbalanced). For the test, participants completed the 20 questions and then notified the experimenter. Next, the experimenter graded the test and told participants how many questions they got correct/incorrect, but not which ones they got correct/incorrect (e.g., "You got 14 out of 20 questions correct"). This feedback was used to create educationally relevant learning conditions in which learners are graded on their work, without affording participants an additional study opportunity. For restudying, participants reread two passages and then notified the experimenter. Pilot testing showed that testing and rereading took a similar amount of time.

Distractor phase 2. Participants worked on 50 multiplication problems for 5 minutes.

Final cued-recall test. The final cued-recall test occurred either after the second distractor task (i.e., the 5-min distractor task was the 5-min retention interval) or 2 days later depending on random assignment. Participants were tested on all 40 randomly ordered questions and were told to fill in the blanks as best as they could.

Results

All analyses with a *p* value below .05 are considered significant and effect sizes for significant *F* and *t* tests are represented by η_p^2 and Cohen's *d*, respectively. Additionally, both initial and final test performance were calculated as percent correct.

Initial Recognition Test Performance

On the initial recognition test in the learning phase, the younger university adults outperformed both community groups. These latter groups, nonetheless, did not differ from each other. A 2 (retention interval: 5-min vs. 2-day) × 3 (age group: younger university adults vs. younger community adults vs. middle-aged to older community adults) between-subjects ANOVA verified this with a significant main effect of age group on initial recognition performance, $F(2, 174) = 5.02$, $p = .008$, $MSE = 165$, $\eta_p^2 = .06$. Post hoc pairwise comparisons (*LSD*) revealed that the younger university adults had higher initial performance than both the younger community adults ($t(177) = 2.98$, $p = .003$, $d = .55$) and the middle-aged to older community adults did, $t(177) = 2.42$, $p = .017$, $d = 0.48$. The latter two groups did not differ significantly, $t(177) = .57$, $p = .57$, $d = -0.10$, and no other differences were found. Taking raw WASI vocabulary and matrix reasoning scores into account, however, eliminated age differences on initial performance (see Appendix 1 for comparison of ANOVA and ANCOVA statistics).

Final Test Performance

As in previous studies, learning phase testing resulted in better final performance than restudying did (i.e., a testing effect occurred). Also, learners performed better after a shorter retention interval than after a longer one, yet testing effects increased with retention interval length. Lastly, the younger university adults outperformed the community groups, but testing effects were similar for all. In fact, 87% of younger university students, 80% of younger community adults, and 80% of middle-aged to older community adults showed testing effects (i.e., initial performance was less than or equal to final performance), with an average testing effect close to 17 points (see Figure 1). Results were validated with a 2 (learning condition: testing vs. restudying) × 2 (retention interval: 5-min vs. 2-day) × 3 (age group: younger university students vs. younger community adults vs. middle-aged to older community adults) repeated measures ANOVA. A significant effect of learning condition was found, such that testing led to better final performance than restudying did, $F(1, 174) = 187.07$, $p < .001$, $MSE = 132$, $\eta_p^2 = .52$). There was also a significant main effect of retention interval on final test performance, such that performance was better on the immediate final test compared to the delayed final test, $F(1, 174) = 41.00$, $p < .001$, $MSE = 403$, $\eta_p^2 = .19$ (see Figure 1). Also, a main effect of age group was found ($F(2, 174) = 7.73$, $p = .001$, $MSE = 403$, $\eta_p^2 = .08$), such that younger university participants outperformed both the younger ($t(177) = 3.07$, $p = .002$, $d = .46$) and middle-aged to older community participants ($t(177) = 3.66$, $p < .001$, $d = .58$). Performance of the community samples, however, did not significantly differ.

Additionally, there was a significant learning condition (testing vs. restudying) by retention interval interaction, $F(1, 174) = 4.65$, $p = .03$, $MSE = 132$, $\eta_p^2 = .03$. A post hoc analysis using the testing effect as the dependent variable revealed that testing effects increased with increasing retention intervals, $F(1, 174) = 4.65$, $p = .03$, $MSE = 264$, $\eta_p^2 = .03$: the effect was 14 points after a 5-min delay and 19 points after a 2-day delay. No other differences were found.

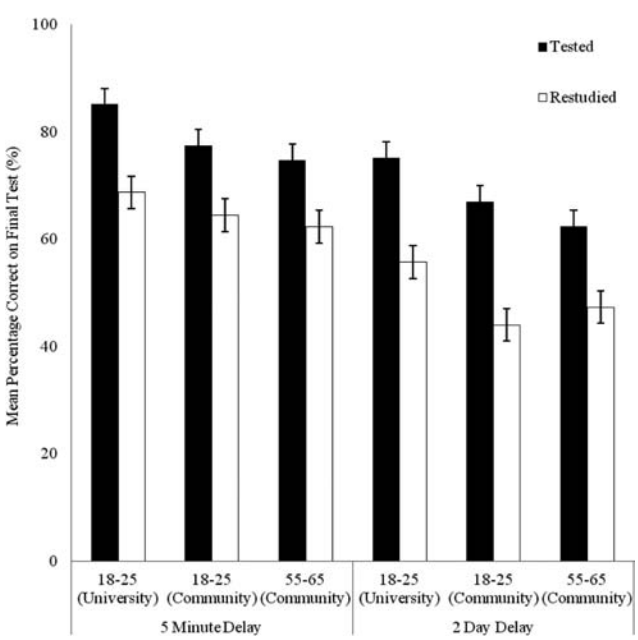

Figure 1. Mean percentage of cued-recall items answered correctly on final cued-recall test as a function of learning condition, final test retention interval, and age group. Error bars represent ± 1 standard error of the mean.

Taking both raw vocabulary and matrix reasoning WASI scores into account, however, removed age effects on performance, yet maintained all testing effect patterns (see Appendix 1 for comparison of ANOVA and ANCOVA statistics). So, not only was age not related to performance after controlling for intelligence, but it did not interact with any other factor, meaning both younger and middle-aged to older adults benefited similarly from testing.

Discussion

Given the prevalence of lifelong learning in a spectrum of ages (at least, for employed people, U.S. Department of Education, 2012), this research examined the benefits associated with a highly touted learning technique—testing—and its use with younger and older learners outside traditional university settings using educationally relevant materials, a previously unexplored topic. In the present study, younger university students outperformed both younger and middle-aged to older community samples on initial and final memory tests, though there was no performance difference between the community groups. Nonetheless, all groups benefited similarly from testing compared to restudying on immediate and delayed tests (even when considering intelligence). These findings imply that testing not only helps different populations of learners, but it helps them to a similar extent. Even learners unaccustomed to regular testing can make educational gains by being tested. Thus, the current study generalizes testing benefits to additional groups of learners, including community adults and older individuals, and shows the effects for educationally relevant materials. As mentioned in the results section, a high percentage of individuals from each group showed testing effects, with an average testing effect of almost two letter grades in education.

A caveat for these findings is that the educationally relevant conditions used may give a memorial advantage to initially tested items. Participants saw the same items on the initial and final tests in the testing condition, but saw the entire study text in the restudy condition without being directed toward the items to be tested later. A memorial advantage might come from increased attention or practice on tested items in the learning phase. Future studies could show only the pieces of information in the restudy condition that appear on the final test to ensure the current findings still apply in less educationally relevant, scientifically rigorous conditions.

An additional finding regarding the powerful effect of testing in this study is that testing effects were found after both 5-min and 2-day delays. Some reports show testing effects after 1 or 2 days, but not after 5 minutes (Roediger & Karpicke, 2006a). Differences may arise from varied distractor periods used after encoding (i.e., the use of 2-min distractors vs. the use of 5-min distractors in the present study may have decreased effortful retrieval processes and increased cursory benefits of restudying in the other study) or the use of free recall versus recognition on the initial test. Testing effects with free recall may manifest themselves later, when briefer benefits of studying wear off. Other studies have found benefits of testing with shorter delays, however (see Roediger et al., 2010 for a review), so the present findings complement many testing effect studies.

Perhaps more interesting is the finding that middle-aged to older adults showed similar testing effects as younger adults did at both intervals, meaning they did not show an increased amount of forgetting over the longer delay after being tested during learning. Testing seemingly helped protect against forgetting in all adults. The fact that the middle-aged to older adults benefited from testing similarly to younger adults may be because older adults often perform well when learning material that is more relevant to their lives and also when effective retrieval operations are induced by the task (e.g., Castel, 2005). This study used learning materials that older adults may be somewhat familiar with (e.g., the human heart), increasing the chances they were able to relate the material to other they knew and engage appropriate encoding processes during study. An open question is whether using cued- or free-recall tests during learning in middle-aged to older adults would have produced these effective encoding/retrieval operations.

Theoretically, the finding that middle-aged to older learners benefited from testing as much as both sets of younger learners did offers mixed evidence toward the hypothesis that older adults may benefit less from testing than younger adults do because of poorer initial performance. Initial performance of both the middle-aged to older adults and the younger community adults was worse than that of the university students. However, both community groups benefited as much as the university students did from testing, which may not seem to support the hypothesis that initial performance is related to testing effects. On the other hand, a correlational analysis shows that testing effects were related to initial performance, $r(178) = .194$, $p < .01$, but not to age specifically. Controlling for intelligence, all samples performed similarly, so the hypothesis that older adults benefit less from testing due to poorer initial performance becomes immaterial. The relatively older adults in this study benefited as much from testing as the younger samples did, regardless of initial performance. This may be due to the high performance level seen in this sample of middle-aged and older adults. Future studies could look at lower-performing older adults to elucidate the relationship between initial performance and testing benefits in aging populations, yet results may just imitate findings with poorer-performing younger adults (Marsh, Agarwal, Roediger, 2009) and may not represent selective aging differences.

Implications

The use of testing as a learning tool has been thoroughly examined in young students. This research builds on that and supports the notion that educators can use tests to increase learning in an expanded population, including younger and middle-aged to older community adults. Nontraditional students themselves can also justifiably use testing to increase their learning. Lastly, this research has positive implications for training working adults who often need to gain new skills/knowledge for work, especially when changing careers or jobs.

References

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy young and older adults. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 395–409). London, UK: Oxford University Press.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34,* 268–276.

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging, 20,* 718–721.

Craigslist. (2010). Online community. Retrieved from http://www.craigslist.org/

Henkel, L. A. (2007). The benefits and costs of repeated memory tests for young and old adults. *Psychology and Aging, 22,* 580–595.

Henkel, L. A. (2008). Maximizing the benefits and minimizing the costs of repeated memory tests for older adults. *Psychology and Aging, 23,* 250–262.

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition, 15,* 257–280.

Logan, J. M., Thompson, A. J., & Marshak, D. W. (2011). Testing to enhance retention in human anatomy. *Anatomy Science Education, 4,* 243–248.

Marsh, E. J., Agarwal, P. K., & Roediger, H. L., III. (2009). Memorial consequences of answering SAT II questions. *Journal of Experimental Psychology: Applied, 15,* 1–11.

National Geographic. (2012a). *Armadillos: Dasypodidae*. Retrieved from http://animals.nationalgeographic.com/animals/mammals/armadillo/?source=A-to-Z

National Geographic. (2012b). *Black holes: A might void*. Retrieved from http://science.nationalgeographic.com/science/space/universe/black-holes-article.html

National Geographic. (2012c). *Heart*. Retrieved from http://science.nationalgeographic.com/science/health-and-human-body/human-body/heart-article.html

National Geographic. (2012d). *Tsunamis: Killer waves*. Retrieved from http://environment.nationalgeographic.com/environment/natural-disasters/tsunami-profile/

Park, D. C., Polk, T. A., Mikels, J. A., Taylor, S. F., & Marshuetz, C. (2001). Cerebral aging: Integration of brain and behavioral models of cognitive function. *Dialogues in Clinical Neuroscience, 3,* 151–165.

Rabinowitz, J. C., & Craik, F. I. M. (1986). Prior retrieval effects in young and old adults. *Journal of Gerontology, 41,* 368–375.

Roediger, H. L., III, Agarwal, P. K., Kang, S. H. K., & Marsh, E. J. (2010). Benefits of testing memory: Best practices and boundary conditions. In G. M. Davies & D. B. Wright (Eds.), *New frontiers in applied memory* (pp. 13–49). Brighton, UK: Psychology Press.

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17,* 382–395. doi:10.1037/a0026252

Roediger, H. L., III, & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Roediger, H. L., III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implication for educational practice. *Perspectives on Psychological Science, 1,* 181–210.

Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1155–1159.

Tse, C. S., Balota, D. A., & Roediger, H. L., III. (2010). The benefits and costs of repeated testing: On the learning of face-name pairs in older adults. *Psychology and Aging, 25,* 833–845.

U.S. Department of Education, National Center for Education Statistics. (2012). *Digest of education statistics 2011* (NCES Report 2012–011). Retrieved from http://nces.ed.gov/pubs2012/2012001.pdf

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI).* San Antonio, TX: Harcourt Assessment.

(*Appendix follows*)

Appendix 1

Analyses of Variance and Covariance for Initial and Final Performance for Younger and Middle-Aged to Older Learners

	Initial test								Final test							
	ANOVA[a]				ANCOVA[b]				ANOVA[c]				ANCOVA[d]			
Effect	*df*	*F*	*p*	η_p^2	*df*	*F*	*p*	η_p^2	*df*	*F*	*p*	η_p^2	*df*	*F*	*p*	η_p^2
Between																
Delay	1,174	1.22	.27	.007	1,172	1.60	.21	.01	**1,174**	**41.00**	**<.001**	**.19**	**1,172**	**66.66**	**<.001**	**.28**
Age group	**2,174**	**5.02**	**.008**	**.06**	2,172	.28	.76	.003	**2,174**	**7.73**	**.001**	**.08**	2,172	1.20	.30	.01
Delay × age group	2,174	.50	.61	.006	2,172	.64	.53	.007	2,174	.30	.74	.003	2,172	.30	.74	.003
Raw WASI vocabulary (Covariate)	—	—	—	—	**1,172**	**9.36**	**.003**	**.05**	—	—	—	—	**1,172**	**33.92**	**<.001**	**.17**
Raw WASI matrix Reasoning (Covariate)	—	—	—	—	**1,172**	**22.75**	**<.001**	**.12**	—	—	—	—	**1,172**	**32.38**	**<.001**	**.16**
Within																
Learning condition	—	—	—	—	—	—	—	—	**1,174**	**187.07**	**<.001**	**.52**	**1,172**	**4.24**	**.04**	**.02**
Learning condition × delay	—	—	—	—	—	—	—	—	**1,174**	**4.65**	**.03**	**.03**	**1,172**	**4.81**	**.03**	**.03**
Learning condition × age group	—	—	—	—	—	—	—	—	2,174	1.34	.26	.015	2,172	1.11	.33	.01
Learning condition × delay × age group	—	—	—	—	—	—	—	—	2,174	.98	.38	.011	2,172	1.03	.36	.01

Note: Mean Square Error (*MSE*) = 154 ([a]); 130 ([b]); 403 for between and 132 for within ([c]); and 255 for between and 133 for within ([d]). Analyses in bold are significant ($p < .05$).

Received January 20, 2012
Revision received September 21, 2012
Accepted September 24, 2012 ■