

A New Look at Retest Learning in Older Adults: Learning in the Absence of Item-Specific Effects

Lixia Yang, Maureen Reed, Frank A. Russo, and Andrea Wilkinson

Department of Psychology, Ryerson University, Toronto, Canada.

We investigated retest learning (i.e., performance improvement through retest practice) in the absence of item-specific effects (i.e., learning through memorizing or becoming familiar with specific items) with older adults. Thirty-one older adults (ages 60–82 years, $M=71.10$, $SD=6.27$) participated in an eight-session self-guided retest program. To eliminate item-specific effects, parallel versions of representative psychometric measures for Inductive Reasoning, Perceptual Speed, and Visual Attention were developed and administered across retest sessions. The results showed substantial non-item-specific retest learning, even controlling for anxiety, suggesting that retest learning in older adults can occur at a more conceptual level.

Key Words: Inductive reasoning—Item-specific effects—Older adults—Perceptual speed—Retest learning—Visual attention.

ACCUMULATED research suggests reserved cognitive plasticity, as indexed by training-induced improvement, in old age (e.g., Ball et al., 2002; Baltes & Lindenberger, 1988). The improvement occurs not only in the tutor-guided training where a tutor explains rules and teaches effective strategies (e.g., Saczynski, Willis, & Schaie, 2002) but also in the self-guided retest program where participants learn from repeatedly practicing the same tasks (i.e., retest learning) without receiving any external guidance or feedback (e.g., Hofland, Willis, & Baltes, 1981; Yang, Krampe, & Baltes, 2006). It has been evidenced that self-guided retest and tutor-guided training can produce an equal amount of improvement (Baltes, Sowarka, & Kliegl, 1989).

Strategy use is evidenced to be an underlying mechanism for tutor-guided training effects in Inductive Reasoning (Saczynski et al., 2002); however, the mechanisms for self-guided retest learning are poorly understood. Salthouse, Schroeder, and Ferrer (2004) speculated that retest learning could be driven by item-specific effects through memorizing/familiarizing with specific items and item-general effects through familiarity with the testing situation, reduced anxiety, and procedural learning. Previous retest learning studies were particularly vulnerable to item-specific effects because the same items were given across sessions (e.g., Yang et al., 2006). It remains unclear whether older adults will show retest learning in the absence of item-specific effects.

To fill this gap, this study focused on the non-item-specific retest learning in older adults. To this end, we developed and administered parallel versions of the psychometric tests, which do not overlap in perceptual items but involve largely the same conceptual rules, at different retest sessions. The manipulation effectively eliminated item-specific effects and maximally encouraged rule-based learning at a more conceptual level. The target abilities (i.e., Inductive Reasoning, Perceptual Speed, and Visual Attention) were the

same as those used in Yang et al. (2006). They all showed reliable age-related declines and substantial retest learning effects in old age (Yang et al., 2006). This enables between-study comparisons to determine the contribution of item-specific effects in different ability domains. Finally, the Beck Anxiety Inventory (BAI; Beck & Steer, 1993) was administered at pretest and posttest to evaluate the possibility that a reduction in anxiety would mediate retest effects.

In sum, this study aims to determine the existence of non-item-specific retest learning and the contributions of item-specific effects and anxiety to retest learning in older adults.

METHODS

Participants

Thirty-one healthy community-dwelling older adults (ages 60–82 years, $M=71.10$, $SD=6.27$), including 26 women and 5 men, participated in this study. They were well educated based on their years of education ($M=17.65$, $SD=4.40$) and scores on the Shipley Vocabulary test ($M=36.61$, $SD=3.48$; Shipley, 1946). They rated themselves as healthy on a 1–10 scale ($M=7.74$, $SD=1.37$). All of them scored below the cutoff of 6 ($M=1.29$, $SD=1.60$) on the Short Blessed test (SBT; Katzman et al., 1983).

Materials

The tests used in the retest sessions were two Inductive Reasoning measures: letter series (LS) and number series (NS; Bliessner, Willis, & Baltes, 1981); two Perceptual Speed tests: digit symbol substitution test (DSST; Wechsler, 1981) and letter comparison (LC; Salthouse, 1991), and one Visual Attention test: D2 test (D2; Brickenkamp, 1994). For each test, we developed parallel versions to be

administered at different sessions. Pilot work was conducted to ensure that the parallel versions of the same test were comparable in difficulty level.

Letter Series and Number Series.—The tests require participants to figure out a pattern/rule based on a series of letters/numbers and then identify one letter/number that could best continue the pattern. The time limit was 6 min. Four parallel versions were developed for each test. Parallel versions were developed following the rules described in a previous study (Allaire & Marsiske, 2005). We first parsed each item in the parent tests into the specific rules/patterns and then developed a new item by shifting each letter alphabetically or shifting each number in numeric order so that the new item and the parent item follow the same pattern/rule. For instance, the item “g a f a e a ___” could be changed into “h b g b f b ___,” and the item “12, 13, 10, 11 ___” could be modified into “14, 15, 12, 13 ___” in a parallel version of LS and NS, respectively. The alternative answer choices were accordingly generated by shifting each letter or number by a certain number of steps in alphabetic or numeric order.

Digit Symbol Substitution Test.—This test requires participants to substitute as many digits as possible with the corresponding symbols within a 90-s time limit. Eight parallel versions were developed by pseudo-randomly assigning each digit a different symbol for each new version so that none of the digit–symbol pairs repeated across versions.

Letter Comparison.—In this test, participants compare and determine whether two letter strings are the same or different within 1 min. Eight parallel versions were developed for LC. To develop parallel versions for LC, we first split the two pages for each of the three sections (containing three-letter, six-letter, and nine-letter strings, respectively) to make the two new parent versions so that each parent version contained one page for each section. Then we replaced each consonant letter in the parent items by a new consonant letter that did not appear in any parent test items (e.g., “C” was replaced by “G”) to develop the first two alternate versions. Finally, we reversed each pair of letters in all the parent items (e.g., “C” reversed with “L”) or in the items of the first two alternate version (e.g., “G” reversed with “B”) to further develop four new alternate versions.

D2 Test.—In this test, participants visually detect two target symbols, for example, a letter “d” with one (‘) or two marks (”) above and/or underneath, from all other similar symbols (e.g., a letter “d” with different mark patterns, a letter “p” with various mark patterns). Four parallel versions were developed, each containing two unique target items. Time limit was 80 s.

Procedure

Data were collected in a 10-session training course, comprising pretest, eight sessions of self-guided retest training (twice a week), and posttest. We used the same self-guided retest paradigm as in Yang et al. (2006) where participants were repeatedly tested without receiving any external guidance or feedback. Different from Yang et al., a unique parallel version of each test was administered at each retest session. To maintain participants’ interest, one of the principal researchers gave a lecture at each session on various topics addressed in an undergraduate Introduction to Psychology course. The training tests were integrated into the 1-hr lecture and administered under standard timed conditions. The order of tests for the three target abilities was counterbalanced across sessions. To minimize fatigue effects, three tests were given at each session. DSST was given at all eight sessions. LC and NS were given at sessions 2, 4, 6, and 8. LS and D2 were given at sessions 1, 3, 5, and 7.

RESULTS

Overall, 4.91% of the 713 attainable data points were missing due to participants’ absence or being late. The missing data points were replaced by the individual linear regression estimates following an established procedure (Lindenberger & Baltes, 1997). Further analyses showed the participants who missed data points ($n=10$) and those who completed all the sessions ($n=21$) did not differ in most demographic variables ($ps>.21$) except that participants with missing data points were older than those without ($p=.001$). Nevertheless, the covariance analyses including the completion status as covariate did not change the main result patterns, and the data analyses based on Listwise deletion of missing data points revealed virtually identical result patterns. To provide a common baseline and enable between-test comparisons, we transformed raw scores (number of correct solutions) into T scores for each test ($M=50$, $SD=10$), standardized to the first training session. Session 5 of the D2 test was excluded from the final analyses due to a technical problem with the stopwatch in precisely timing the task.

Retest Learning Effects

Number of correct solutions (T scores).—Figure 1 shows the mean T scores for the number of correct solutions. The repeated-measures analyses of variance (ANOVAs) were conducted in each ability domain, with linear (i.e., incremental improvement) and quadratic (i.e., saturation) contrasts specified. Both Perceptual Speed tests (DSST and LC) showed significant retest learning in linear, $F_s>24.35$, $ps<.001$, $\eta^2_s>.44$, and quadratic trends, $F_s>15.90$, $ps<.001$, $\eta^2_s>.34$. The composite Inductive Reasoning scores (i.e., average T scores of LS and NS) showed a significant linear effect, $F(1, 30)=20.56$, $p<.001$, $MSE=10.15$, $\eta^2=.41$; but

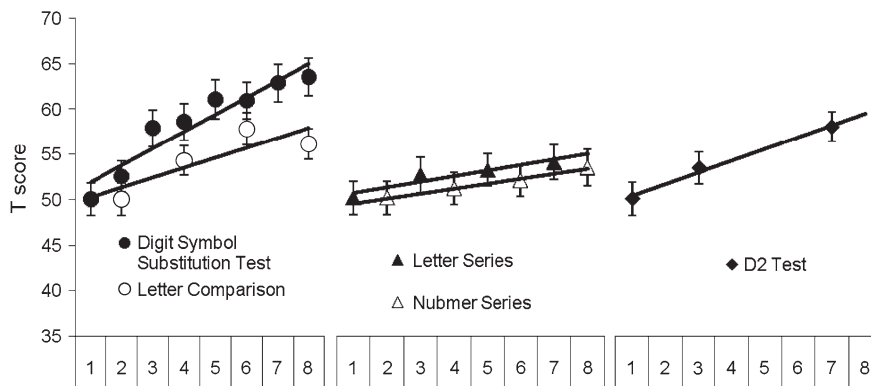


Figure 1. Mean T scores and linear regression fits as a function of retest session for each test and ability domain. Error bars denote standard errors. Data for the three abilities were presented in separate panels. The curve fits were based on average individual estimates from the linear regressions.

no quadratic effect, $F < 1$. The Visual Attention test (D2) also showed a reliable linear effect, $F(1, 30) = 22.19$, $p < .001$, $MSE = 43.66$, $\eta^2 = .43$; but no quadratic effect, $F < 1$.

Accuracy.—The same sets of ANOVAs on accuracy (i.e., proportion of correct solutions out of all attempted items) revealed that only the linear improvement in Inductive Reasoning was significant, $F(1, 30) = 5.17$, $p < .05$, $MSE = 0.01$, $\eta^2 = .15$. All other effects were not significant, $F_s < 2.39$, $p_s > .13$. The increase in accuracy suggests training-induced improvement in strategy use in Inductive Reasoning (Boron, Turiano, Willis, & Schaie, 2007).

Retest learning rate (slope).—The analysis on the individual slope estimates revealed different learning rates across ability domains, $F(2, 60) = 7.14$, $p < .01$, $MSE = 8.49$, $\eta^2 = .19$. Visual Attention ($M = 3.95$, $SD = 4.67$) and Perceptual Speed ($M = 2.42$, $SD = 1.49$) did not differ in learning rate ($p = .09$); however, both showed a steeper rate than Inductive Reasoning ($M = 1.16$, $SD = 1.42$), $F_s(1, 30) > 10.37$, $p_s < .01$, $\eta^2_s > .25$.

The Contribution of Item-Specific Effects

To examine the contribution of item-specific effects in different ability domains, we compared the retest learning effect size (i.e., the gain T scores divided by baseline standard deviation in each test) across the equal number of training sessions between this study and Yang et al. (2006) on Perceptual Speed (DSST), Inductive Reasoning (LS and NS), and Visual Attention (D2). The analyses involved participants who overlapped in age range (aged 70–82 years), including 16 from this study and 47 from Yang et al. The overall 3 (ability: Speed, Reasoning, vs. Attention) \times 2 (study: Present Study vs. Yang et al.) ANOVA revealed a significant ability effect, $F(2, 122) = 11.04$, $p < .001$, $MSE = 0.33$, $\eta^2 = .15$; qualified by a reliable interaction, $F(2, 122) = 4.68$, $p < .05$, $MSE = 0.33$, $\eta^2 = .07$. Post-hoc analyses revealed that the between-study difference was only reliable

in Reasoning, $F(2, 60) = 7.14$, $p < .01$, $MSE = 8.49$, $\eta^2 = .19$, with a larger learning effect in Yang et al. ($M = 0.80$, $SD = 0.78$) than in the current study ($M = 0.29$, $SD = 0.55$). Figure 2 illustrates this interaction.

The Contribution of Anxiety

Two participants missed BAI at posttest and were thus excluded from the analysis. Overall, BAI scores (collapsed over pretest and posttest) showed consistent negative correlations with the average T scores and accuracy (collapsed across retest sessions) in all the tests ($r_s = -.14$ to $-.72$). The correlations were significant in Reasoning ($r_s < -.35$, $p_s \leq .05$) and Attention ($r_s < -.55$, $p_s < .001$). However, BAI scores did not change from pretest ($M = 6.71$, $SD = 6.59$) to posttest ($M = 6.10$, $SD = 6.69$), $t < 1$, and they did not correlate with the retest learning effect size ($r_s < 0.26$) in all the tests. Moreover, the retest learning effects remained significant even after controlling for anxiety in the covariance

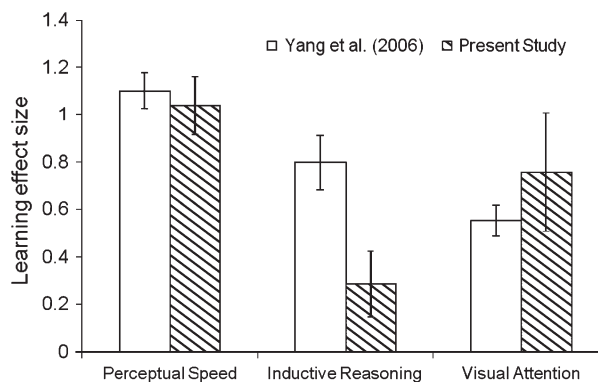


Figure 2. The retest learning effect size (i.e., gain T score from the first to the last training sessions divided by the baseline standard deviation at the first training session) for the three ability domains: Perceptual Speed (performance on digit symbol substitution test across the first six sessions), Inductive Reasoning (performance on letter series and number series across the first four sessions), and Visual Attention (performance on D2 test across the first three sessions) between Yang et al. (2006) and the current study.

analyses. In sum, anxiety neither changes with retest practice nor contributes to the non-item-specific retest learning.

DISCUSSION

This study provides the first direct empirical evidence that retest learning in older adults may occur in the absence of perceptual item-specific effects. Overall, the training gain corresponds to 0.33–1.35 *SD* units from baseline performances (DSST: 1.35; LC: 0.61; LS: 0.39; NS: 0.33; D2: 0.74).

Although the specific nature of the non-item-specific retest learning remains to be specified, our data suggest that anxiety is not a critical factor. Based on previous findings that training-induced increases in accuracy reflect improved strategy use (Boron et al., 2007), we infer from the result of training-induced accuracy improvement in Inductive Reasoning that retest learning in Inductive Reasoning is driven by spontaneous strategy use through mastering and applying the rules to solve new items at a conceptual level. In contrast, the retest learning in Perceptual Speed and Visual Attention was mainly manifested in response speed rather than accuracy, suggesting that retest learning in these domains may be primarily driven by item-general effects through familiarity with the testing situation or skill-based procedural learning (e.g., better eye–hand coordination and visual scanning). Furthermore, the flatter learning rate (slope) in Inductive Reasoning than in the other two domains may be driven by the differentially higher complexity of the Inductive Reasoning tests, which involve multiple complex rules (such as repetitions, skips, forward and backward orders identified in Saczynski et al., 2002). In contrast, the Perceptual Speed and Visual Attention tests involve very simple rules that apply to all the items in each test.

The between-study comparison suggests that item-specific effects only occur in Inductive Reasoning, perhaps due to the relatively complex surface structures of the items and the multiple solution rules. In contrast, the simple items and the solution rules in the Perceptual Speed and Visual Attention tests expose minimal requirement on item-specific effects.

Some limitations should be noted. First, the current sample involves well-educated and highly functioning participants and thus limits the generalization of the findings. Second, direct within-study manipulation will certainly strengthen our current between-study comparison that was based on a relatively small sample. Nevertheless, our results made significant contributions to the cognitive training literature by providing direct empirical evidence that retest learning in older adults is driven, at least in part, by item-general factors not captured by item-specific learning or anxiety.

FUNDING

This study was supported by the Ryerson New Faculty SRC Development Fund awarded to the lead author, Dr. Lixia Yang. We thank Ann

Manderville, Linda Truong, Eva Tan, and Sherry Weng for their assistance in collecting data.

CORRESPONDENCE

Address correspondence to Lixia Yang, PhD, Department of Psychology, Ryerson University, 823A-Jorgenson Hall, 350 Victoria Street, Toronto, Ontario, Canada M5B 2K3. Email: lixiay@ryerson.ca

REFERENCES

- Allaire, J. C., & Marsiske, M. (2005). Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychology and Aging, 20*, 390–401.
- Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., et al. (2002). Effects of cognitive training interventions with older adults. *Journal of the American Medical Association, 288*, 2271–2281.
- Baltes, P. B., & Lindenberger, U. (1988). On the range of cognitive plasticity in old age as a function of experience: 15 years of intervention research. *Behavior Therapy, 19*, 283–300.
- Baltes, P. B., Sowarka, D., & Kliegl, R. (1989). Cognitive training research on fluid intelligence in old age: What can older adults achieve by themselves? *Psychology and Aging, 4*, 217–221.
- Beck, A. T., & Steer, R. A. (1993). *Beck Anxiety Inventory manual*. San Antonio, TX: Psychological Corporation.
- Blieszner, R., Willis, S. L., & Baltes, P. B. (1981). Training research in aging on the fluid ability of inductive reasoning. *Journal of Applied Developmental Psychology, 2*, 247–265.
- Boron, J. B., Turiano, N. A., Willis, S. L., & Schaie, K. W. (2007). Effects of cognitive training on change in accuracy in inductive reasoning ability. *Journal of Gerontology: Psychological Sciences, 62B*, P179–P186.
- Brickenkamp, R. (1994). *Test D2: Aufmerksamkeits-Belastungs-Test. (Handanweisung)*. Göttingen, Germany: Hogrefe-Verlag GmbH & Co. KG.
- Hofland, B. F., Willis, S. L., & Baltes, P. B. (1981). Fluid intelligence performance in the elderly: Intraindividual variability and conditions of assessment. *Journal of Educational Psychology, 73*, 573–586.
- Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a short orientation-memory-concentration test of cognitive impairment. *American Journal of Psychiatry, 140*, 734–739.
- Lindenberger, U., & Baltes, P. B. (1997). Intellectual functioning in old and very old age: Cross-sectional results from the Berlin Aging Study. *Psychology and Aging, 12*, 410–432.
- Saczynski, J. S., Willis, S. L., & Schaie, W. (2002). Strategy use in reasoning training with older adults. *Aging Neuropsychology and Cognition, 9*, 48–60.
- Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science: A Journal of the American Psychological Society/APS, 2*, 179–183.
- Salthouse, T. A., Schroeder, D. H., & Ferrer, E. (2004). Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age. *Developmental Psychology, 40*, 813–822.
- Shipley, W. C. (1946). *Institute of Living Scale*. Los Angeles: Western Psychological Services.
- Wechsler, D. (1981). *WAIS-R manual*. New York: Psychological Corporation.
- Yang, L., Krampe, R. T., & Baltes, P. B. (2006). Basic forms of cognitive plasticity extended into the oldest-old: Retest learning, age, and cognitive functioning. *Psychology and Aging, 21*, 372–378.

Received August 7, 2008

Accepted March 1, 2009

Decision Editor: Elizabeth Stine-Morrow, PhD