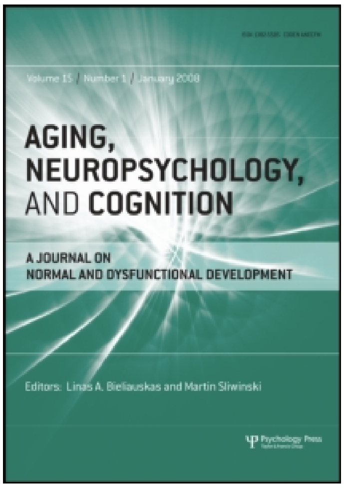


Aging, Neuropsychology, and Cognition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713657683

How to Gain Eleven IQ Points in Ten Minutes: Thinking Aloud Improves Raven's Matrices Performance in Older Adults

Mark C. Fox[a]; Neil Charness[a]
[a] Department of Psychology, Florida State University, Tallahassee, FL, USA

First published on: 03 August 2009

Aging, Neuropsychology, and Cognition, 17: 191–204, 2010
http://www.psypress.com/anc
ISSN: 1382-5585/05 print; 1744-4128 online
DOI: 10.1080/13825580903042668

Psychology Press
Taylor & Francis Group

How to Gain Eleven IQ Points in Ten Minutes: Thinking Aloud Improves Raven's Matrices Performance in Older Adults

MARK C. FOX AND NEIL CHARNESS
Department of Psychology, Florida State University, Tallahassee, FL, USA

ABSTRACT

Few studies have examined the impact of age on reactivity to concurrent think-aloud (TA) verbal reports. An initial study with 30 younger and 31 older adults revealed that thinking aloud improves older adult performance on a short form of the Raven's Matrices (Bors & Stokes, 1998, *Educational and Psychological Measurement, 58*, p. 382) but did not affect other tasks. In the replication experiment, 30 older adults (mean age = 73.0) performed the Raven's Matrices and three other tasks to replicate and extend the findings of the initial study. Once again older adults performed significantly better only on the Raven's Matrices while thinking aloud. Performance gains on this task were substantial (d = 0.73 and 0.92 in Experiments 1 and 2, respectively), corresponding to a fluid intelligence increase of nearly one standard deviation.

Keywords: Think-aloud; Concurrent verbalization; Older adults; Aging; Problem solving; Raven's Matrices; Matrix reasoning.

INTRODUCTION

Concurrent verbal reports allow for a more detailed analysis of cognitive processes than do traditional behavioral measurements such as response time and accuracy because they provide information that can be used to infer specific strategies and identify the characteristics of a task that give rise to errors in problem solving (Ericsson & Simon, 1980, 1993). One concern is

This research served in part to fulfill the Master's thesis requirement for the first author and was supported by a grant to the second author from NIA: 2P01 AG17211. The authors would like to thank Anders Ericsson and Joyce Ehrlinger for many helpful comments and suggestions.

Address correspondence to: Mark C. Fox, Department of Psychology, Florida State University, Tallahassee, FL 32306-1270, USA. E-mail: fox@psy.fsu.edu

that concurrent verbalization might be reactive, altering the way in which a task would normally be accomplished. Ericsson and Simon (1980) proposed a process model of concurrent verbalization predicting that reactivity is more likely to occur when instructions solicit verbalization beyond the relatively automatic linguistic activation that occurs naturally during thought processes. They recommend that researchers use non-manipulative instructions that direct participants to just think aloud (TA verbalization), and discourage descriptive and introspective verbalizations.

There is considerable evidence supporting the position that TA is non-reactive, unlike other modes of concurrent verbalization that have been shown to alter performance (Ericsson & Simon, 1993). However, some have argued that even TA could impact performance if participants modify their strategies to match the presumed demands of the experimenter, perform better because of enhanced working memory resulting from auditory rehearsal, or suffer decrements in performance as a result of processing resources lost to verbalizing (Biehal & Chakravarti, 1989; Russo, Johnson, & Stephens, 1989). Others have argued that TA could encourage participants to focus more on verbalizable information (Schooler, Ohlsson, & Brooks, 1993).

Despite the fact that the TA method has been used in aging research to investigate problem solving processes (e.g., Charness, 1981), remarkably few studies have examined whether older adults – a population characterized by general declines in fluid ability and processing speed – react differently to TA than do younger adults. However, these studies focused on single tasks, limiting the extent to which conclusions about older adult reactivity can be drawn. Johnson (1993) found that TA increased reaction times but did not impact apartment-search decisions in older or younger adults. Perfect and Dasgupta (1997) had older adults think aloud or remain silent during encoding and found that the groups did not differ in familiarity judgments. This study did not test for the possibility of reactivity differences between age groups, because there was no silent younger group. Gilhooly, Phillips, Wynn, Logie, and Della Salla (1999) reported that verbalization did not impact older or younger adult performance on a five-disc version of the Tower of London (number of moves to completion), though it did slow planning time and movement time. While Johnson's sample was large (64 older adults and 116 participants total), sample sizes in the Perfect and Dasgupta, and Gillhooly et al. studies were relatively small (20 participants per group in both studies), leaving open the possibility that minor TA effects went undetected.

Thus, it remains to be seen whether other tasks entailing a wider variety of processing requirements are susceptible to reactivity from TA. It is possible that specific cognitive deficits experienced by older adults may inform the question of which tasks are most likely to elicit reactivity in this group. Older adults experience declines in processing speed that lead to slower performance in general and lower accuracy on complex cognitive tasks (Salthouse, 1996). It is possible that older adults may have greater increases in solution times when thinking aloud relative to remaining silent than younger adults. Because TA is essentially a concurrent task, previous studies of older adult dual-task performance may be relevant to predicting how older adults should respond to TA. A meta-analysis on age differences in dual-task performance by Riby, Perfect, and Stollery (2004) revealed that controlled and automatic processes differentially affect dual-task performance of older adults and younger individuals. Dual-tasks involving controlled processes tend to lead to greater age-related decrements than those involving automatic processes. The relatively low linguistic demands of TA (i.e., admittance of phrases, single words, long pauses, etc.) would seem to require little or no controlled processes under most circumstances, however, the need to translate abstract information into words may involve more controlled processing. Such recoding may be necessary when thinking aloud during abstract reasoning tasks such as the Raven's Advanced Progressive Matrices or spatial tasks such as cube comparison. Ericsson and Simon (1980) suggest that recoding of non-verbal information may prolong solution times, but should not impair accuracy when proper TA instructions are used; however, their model has not been explicitly tested with older adults.

EXPERIMENT 1

The purpose of Experiment 1 was to test the extent of reactivity in older adults relative to younger adults on a variety of tasks that are assumed to differ systematically in processing requirements. Four tasks were chosen that varied with respect to how much processing of spatial or abstract stimuli were entailed. Older adults should have greater difficulty with abstract or spatial tasks when thinking aloud.

Method

Participants

Thirty-one older (18 female; mean age 72.3) and 30 younger (18 female; mean age 19.0) adults participated in the study. Older adults were recruited from the community and were paid $10 to participate. Older participants were screened for dementia with the Wechsler Memory Scale III and the Short Portable Mental Status Questionnaire. The first test requires the participant to correctly recall at least seven pieces of information from at least one of two short narratives read by an experimenter. Participants had to make two or fewer errors on SPMSQ for participation in the study. Younger adults were recruited from the department subject pool and received course credit.

Materials

Tasks investigated were paired associate learning with instructions to generate mediators, Cube Comparison, an abbreviated set of the Raven's Advanced Progressive Matrices (Bors & Stokes, 1998; Raven, 1965), and mental multiplication. The paired associates task entails generating a verbal association between two words to facilitate later recall of one of the words when the other is presented. The task consists of 24 pairs of concrete words such as 'bicycle-snake', that are presented for five seconds. Instructions are given to generate a mediator (e.g., 'A bicycle ran over the snake') to facilitate retrieval when the first word of each pair is presented in a subsequent recall test. Cube Comparison is a spatial task in which pictures of two cubes with markings on each face must be compared to determine if they are the same cube in different orientations or different cubes. The Raven's Matrices is an inductive reasoning test in which the goal is to decide which of eight possible choices best completes a matrix of abstract figures. In mental multiplication two 2-digit numbers are presented that must be multiplied mentally. Paired associates and mental multiplication were assumed to be least affected by concurrent verbalization among older adults because these tasks are involve minimal recoding into verbal strings, whereas cube comparison and the Raven's Advance Progressive matrices were expected to demonstrate reactivity because verbal labels for elements are not as easy to generate.

Cube Comparison and the Raven's Matrices were administered in paper and pencil format, while other tasks were administered on a computer. Participants were given as long as necessary to complete each task other than Cube Comparison so that the effect of verbalization condition on solution times could be assessed. Because Cube Comparison was timed (according to instructions), solution times were determined by the number of problems attempted in each of two 3-min problem sets.

Procedure

Each experimental session began with a warm-up procedure from Ericsson and Simon (1993) to acquaint participants with providing TA verbal reports. This procedure (p. 378) provides instructions for how to verbalize without describing or introspecting, and includes practice problems to acquaint participants with the method and alleviate initial evaluation apprehension. These practice problems (e.g., 'What is the result of multiplying 24×36?') allow the experimenter to establish that the participant understands the difference between thinking aloud and describing thought processes.

Participants performed the tasks in the following fixed order: paired-associates, Cube Comparison, Raven's Matrices, and mental multiplication. Each participant completed two tasks silently and two tasks while thinking aloud. Four overall verbalization orders were used to vary the order in which participants performed silently or while thinking aloud during an experimental session lasting about 1 h.

Results

Paired associates were scored in terms of total number of words recalled. Cube Comparison dependent variables were total number of items correct and total number of items attempted in two 3-min problem sets. Raven's Matrices and mental multiplication were scored with respect to overall scores and solution times in seconds. A multivariate analysis of variance (MANOVA) revealed no differences between the four verbalization orders on any of these dependent variables, $p > .05$. The order in which participants verbalized or remained silent on the four tasks did not affect scores or the speed at which tasks were performed.

A series of 2 (verbalization condition: TA/silent) × 2 (age group: older/younger) multivariate analyses of variance (MANOVA) were used to test for differences between groups on performance and speed for each task. Analyses of accuracy revealed a main effect for age on paired associates (younger: $M = 10.90$, $SD = 4.60$; older: $M = 5.39$, $SD = 4.14$), $F(1, 57) = 23.18$, $MSE = 19.73$, $p < .01$; Cube Comparison in part 1 (younger: $M = 9.87$, $SD = 3.61$; older: $M = 6.30$, 3.40), $F(1, 57) = 17.05$, $MSE = 11.63$, $p < .01$; and part 2 (younger: $M = 11.27$, $SD = 4.31$; older: $M = 7.90$, 3.78), $F(1, 57) = 12.24$, $MSE = 14.99$, $p < .01$; and mental multiplication (younger: $M = 2.33$, $SD = 1.56$; older: $M = 1.16$, 1.29), $F(1, 57) = 10.57$, $MSE = 2.05$, $p < .01$, as younger adults outperformed older adults on these tasks. Age group differences were not found for solution time on any of these tasks as younger adults did not attempt more problems in Cube Comparison or perform mental multiplication faster, $p > .05$, and paired associates was not timed. Verbalization condition did not impact scores or solution times on these tasks and did not interact with age group, $p > .05$. No significant reactivity emerged on paired associates, Cube Comparison, or mental multiplication.

Age group interacted with verbalization condition on the Raven's Matrices as older adults scored higher, $F(1, 57) = 5.64$, $MSE = 5.00$, $p < .05$, and took longer, $F(1, 57) = 4.28$, $MSE = 82{,}900$, $p < .05$, when thinking aloud relative to younger adults (see Figure 1). Simple effect contrasts revealed that TA and silent older adult differences in score (TA: $M = 4.07$, $SD = 2.76$; silent: $M = 2.35$, $SD = 1.90$), $F(1, 57) = 4.12$, $MSE = 5.64$, $p < .05$, $d = 0.73$; and solution time (TA: $M = 1080$, $SD = 477$; silent: $M = 719$, $SD = 224$), $F(1, 57) = 11.77$, $MSE = 82900$, $p < .05$, $d = 0.96$, were significant. Conversely, younger adults groups did not differ in scores (TA: $M = 5.40$, $SD = 2.72$; silent: $M = 6.67$, $SD = 2.40$) or solution times (TA: $M = 625$, $SD = 169$; silent: $M = 574$, $SD = 201$), $p > .05$. Contrary to predictions, TA older adults actually performed better than the silent group on this task. This finding was not predicted for the nonverbal Raven's Matrices test.

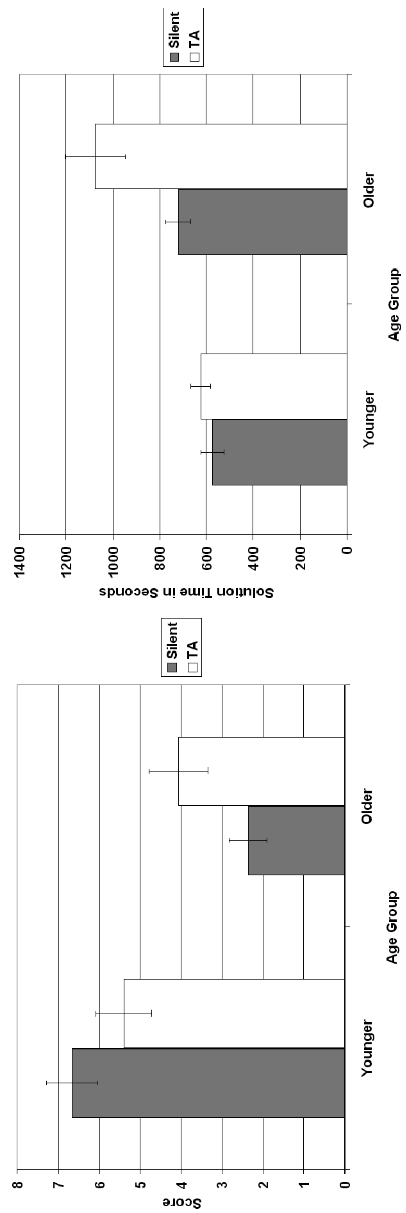


FIGURE 1. Scores and solution times on the Raven's Matrices from Experiment 1 with ± 1 standard error bars.

Discussion

Thinking aloud had little impact on total scores and the amount of time needed to complete three tasks. The only task in which reactivity was observed was the Raven's Matrices where older adults performed better when thinking aloud. Aside from this facilitation effect on the Raven's, these findings are consistent with previous studies (Gilhooly et al., 1999; Johnson, 1993; Perfect & Dasgupta, 1997) in showing older adults do not exhibit negative reactivity on a variety of tasks and generally support Ericsson and Simon's model. However, longer solution times were not observed for Cube Comparison as would have been predicted based on this task's nonverbal content. Moreover, older adults, if anything, may benefit from TA under certain conditions as their performance was better on the Raven's Matrices when thinking aloud. These findings seem to show that any additional processes involved in thinking aloud are resistant to cognitive decline, even on complex tasks with abstract and spatial stimuli.

An explanation for superior performance on the Raven's Matrices for TA older adults must account for the absence of reactivity on other tasks in both age groups as well as the finding that younger adult performance does not differ with condition on the Raven's Matrices. For example, it is unlikely that experimenter cues played a role because only one task was affected. One possibility is that TA engendered more systematic scanning and processing of the matrices as verbalizing may have alerted older adults to random scanning and processing as it occured. Previous studies have shown that verbalization can improve performance when it involves reasoning about a task prior to responding (e.g., Davis, Carey, Foxman, & Tarr, 1968; Gagne & Smith, 1962; Ray, 1957), presumably because attempting to articulate one's strategy verbally can expose it as incomplete or otherwise inadequate. However, non-descriptive TA reports like those collected in the present study have seldom been shown to improve performance in this fashion and such a theory cannot account for null findings among younger adults unless it is only the lowest performing younger adults.

A more age-specific theory is that that TA mitigates age-related declines in ability to inhibit irrelevant visual information. Hasher and Zacks (1988) proposed that many performance declines in older adults can be explained by an age-related inability to inhibit information that is irrelevant to task performance that leads to a lower signal-to-noise ratio during information processing. The typical Ravens problem shows a matrix with eight complex objects and a set of multiple choices with an additional eight such objects. Thus, there are many opportunities for attention to be attracted by answer objects when viewing problem objects, and vice versa. It is possible that conversion to a verbal code helped older adults to process information more efficiently from the matrices by avoiding visual distraction. An inhibition account of the pilot findings would lead to several predictions about performance on matrix reasoning problems and TA. If concurrent verbalization helps older adults to attend more selectively to relevant information, control and TA participants should perform more similarly when less irrelevant information is present in the matrix. For example, matrix reasoning problems consisting of two rows by two columns should elicit less facilitation for TA participants than conventional problems consisting of three rows by three columns because they contain less visual information. Moreover, if a third row and column containing information that will never be useful, are added to a two-by-two matrix, TA participants should perform better than silent participants.

This inhibition explanation was tested in Experiment 2 because it is consistent both with the Raven's Matrices being the only task affected by TA, and older, but not younger, adults benefiting. This account was tested in Experiment 2 with pairs of modified matrix reasoning problems that were identical except that some contained additional irrelevant figures and some did not. Because the hypothesis was not supported it will only be mentioned with respect to how it impacted the design of Experiment 2. The primary purpose of Experiment 2 was to rule out the possibility of type I error by replicating the findings of Experiment 1. Additionally we tested the generalizability of TA facilitation by changing how the Raven's Matrices were administered. This task was always performed third in Experiment 1 but was performed with position counterbalanced among the other tasks in Experiment 2. In addition to this change, we collected retrospective verbal reports after some of the problems in silent and TA groups, changing the task even more. We also re-administered another task from Experiment 1 (mental multiplication) for which no TA facilitation effect was observed. Higher scores for TA older adults on the Raven's Matrices but not mental multiplication would replicate Experiment 1 and suggest that TA facilitation of older adults is indeed specific to one task or one class of tasks.

EXPERIMENT 2

Method

Design

The experimental portion of the experiment was a 2 (verbalization condition: TA/silent) $\times$ 2 (visual distraction condition: distraction/control) mixed factorial that will not be described in detail because of null results. Participants performed the same 12 Raven's Matrices items and the mental multiplication task from Experiment 1 along with two experimental matrix reasoning tasks. Verbalization condition (silent/TA) was a between-subjects variable for these tasks. Dependent variables of interest were total scores and solution times on each of the four tasks.

Participants

Thirty older adults (16 female; mean age 73.0) were recruited from the community to participate in a study lasting approximately 1 h and were compensated $10 for participating. All participants were screened for dementia with the Wechsler Memory Scale III and the Short Portable Mental Status Questionnaire.

Materials

Testing materials were the abbreviated set of Raven's Matrices used in the initial study, two sets of modified matrix reasoning problems, and the mental multiplication task from the initial study. The Raven's Matrices were administered in paper and pencil format while all other tasks were computer administered.

Procedure

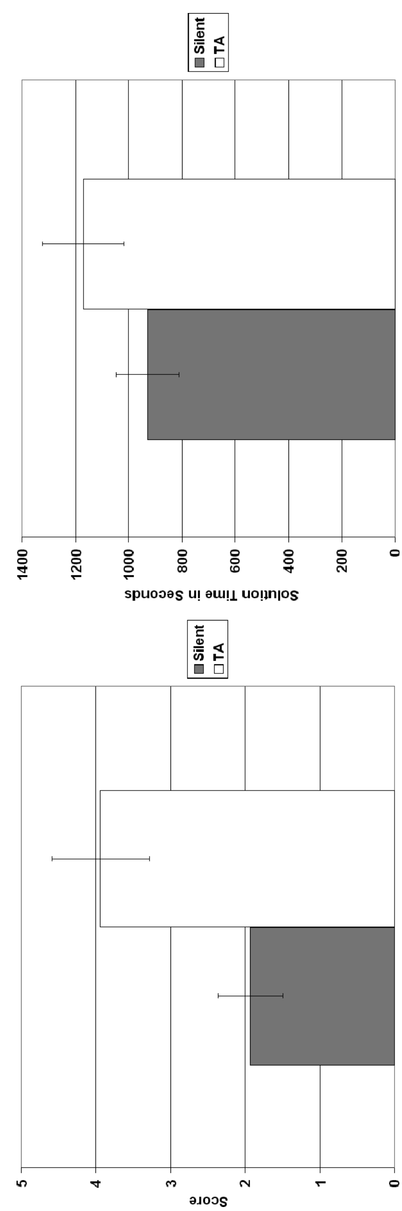
Each experimental session began with the same warm-up procedure outlined by Ericsson and Simon (1993) to acquaint participants with providing concurrent and retrospective verbal reports. The four tasks were completed in four different orders to test whether the pilot findings generalize beyond the fixed ordering of tasks in the first study.

Manipulation of TA consisted of dividing participants into two groups: One group thought aloud during the Raven's Matrices and mental multiplication and remained silent for the modified matrix reasoning tasks, and another group completed the matrix reasoning tasks aloud and remained silent on the Ravens Matrices and mental multiplication. All tasks were administered without a time limit to assess the impact of TA on solution times. Additionally, participants provided retrospective reports for four of the 12 problems from the Raven's Matrices (problem numbers 3, 15, 22, and 34).

Results

Multivariate analysis of variance (MANOVA) revealed no difference between the four orders on accuracy or solution time for any tasks, $p > .05$. A one-way ANOVA revealed a significant difference between TA (M = 3.94, SD = 2.62) and silent (M = 1.93, SD = 1.64) groups on Raven's Matrices Scores $F(1, 28) = 6.12$, $MSE = 4.92$, $p < .05$, $d = .92$, but not solution time (TA: M = 1170, SD = 610; silent: M = 929, SD = 436), $p > .05$. This finding partially replicates Experiment 1 as TA participants had higher scores on both measures. However, the null finding for solution time may be due to lack of power because there was a trend toward replication of longer solution times for the TA group. The effect size was smaller than before ($d = 0.45$ as opposed to $d = 0.96$ in the Experiment 1), resulting in less power ($1 - \beta = .22$) and thus a high probability of Type II error. Mean scores and overall solution times are depicted in Figure 2. Null findings for

FIGURE 2. Scores and solution times on the Raven's Matrices from Experiment 2 with ± 1 standard error bars.



mental multiplication were also replicated, as scores (TA: $M = 1.75$, $SD = 1.65$; silent: $M = 1.50$, $SD = 1.29$), and solution times (TA: $M = 48.7$, $SD = 21.5$; silent: $M = 58.6$, $SD = 27.9$) were unaffected by verbalization condition, $p > .05$.

The inhibition hypothesis predicted that TA should facilitate performance in the distraction condition of the modified matrix reasoning task and have little effect in the control condition. Contrary to this prediction, modified matrix reasoning was not impacted by verbalization condition or visual distraction condition as TA did not significantly improve scores, prolong solution time per problem, or interact with visual distraction, $p < .05$. One possible explanation for not detecting the expected interaction is insufficient power as there was a minor trend toward the expected finding. However, the experiment should have had adequate power with its within-subject design to detect the inhibition effect if it is indeed the explanation behind the substantial effect sizes observed on the Raven's Matrices from TA. Once again, the Raven's Matrices was the only task impacted by thinking aloud.

Retrospective reports did not reveal strategic differences between conditions for solving Ravens problems. It is possible that only four retrospective reports for the twelve items were not enough to obtain precise process data. For example, TA reports contained more descriptions of correct answer elements than silent reports, as would be expected if TA participants inferred more correct answers. However, the difference was not statistically significant.

Discussion

Experiment 2 partially replicated the initial study as older participants performed better on the Raven's Matrices while thinking aloud, but not the other tasks. However, unlike Experiment 1, the TA group did not take significantly longer to complete the task. Total scores and solution times were unaffected by verbalization condition on the mental multiplication task and the modified matrix reasoning task. Experiments 1 and 2 provide converging evidence that TA facilitates performance on the Ravens but does not impact other tasks.

It is unclear why the TA group did not have significantly longer solution times in Experiment 2 as they did in Experiment 1. However it is notable that results were trending in this direction and it is possible that lack of power played a role in this null result.

Experiment 2 findings demonstrate that TA facilitation of the Raven's Matrices among older adults generalizes to experimental conditions differing from those of Experiment 1. Participants in Experiment 2 performed the Raven's matrices either first, second, third, or fourth among other tasks as opposed to the initial study in which this task was always performed third. Moreover, participants generated retrospective verbal reports in Experiment 2 while not doing so in Experiment 1. It appears that older adult facilitation on the Raven's matrices is not only replicable, but robust enough to manifest under a variety of experimental conditions.

GENERAL DISCUSSION

Findings from these experiments suggest that age and type of task influence reactivity as it was shown that older adults perform better on at least one task when thinking aloud. This finding was then replicated in a second experiment. Conversely, younger adults did not exhibit reactivity on any task. Older adults appear to perform the same on a variety of cognitive tasks whether silent or thinking aloud. Null results on tasks other than the Raven's Matrices are consistent with previous studies (Gilhooly et al., 1999; Johnson, 1993; Perfect & Dasgupta, 1997) showing that older adults rarely exhibit reactivity to concurrent verbal reports. While Ericsson and Simon's process model cannot account for older adult facilitation on the Raven's Matrices, it is generally consistent with other findings in both experiments.

The initial hypothesis that older adults may perform worse while thinking aloud due to declines in processing speed and deficits in dual-task performance was not supported. They did not exhibit any negative reactivity and, in fact, showed improved performance on the Raven's Matrices when thinking aloud. An inhibition-related explanation was tested to explain the effects of thinking aloud on older adult performance and was unsupported. It is possible that further experiments may implicate an inhibition account of why older adults perform better on the Raven's Matrices while thinking aloud, but no such evidence was forthcoming in the second experiment. Future studies using alternative process-tracing methods like eye-tracking will allow specific hypotheses to be tested for why TA seems to improve older adult performance on this task.

The possibility remains that TA facilitation is explicable without recourse to an age-related explanation. Several studies (Short, Evans, Friebert, & Schatschneider, 1991; Short, Schatschneider, Cuddy, & Evans, 1991) have reported higher performance among children who verbalize, but not college-age adults on verbal and spatial analogies. A relationship between these studies and the findings of the current investigation should be interpreted with caution given that Short et al.'s descriptive verbal reports did not conform to the guidelines of Ericsson and Simon. Nonetheless, these studies provide some evidence that individuals with lower fluid ability (e.g., children and older adults) may benefit most from concurrent verbalization. Future studies may reveal that TA older adults benefit on the verbal and spatial analogies tests used in Short et al.'s studies. It is possible that the present effects could be increased using explanatory reports, as previous studies show a general positive impact of explanatory verbalization on performance in younger populations (e.g., Davis et al., 1968; Gagne & Smith, 1962; Ray, 1957).

Of note were the effect sizes for TA on Raven's scores (d = 0.73 and 0.92 for Experiments 1 and 2, respectively). The standard deviations provided by Bors and Stokes (1998), and those obtained in the present study, are all roughly 2.5. Using an IQ scale with a standard deviation of 15, these effects (averaged) correspond to an IQ gain of about 11 points. A difference of this magnitude wields significance in both theoretical and applied domains. Previous studies (Bors & Vigneau, 2003; Raven, 1938) have revealed gains of one testing session to be about the same as the facilitation effect described here. Importantly, TA facilitation may be less test-specific than practice or training and thus generalize beyond the Raven's Matrices. It is possible the effect may influence other psychometric tests and have practical applications in training and usability. Older adults acquire new skills more slowly and are less proficient with many technologies that could improve their quality of life (Fisk, Rogers, Charness, Czaja, & Sharit, 2009). Interventions may help older adults adapt to these technologies (Fisk et al., 2009), and verbalization could provide a cheap and relatively instantaneous supplement to these programs if it is found to facilitate performance more generally.

Nevertheless, the extent to which the facilitation effect generalizes remains unknown. Participants attempted to solve the same 12 problems in both studies and it is possible that specific characteristics of these problems drove the effect. Additional studies will be needed to determine whether the effect persists with the full set of 36 Raven problems or other matrix reasoning problems. Whether or not TA facilitates real-world tasks remains to be discovered.

The major contribution of this study has been to demonstrate that potential for reactivity is determined in part by age group and task, and to show that negative reactivity among older adults is unlikely to occur on most cognitive tasks, at least when instructions are used that elicit non-manipulative reports (as opposed to explanatory or introspective reports). Moreover, older adults have been shown to perform better on at least one task while thinking aloud. Future studies will be needed to identify the determinants of improvement while thinking aloud and establish whether this relatively simple-to-follow instruction can be used to improve other tasks with greater practical significance.

Original manuscript received September 26, 2008
Revised manuscript accepted May 11, 2009
First published online August 3, 2009

REFERENCES

Biehal, G., & Chakravarti, D. (1989). The effects of concurrent verbalization on choice processing. *Journal of Marketing Research, 26,* 84–96.

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement, 58*, 382–398.

Bors, D. A., & Vigneau, F. (2003). The effect of practice on the Raven's Advanced Progressive Matrices. *Learning and Individual Differences, 13*, 291–312.

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 467–476.

Davis, J. H., Carey, M. H., Foxman, P. N., & Tarr, D. B. (1968). Verbalization, experimenter presence, and problem solving. *Journal of Personality and Social Psychology, 8*, 299–302.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: The MIT Press.

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2009). Designing for older adults: Principles and creative human factors approaches. Boca Raton, FL: CRC Press.

Gagne, R. H., & Smith, E. C. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology, 63*, 12–18.

Gilhooly, K. J., Philips, L. H., Wynn, V., Logie, & Della Sala, S. (1999). Planning processes and age in the five-disce Tower-of-London task. *Thinking and Reasoning, 5*, 339–361.

Hasher, L. & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 193–225). San Diego, CA: Academic Press Inc.

Johnson, M. S. (1993). Thinking about strategies during, before, and after making a decision. *Psychology and Aging, 8*, 231–241.

Perfect, T. J., & Dasgupta, Z. R. R. (1997). What underlies the deficit in reported loss of recollective experience in old age? *Memory & Cognition, 25*, 849–858.

Raven, J. C. (1938). *Progressive Matrices: A perceptual test of intelligence. 1938, sets, A, B, C, D, and E*. London: H. K. Lewis.

Raven, J. C. (1965). *Advanced Progressive Matrices, Sets I and II*. London: H. K. Lewis.

Ray, W. S. (1957). Verbal compared with manipulative solution of an apparatus-problem. *American Journal of Psychology, 70*, 289–290.

Riby, L. M., Perfect, T. J., & Stollery, B. T. (2004). The effects of age and task domain on dual task performance: A meta-analysis. *European Journal of Cognitive Psychology, 16*, 868–891.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition, 17*, 759–769.

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

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*, 166–183.

Short, E. J., Evans, S. W., Friebert, S. E., & Schatschneider, C. W. (1991). Thinking aloud during problem solving: Facilitation effects. *Learning and Individual Differences, 3*, 109–122.

Short, E. J., Schatschneider, C. W., Cuddy, C. L., & Evans, S. W. (1991). The effect of thinking aloud on the problem-solving performance of bright, average, learning disabled, and developmentally handicapped students. *Contemporary Educational Psychology, 16*, 139–153.