Do intensive studies of a foreign language improve associative memory performance?

Johan Mårtensson* and Martin Lövdén1,2,3

1 Department of Psychology, Lund University, Lund, Sweden
2 Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee, Berlin, Germany
3 Aging Research Center, Karolinska Institutet and Stockholm University, Gävlegatan, Stockholm, Sweden

Edited by:
Wilfried Kunde, Julius-Maximilians-Universitaet Wuerzburg, Germany

Reviewed by:
Markus Janczyk, University of Würzburg, Germany
Ralf Rummer, Universität Erfurt, Germany

*Correspondence:
Johan Mårtensson, Department of Psychology, Lund University, P.O. Box 213, SE-221 00 Lund, Sweden.
e-mail: johan.martensson@psychology.lu.se

INTRODUCTION

Education has been proposed to shape life-long cognitive development. For example, formal schooling might improve intelligence (Ceci, 1991), be responsible for generational increases in memory performance (cf. Flynn, 1987; Rönnlund et al., 2005), and serve as a protective factor against dementia (Stern, 2002). The causal mechanisms are however uncertain. Indirect effects are plausible, including that lack of schooling leads to life conditions that negatively affect intellectual health. More direct effects are also conceivable: Education may convey knowledge (e.g., metacognitive strategies) or foster improved efficiency of cognitive processes (e.g., sustained attention) relevant for performing tests of intellectual capacity (Ceci, 1991).

Recent experimental studies administering practice on working memory (Olesen et al., 2004; Dahlin et al., 2008; Jaeggi et al., 2008) and executive tasks (Persson and Reuter-Lorenz, 2008; Karbach and Kray, 2009) in a controlled setting to younger university students have reported that practice-related gains in performance can transfer to tasks where acquired knowledge appears to be of relatively minor relevance. For example, Jaeggi et al. (2008) reported improvements on a measure of figural reasoning from practicing a demanding working-memory task (performing two n-back tasks simultaneously). Notably, participants trained between 8 and 19 days with transfer to the reasoning task varying as a function of training time. These findings, and other similar ones (Schmiedek et al., 2010), indicate (a) that training dosage is a critical factor for transfer to occur; and (b) lend some credibility to the notion that formal education might exert some parts of its effects on intellectual capacity through direct mechanisms that improve processing efficiency.

Notably, other types of cognitive training, in particular mnemonic instructions and practice on associative memory tasks, have, in contrast to the more successful studies on working memory and executive control practice, been largely unsuccessful in producing transfer effects (see Noack et al., 2009; Lövdén et al., 2010, for reviews). For example, a large body of literature have administered instructions and short practice on the method-of-loci mnemonic, which requires imagery and associative binding of to-be-learned concrete words to imagined locations, and have reported negligible transfer effects (see Verhaeghen et al., 1992, for meta-analysis). Associative memory is a component of episodic memory, which refers to mechanisms that bind different aspects of an event into an episode during encoding, storage, and retrieval (Zimmer et al., 2006). Considering that the key neural substrate of associative memory and knowledge acquisition, the hippocampus, is widely regarded to be a particularly plastic brain region, and displays remarkable structural plasticity in response to intensive periods of formal education (Draganski et al., 2006), it is unsettling that previous studies of associative memory training have not detected transfer effects. This state may suggest that previous studies have been sub-optimally designed. Specifically, the focus of training has often been on teaching mnemonic strategies rather than on rote practicing, and the dosage of practice may have been too low.

We investigated whether extremely intensive study of a foreign language, which involves demands on associative memory, causes improvements in associative memory performance. To this end, we studied conscript interpreters at the Swedish Armed Forces Intelligence and Security Centre. These individuals are handpicked (based on intelligence, language proficiency, and mental stability) from the drafted population consisting of all Swedish males, along with women who choose to undergo military training. The extreme intensity of the studies makes the interpreters’ situation different from an average university students’. Starting from scratch, the interpreters become fluent in a foreign language (Arabic or Persian).
and individual studies, interleaved with basic soldier training, from 08.00 until bedtime. Weekends are even more intense with individual studies throughout the day.

Studies at the academy, like most language studies, strongly focus on acquiring the foreign vocabulary by studying new words that have to be associated with prior semantic knowledge. According to the professors at the academy, the training contrasts with normal language studies in that the interpreters acquire between 300 and 500 new words each week. If studies in a formal setting can, in principle, lead to improvements in associative memory performance, then it should emerge in this highly specialized learning environment with its extreme demands.

To investigate the effects of interpreter training we employed a pretest-training-post-test design. In order to separate interpreter-related effect on performance from mere retest effects, maturation, or other confounds, we included a control group of students from non-language introductory university courses. Pretest and post-tests included a battery of cognitive tasks. In this test battery, we include measures of associative memory performance, working memory performance, strategic episodic memory performance, and reasoning. Working memory measures were included because we, in addition to investigating effects of language acquisition on associative memory, wanted to explore if the practice of switching between different languages during interpretation sessions, which place large demands on working memory processing (e.g., Daro and Fabbro, 1994), would lead to improved working memory performance. The measure of reasoning and the strategy-loaded episodic-memory tasks were included in order to demonstrate the equivalence of the interpreter and control groups in terms of the general magnitude of retest effects. We did not expect larger improvements for interpreters than for controls on these tasks. Otherwise, we predicted larger improvements from educational training for the conscript interpreters than for controls on these tasks. Otherwise, we predicted larger improvements from educational training for the conscript interpreters than for the university students.

**MATERIALS AND METHODS**

**PARTICIPANTS**

Fifteen conscript interpreters (mean age = 19.6; SDage = 0.4; 4 women) and 19 university-student controls (mean age = 20.1; SDage = 0.2; 9 women) volunteered to participate in assessments of cognitive performance before (pretest) and after (post-test) the three first months of studies at the interpreter academy in Uppsala and at introductory courses at Lund university, respectively.

**MATERIALS AND PROCEDURE**

Eight cognitive tasks were administered at pretest and post-test: three paired-associates tasks targeting associative memory (word–non-word cued recall, word–word recognition, and face–name cued recall), followed by two strategic episodic memory tasks (word free recall and pictorial object free recall), two working memory tasks (numerical n-back; Kirchner, 1958 and alpha span; Craik, 1986), and a fluid intelligence task (Raven’s matrices; Raven, 2000).

**Word–non-word cued recall**

This task consisted of 44 combinations of Swedish words and non-words presented on a computer screen as pairs for 4 s each with an equal number of combinations starting with a Swedish word and a non-word. During the test phase the first word in the word-pair was presented, and the participant had to type in, without time constraints, the second word. The dependent variable was the number of correctly recalled word-pairs.

**Word–word recognition**

This task consisted of 20 Swedish word–word pairs presented in pairs for 4 s each. During the test phase 60 word-pairs consisting of (1) presented combinations of words, (2) rearranged combinations of words, and (3) new word-pairs were presented and the participant was instructed to respond whether or not the exact presented combination had been present during the study phase. There was no time-limit. The dependent variable was computed by subtracting the percentage of correctly recognized presented word-pairs (Hits); correct categorization of new items (i.e., distracters) were not used.

**Face–name cued recall**

In the encoding phase, 44 face–name combinations were presented for 4 s each. The faces were displayed in black and white with the given name presented beneath. During the test phase the faces appeared again, in rearranged order, and the participant was instructed to type in the correct name for each individual, with 30 s allotted per presented face. The dependent variable was the number of correctly recalled names.

**Word free recall**

This test consisted of 30 Swedish words presented for 4 s each at encoding. At retrieval, participants typed in as many words as they could recall during 4 min. The dependent variable was the number of correctly recalled words.

**Pictorial object recall**

In this task 35 images were shown for 4 s each. Every image contained a central object (such as a dolphin, a bus, or a chair). During the test phase, 4 min were given to type in the names of the central objects. The dependent variable was the number of correctly recalled objects.

**Numerical n-back**

In the numerical n-back task the participants were shown four blocks of 39 numbers presented for 2 s each, with a break in between the blocks. The first two blocks were 3-back tasks. The participant responded with a yes response (via the keyboard) if the number shown on the screen was the same as the number shown three steps back in the series of digits and with a no response if it was not. The last two blocks consisted of a 4-back task. Dependent variables were the number of correct decisions and the mean reaction time for correct decisions.

**Alpha span**

Five blocks of 10 combinations, each consisting of a letter followed by a digit and a question mark (e.g., K 4?), were presented. The participant was instructed to remember the letters (in this case K) and change their order amongst the prior shown items according to the alphabet (in case K was preceded by B and F in a series, K would now be number 3 in order), following which they were to decide whether the presented number was the correct order based on the
earlier presented combinations of a given series (and in this case, if 
K is number 3, the correct response to K 4? would be no). For each 
combination of letter, digit, and question mark, 4 s were given for the 
4 first blocks, whilst for the fifth block 3 s per combination was allot-
ted. The dependent variable was the number of correct responses.

**Raven’s matrices**
The 18 odd-numbered items from set II of this test were presented. 
Ten minutes in total were allotted to complete the task, and the 
dependent variable was the number of correctly selected patterns. 
The same items were used at both pre- and post-test.

**STATISTICAL ANALYSES**
Separate 2 (group; interpreter vs. control) × 2 (time; pretest vs. 
post-test) mixed ANOVAs for each task were performed, target-
ing the critical group by time interaction indicating differential 
changes over time for one group as compared with the other. The 
threshold for statistical significance was \( p < 0.05 \). Unfortunately, 
the word–non-word task displayed a floor effect and the word– 
word test displayed a ceiling effect (see Figure A1 in Appendix for 
illustration of floor/ceiling effects and Table A1 for a full ANOVA 
table). These tests were thus not further analyzed.

**RESULTS AND DISCUSSION**
A significant group by time interaction was observed for the face– 
name cued recall task only, \( F(1,32) = 5.16, p = 0.03, \eta^2 = 0.14 \), 
depending larger improvements for interpreters than for students 
(see Figure 1). This face–name task, which is strongly tied to hip-
pocampal functioning (Sperling et al., 2003), is in the format quite 
different from the tasks that the interpreters are exposed to during 
the training, although it is the administered test, of those that show 
acceptable psychometric properties, that overlaps most closely with

![Figure 1](https://example.com/figure1.png)

**FIGURE 1** | Face–name paired-associates performance increase more for interpreters than for university students (A). There were no group-differences in change for working memory (B, C, and G), strategic episodic memory (D and E), or fluid intelligence (F). Error bars represent SEM.
the associative memory processes required when acquiring a foreign vocabulary. The net effect size of the observed transfer effect, subtracting the improvements for the university students and normalized against pretest SD for both groups, was 0.8 SD, corresponding to an effect that is considerably larger than the transfer effects reported in interventions studies including mnemonic instructions or other types cognitive practice (Noack et al., 2009). If the results were to be put in a real life situation corresponding to the face–name test, such as attending a cocktail party and attempting to learn the names of all the guests, the interpreters would remember 28% more names relative to controls after 3 months of training.

The observed floor effect of the word–non-word task as well as the ceiling effect of the word–word task deserves further mention. Since no prior study had been conducted on this highly select group, the task of designing each test with the aim to tax the interpreters whilst still being manageable for the control group came with both some challenge and uncertainty. These two tasks ended up as too difficult and too trivial, respectively, which unfortunately makes it difficult to interpret changes in these tasks, and to validate the findings on the face–name task with other associative memory tasks.

Though the interpreters significantly outperformed the students on most of the tasks (see Figure 1), group differences in performance of the face–name task at pretest did not account for the group-differential changes over time in this task, as revealed by a significant group by time interaction in a repeated measures ANCOVA with pretest performance as covariate, \( F(1,31) = 5.8, p = 0.02, \eta^2 = 0.16 \). Moreover, no selective increases for interpreters as compared to students were observed on the other tests, all \( F_s < 1.06 \), indicating that the two groups benefited equally from the mere repeated testing experience. Thus, a group difference in the magnitude of retest effects is not a viable explanation for the interpreter-selective effects on the face–name task. In addition, the absence of transfer effects on the strategic episodic memory tasks (word free recall and pictorial object free recall), in which mnemonic strategies may play a sizable role, suggests that selective strategy changes for the interpreters as compared to the students might not play a major role for the improvements in the face–name task. Note, however, that it is difficult to exclude that the interpreters acquired strategies that were selectively applicable to associative memory tasks.

We conclude that intensive language studies in a formal setting may improve associative memory performance. Formal education can improve cognitive performance through direct mechanisms and associative memory performance can be improved in early adulthood. Though this finding is important, the study rate at the interpreter academy is extreme, with very motivated students and highly skilled teachers dedicated to the student’s progression. Future studies should examine effects of less extreme educational situations, including the direct effects of less intensive language studies on associative memory ability and life-long development of memory performance.

ACKNOWLEDGMENTS

Financed by the Sofja Kovalevskaja Award to Martin Lövdén and by the Swedish Research Council. Thanks to Mikael Johansson, Ulman Lindenberger, Magnus Lindgren for comments. Thanks also to Stefan Kindblad and the teachers at the Interpreter Academy.

REFERENCES


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 August 2010; accepted: 12 January 2011; published online 26 January 2011.


This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2011 Mårtensson and Lövdén. This is an open-access article subject to an exclusive license agreement between the authors and Frontiers Media SA, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.
APPENDIX

A

Interpreters: Word–non-word Pretest

Controls: Word–non-word Pretest

Interpreters: Word–non-word Post-test

Controls: Word–non-word Post-test

FIGURE A1 | Continued.
FIGURE A1 | There was restriction in range for the word–non-word task (A), which showed a floor effect as well as in the word–word task (B) which showed a clear roof effect.
Table A1 | Repeated measures ANOVA of main effects and Interaction effects for all measures.

<table>
<thead>
<tr>
<th>Measure and effects</th>
<th>df</th>
<th>F</th>
<th>Partial η²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WORD–NON-WORD CUED RECALL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>45.75</td>
<td>0.59</td>
<td>0.00**</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>1.36</td>
<td>0.04</td>
<td>0.25</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>32</td>
<td>(3.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WORD-WORD RECOGNITION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>9.03</td>
<td>0.22</td>
<td>0.00**</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>0.79</td>
<td>0.03</td>
<td>0.38</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>32</td>
<td>(5.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FACE–NAME CUED RECALL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>18.08</td>
<td>0.36</td>
<td>0.00**</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>5.16</td>
<td>0.14</td>
<td>0.03*</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>32</td>
<td>(7.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WORD FREE RECALL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>36.49</td>
<td>0.53</td>
<td>0.00**</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.98</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>32</td>
<td>(9.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PICTORIAL OBJECT RECALL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>59.12</td>
<td>0.65</td>
<td>0.00**</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>0.30</td>
<td>0.01</td>
<td>0.30</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>32</td>
<td>(5.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NUMERICAL n-BACK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>8.64</td>
<td>0.21</td>
<td>0.01**</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>1.06</td>
<td>0.03</td>
<td>1.06</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>32</td>
<td>(118.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ALPHA SPAN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>7.95</td>
<td>0.20</td>
<td>0.01**</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>0.08</td>
<td>0.00</td>
<td>0.78</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>32</td>
<td>(6.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RAVEN’S MATRICES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>13.71</td>
<td>0.3</td>
<td>0.00**</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>0.55</td>
<td>0.02</td>
<td>0.46</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>32</td>
<td>(3.40)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values enclosed in parentheses represent mean square errors. *p < 0.05, **p < 0.01.