



Universität Potsdam

Reinhold Kliegl, Jacqui Smith, Paul B. Baltes

Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill

first published in:
Developmental Psychology, 25 (1989) 2, S. 247-256, ISSN 1939-0599,
DOI 10.1037/0012-1649.25.2.247

Postprint published at the Institutional Repository of the Potsdam University:
In: Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 153
<http://opus.kobv.de/ubp/volltexte/2009/4031/>
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-40311>

Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 153

Testing-the-Limits and the Study of Adult Age Differences in Cognitive Plasticity of a Mnemonic Skill

Reinhold Kliegl, Jacqui Smith, and Paul B. Baltes
Max Planck Institute for Human Development and Education
Berlin, Federal Republic of Germany

Investigated the range and limits of cognitive reserve capacity as a general approach to the understanding of age differences in cognitive functioning. *Testing-the-limits* is proposed as a research strategy. Data are reported from 2 training studies involving old (65 to 83 years old) and young adults (19 to 29 years old). The training, designed to engineer an expertise in serial word recall, involved instruction and practice in the Method of Loci. Substantial plasticity was evident in pretest to posttest comparisons. Participants raised their serial word recall several times above that of pretest baseline. Age-differential limits in reserve capacity were evident in amount of training gain but not in responses to conditions of increased test difficulty (speeded stimulus presentation). Group differences were magnified by the training to such a degree that age distributions barely overlapped at posttests. Testing-the-limits offers promise in terms of understanding the extent and nature of cognitive plasticity.

One of the perennial and unresolved questions of cognitive aging research is the distinction between performance and what has been variously labeled as *competence*, *latent potential*, or *reserve capacity*. Proposals for an empirical investigation of these concepts have focused on the study of the range and conditions of plasticity (Baltes, 1987; Brown, 1982; Denney, 1984; Engeström, 1986; Gollin, 1981; Kliegl & Baltes, 1987; Lerner, 1984; Vygotsky, 1978; Wiedl, 1984; Willis, 1985). There is agreement that tasks and experiential contexts need to be identified that reflect a person's potential range of cognitive functioning. With respect to developmental and aging research, a main goal is to examine whether there are age changes in the range (and limits) of plasticity.

In past work we have formulated a heuristic framework to guide the study of cognitive plasticity (Baltes, 1987; Kliegl & Baltes, 1987). We have distinguished, for example, between three levels of information about performance and latent potential: baseline performance, baseline reserve capacity, and developmental reserve capacity. Baseline performance refers to what individuals can do on a given task under standardized conditions of assessment. Baseline reserve capacity is what individuals are capable of performing if the conditions of assessment

are optimized without any effort aimed at altering the cognitive and motivational repertoire available to a person. It is a measure of current maximum performance potential or plasticity. Developmental reserve capacity, finally, involves the assessment of performance following interventions or experiences that are aimed at optimizing an individual's cognitive and motivational potential or reserves.

In order to obtain information about the range of performance associated with these levels of reserve capacities, we have adopted a general methodology called *testing-the-limits* (M. Baltes & Kindermann, 1985; Baltes, 1987; Guthke, 1982; Kliegl & Baltes, 1987; Schmidt, 1971). In our work, testing-the-limits involves a class of intervention strategies aimed at the systematic exploration of the effects of theory-guided practice and other procedures of cognitive engineering. By the use of testing-the-limits methodologies it is expected that we approximate, step by step, maximum levels of performance potential (developmental reserve capacity). As a result, we obtain a fine-grained picture of what participants would be able to do under "idealized" experiential conditions. We have also argued that developmental age-related differences in potential or reserve capacity should be most clearly identifiable near limits of performance (Kliegl & Baltes, 1987).

The purpose of the present experiments is to illustrate the use of the testing-the-limits methodology in the domain of memory functioning. The paradigmatic task chosen for the experiments that are reported here was the acquisition of a mnemonic skill known as the Method of Loci (Bower, 1970; Volkman, 1929; Yates, 1966). Skilled use of the Method of Loci involves a complex range of operations. For successful serial word recall involving the Method of Loci, both imagery-based elaborative encoding and the overlearning of a mental map as a reusable set of encoding and retrieval cues are critical. The fact that old adults can acquire this mnemonic device and increase their level of recall has been demonstrated in a number of short-term

This research was conducted in the context of the "Expertise and Cognitive Aging" project codirected by Paul B. Baltes and Reinhold Kliegl. We wish to express our gratitude to Annette Rentz and Werner Scholtysik for research assistance and coordination. We also thank the following research assistants: Boris Baltes, Gregor Caregnato, Jackie Durfee, Anita Günther, Sabine Edler, Stephan Lempert, Stefan Pfeffer, Elvira Pichler, and Dirik Sahn. Comments by K. Anders Ericsson and Laura A. Thompson are gratefully acknowledged. We are especially grateful to our participants for their patience, interest, and persistence.

Correspondence concerning this article should be addressed to Reinhold Kliegl, Max Planck Institute for Human Development and Education, Lentzeallee 94, D-1000 Berlin 33, Federal Republic of Germany.

training studies (Anschutz, Camp, Markley, & Kramer, 1985, 1987; Heineken & Gekeler, 1985; Robertson-Tschabo, Hausmann, & Arenberg, 1976; Rose & Yesavage, 1983; Yesavage & Rose, 1984). Of these studies only that by Rose and Yesavage (1983) compared the training gains of several age groups. Young adults were shown to benefit slightly but significantly more (i.e., 7%) from three training sessions than old adults. The direction of this effect is in agreement with our expectations for the present study: We expect that providing participants with extended training should lead to a clear separation of age groups, involving minimal distributional overlap.

In this article, we report data from two experiments. In the first experiment, young and old adults were trained to use the Method of Loci to encode and retrieve ordered information. Although participants in this study practiced using the Method of Loci technique to recall ordered lists of various item types, we will report here only the results concerning the application to serial word recall.¹ In general, and in line with our conceptualization of testing-the-limits methodology, we expected to obtain two outcomes. First, we expected a sizeable degree of reserve capacity (plasticity) in both age groups as evident in high levels of task performance following training. Second, we expected that, despite such sizeable reserve capacity, age differences would be magnified in task conditions that required functioning near limits of reserve capacity (i.e., in this instance, for serial word recall using a well-practiced mnemonic skill under speed constraints). The second experiment was designed to replicate the results of the first experiment with a different training program.

Experiment 1

Method

Participants

Two groups participated in the first training program: 20 healthy old adults ($M = 71.7$ years; range = 65–83 years) and 4 young university students ($M = 22.8$ years; range = 20–24 years). Of the original 22 old participants, 2 withdrew early in the program for health reasons. Because our general expectation was one of fairly robust aging loss at high levels of functioning, we decided to sample in a manner that favored falsification of the expectation of conspicuous aging loss. Thus, we selected only well-functioning adults and a larger number of old adults than young adults, thereby increasing the likelihood that some old adults might outperform the young. All participants were naive to memory training.

The old adults had participated in an earlier cognitive training study in Berlin and had scored above the 70th percentile on an intelligence test administered as part of that project (Baltes, Dittmann-Kohli, & Kliegl, 1986). They reported being in good health, were highly motivated, displayed a high degree of involvement, and perceived the training program as a personal challenge. On the Hamburg–Wechsler Intelligence Scale (HAWIE) the old adults performed at an IQ level that was one to two standard deviations above the average for their age group ($M = 119.2$; $SD = 4.9$). In terms of IQ components, they conformed to the usual pattern of healthy aging, with high verbal ($M = 58.4$; $SD = 4.8$) but only average performance ($M = 49.5$; $SD = 6.1$) scores. Their average forward memory span was 6.6 ($SD = .7$).

The young adults were recruited through a newspaper advertisement and were screened according to their performance on the HAWIE. The selection criterion was that they scored in the same range as the old

adults (young $M = 119.5$; $SD = 4.2$). Their verbal score ($M = 60.8$; $SD = 8.6$) and their memory span ($M = 7.3$; $SD = 1.5$) were comparable with those of the old adults; their performance scores, in contrast, were significantly higher than those of old adults ($M = 66.3$; $SD = 4.9$), $t(22) = 5.2$, $p < .01$. All participants were paid 20 Deutsche Mark per session.

Materials: Method of Loci and Stimuli

Method of Loci. The Method of Loci is a mnemonic technique. It involves the use of a highly familiar ordered sequence of locations of a mental map as a structure for encoding and retrieving new information (cf. Bower, 1970; Volkmann, 1929; Yates, 1966). Items that are to be remembered are successively linked (e.g., via an imagined association) with locations from the mental map. At recall, one mentally revisits the ordered locations and retrieves the items imagined.

The locations selected for this study consisted of a set of 40 well-known landmarks in West Berlin (e.g., the Berlin Wall, Memorial Church, Television Tower, etc.). These landmarks were learned according to the order in which they would be visited on a sightseeing trip of the city. Photographs of different aspects of each landmark were used in training.

Stimuli. Several different sets of item lists (words, historical events, and the dates of these events) were used in training and in testing participants' acquisition of the Method of Loci. The assumption was that use of the mnemonic with varying forms of content would facilitate acquisition and effective use.

Lists of words were used in the first stage of the training program and for the various pretests and posttests described below. These lists were generated from a pool of 1,560 concrete nouns extracted from a German dictionary. An estimation of imagery and concreteness ratings, analogous to the Paivio, Yuille, and Madigan (1968) ratings, was available for 150 of these words (Baschek, Bredenkamp, Oehrle, & Wippich, 1977). For this subsample of 150, all were rated higher than 5 on both scales (the maximum score was 7). The number of words per list varied depending on the training stage. For the test sessions, 40 words were included in each list (i.e., 1 for each location on the mental map).

Lists of historical events (e.g., the French Revolution, Columbus discovers America, the end of World War II) were also used in training the skilled use of the Method of Loci mnemonic. Participants learned the years associated with 100 historical events in preparation for a later training stage. The lists of historical events presented to subjects to recall were randomly selected (with replacement) from this set of 100. Again, the number of events per list varied (to a maximum of 40 events) depending on participants' performance.

Training and test sessions were conducted on Apple-IIe computers. Words and to-be-remembered historical event titles were presented in standard 40-column Apple font in the center of the computer monitor.

Procedure

Participants were scheduled to participate in a maximum of 26 individual training and 6 standard test sessions (2 pretests and 4 posttests) lasting about 90 min each. The full set of 26 training sessions was implemented only if subjects did not reach a criterion of 80% correct serial recall with the maximum number of locations and items (40) under self-paced presentation.

Two to four sessions were conducted per week per participant. Four groups of 4 to 7 old adults were trained in Fall 1984, Spring and Fall

¹ For the reader interested in the overall training program, a description of the rationale of the multicomponent model dealing with skilled digit memory can be found in Kliegl, Smith, Heckhausen, and Baltes (1987).

1985, and Spring 1986, respectively. The group of young adults participated in the Fall of 1985. Five tutors conducted the sessions. Because the present study was part of a larger research program investigating an expertise involving digit memory (Kliegl, Smith, & Baltes, 1986), some of the procedural developments described in the following sections were defined by considerations involving later performance on digit memory tasks rather than on the serial word recall task alone, which is the present focus. In our view, this fact reduces achievement of maximum possible performance on the word recall task but does not detract from the objective of the present study.

Procedure for pretest and posttest assessment. Prior to training, the following baseline measures were taken: (a) HAWIE intelligence test (the German equivalent of WAIS), (b) computer-administered auditory and visual digit span, and (c) serial word recall of two lists of 40 concrete nouns presented at a rate of 10 s and 4 s per word. Posttest measures of serial word recall consisted of performance on three lists of 40 words presented at a self-paced, a 10-s, and a 4-s rate respectively. For all word recall tests, participants were asked to recall as many words as possible and to write them on a sheet of paper in the order in which they had been presented. There was no specific time limit for completion of responses.

Procedure of training sessions. Each training session contained a variety of practice tasks and instruction periods. Specifically, there were three types of training activities: (a) acquisition of the Method of Loci for serial recall of concrete nouns, (b) acquisition of 100 historical events and their associated dates, and (c) transfer of the use of the Method of Loci from recall of random sequences of concrete nouns to historical events. The first and second activities were practiced in the same experimental sessions. The third task was introduced after participants had reached a stable level of performance in serial recall of lists of 40 nouns. All training sessions were self-paced; participant-controlled timing between items was recorded.

The first session of the training program was devoted to a sightseeing tour visiting the 40 selected West Berlin locations in a prearranged order. Then, at home, participants memorized these landmarks and the order in which they were visited on this trip. In the next session, they were tested to perfect forward and backward recall of this ordered map.

Participants then practiced using the Method of Loci to remember lists of concrete nouns (randomly selected from the pool of 1,560 nouns). Instruction in the Method of Loci followed the recommendations of Bower (1970). Participants were told to visit the Berlin locations mentally, always in the same fixed sequence, and to generate funny, dynamic, or interactive mental images or stories combining the to-be-remembered words and the locations. Participants verbalized their images; the tutor would occasionally offer suggestions for improvement. Records of each list consisted of the specific items presented (i.e., images to be constructed) and the number of items recalled in the correct position.

Starting with 5 landmarks, participants were introduced to longer lists in increments of 5 or 10 landmarks as their skill developed. At least one perfect recall of a list (up to 30 items) had to occur before a longer list was introduced. For list lengths of 30 to 40 items, the criterion for progression was 80% correct serial recall.

A secondary activity of these early training sessions was the acquisition of historical knowledge. Although this knowledge was a prerequisite for skilled digit memory (a task not covered in this report), it was also relevant to practicing the Method of Loci (see below). Participants were shown titles of historical events and had to enter the corresponding year on a numeric keypad. The computer provided immediate feedback about accuracy.

The final training activity involved transfer of the Method of Loci to remember lists of historical events. The training and test procedure for the historical event task was analogous to that used for the concrete noun lists, the only difference being that, rather than concrete nouns, titles of historical events (already familiar from the acquisition of histor-

Table 1
Means and Standard Deviations of Serial Word Recall Under Various Presentation Rates Prior to and Following Instruction in the Method of Loci

Group	Pretest		Posttest		
	10 s	4 s	Self-paced	10 s	4 s
Young					
<i>M</i>	4.8	2.5	39.8	37.8	24.8
<i>SD</i>	3.3	0.6	0.5	2.1	9.7
Old					
<i>M</i>	3.1	1.6	32.4	23.0	5.7
<i>SD</i>	1.9	1.4	6.0	9.9	5.4

Note. *N* young = 4. *N* old = 20. Maximum score was 40. Only words that were recalled in the original serial position were scored as correct answers. Average encoding rate in self-paced format at posttest was 24 s for both groups.

ical knowledge) were presented in random sequences. Participants were instructed to form images linking the historical events and the Berlin landmarks.

An effort was made to provide a relaxed atmosphere during the training sessions. Participants were not rushed and could take breaks for coffee and cookies as well as for conversation with the tutor.

Results

In the following sections, results will be presented both for pretest and posttest assessment of serial word recall using the Method of Loci and for progress through the training program. The range of performance plasticity is reflected in the gains that are achieved from pretest to posttest performance. Comparative "limits" to plasticity are reflected in the age differences that are exposed by the training program.

Serial Word Recall

Results from pretest and posttest assessments of serial word recall are displayed in Table 1. Only answers that were recalled in correct absolute position within the list were scored as correct. There were no statistically significant group differences in the word recall measure at pretest: An Age (2) \times Presentation Rate (2; 10 s vs. 4 s) analysis of variance (ANOVA), with repeated measures on the second factor, revealed only a significant effect of presentation rate, $F(1, 22) = 7.9, p < .02$.²

Training produced large significant improvements in serial word recall in both groups. Under the test condition of least

² It is possible that this nonsignificance reflects a lack of sensitivity of measurement. When these data were rescored for recall disregarding order, significant differences were revealed that were due to age, $F(1, 22) = 20.5, p < .01$, and presentation rate, $F(1, 22) = 6.6, p < .02$. Consistent with previous research, young adults recalled more words ($M = 20.7$ and 18.0 , in the 10-s and 4-s conditions, respectively) compared with the old adults ($M = 12.3$ and 8.0 , respectively). There were no significant interactions. This scoring procedure, however, eliminates an important component identified for training in the present study and prominent in the test instructions, namely, memory for order. Our pre-post comparisons, therefore, focus on serial word recall.

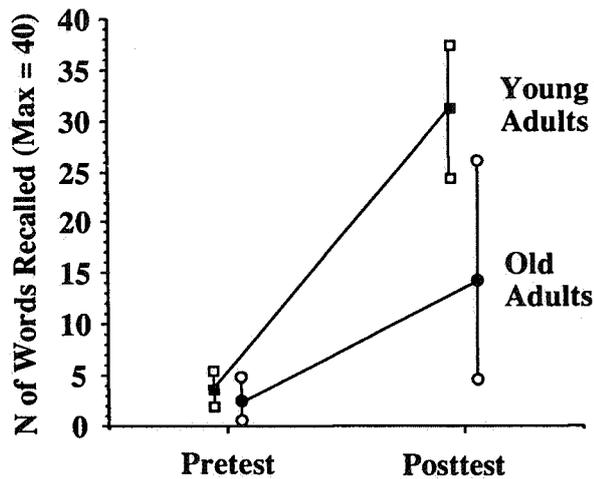


Figure 1. Age \times Time of Assessment interaction for serial recall of words in Experiment 1 (combining performance at 10-s and 4-s rates), showing the magnification of age and individual differences at posttest. (Bar lines indicate range of scores.)

difficulty (i.e., with self-paced presentation), the 4 young adults performed at or near ceiling (maximum of 40 words correct), whereas the old adults ranged between 19 and 40 words recalled ($M = 32.4$). The age groups did not differ in mean encoding latency in the self-paced test: Young adults took 24.4 s and old adults took 24.5 s per word on average.

Training also exposed age differences in serial word recall that were not visible in the baseline cognitive measures. First, as can be seen in Figure 1, performance increases were larger among the young adults than among the old adults. An Age (2) \times Time of Assessment (2) \times Presentation Rate (2) ANOVA, with repeated measures on the second and third factor, that was based on performance on the 10-s and 4-s word recall tests confirmed this significant Time of Assessment \times Age interaction, $F(1, 22) = 23.4, p < .01$. As expected, there were significant main effects resulting from age, $F(1, 22) = 24.1, p < .01$, time of assessment, $F(1, 22) = 149.8, p < .01$, and presentation rate, $F(1, 22) = 47.5, p < .01$. No other interactions reached significance. The bars in Figure 1 indicate the full range of scores. With the exception of 1 old and 1 young adult, there was perfect separation between the age groups at posttest. Considering the fact that the old adults had more practice and had spent more time in the training phase (see the following section), this finding is especially important because it is a conservative estimate of the age difference in developmental reserve capacity.

Second, and contrary to data gathered at pretest, cross-sectional age differences were revealed in all test conditions at posttest, that is, following training in a mnemonic skill to reveal differences in reserve capacity. An Age (2) \times Presentation Rate (3) ANOVA, with repeated measures on the second factor, resulted in a significant effect of age, $F(1, 22) = 25.7, p < .01$, a significant effect of presentation rate, $F(2, 44) = 37.6, p < .01$, and a marginally significant interaction between these two factors, $F(2, 44) = 2.9, p < .07$. As can be seen in Table 1, this interaction was possibly caused by a ceiling for young adults in the self-paced condition.

Progress Through the Training Program

Because rate of progression through the program was individually determined, participants had experienced different amounts of practice (in terms of number and type of lists and number of images to be constructed) before the scheduled posttests in serial word recall. Furthermore, because practice had been self-paced, participants had also spent different amounts of time constructing each image. Table 2 summarizes this practice history (combining data covering training with words and historical events) for both age groups prior to the posttest sessions.

Old adults had an average of about twice as many training sessions as young adults, recalled 31 compared with 19 lists, formed 699 compared with 436 mental images, and spent more than twice as much time in constructing these mental images. The last parameter was determined by summing encoding times (i.e., presentation times) of individual items. Of course, all these age differences were highly significant. Old adults, then, spent considerably more time practicing the task.

Discussion

The results of Experiment 1 were in general agreement with our theoretical expectations. First, old and young adults exhibited a large range of developmental reserve capacity (cognitive plasticity). Old adults, on the average, were able to repeat 32 of 40 words that they had seen only once at a self-paced presentation rate. Such a high level of memory performance has not been reported in past research. Young adults, however, demonstrated an even larger developmental reserve capacity. In line with previous research using mnemonic instructions, young adults produced remarkably high levels of performance, even with a 4-s presentation rate.

Second, there was a clear emergence of cross-sectional age differences following training; age differences had not been observed at pretest. The posttest performance distributions of the two age groups under speed conditions overlapped only minimally in contrast with the considerable overlap at pretest (see Figure 1). This pattern of results gains in significance because it is based on a training schedule that favored old adults in terms of amount of practice.

One finding, however, is not in full agreement with our theoretical expectations. We had expected to observe larger age

Table 2
Training History: Means and Standard Deviations of Image Constructions Associated With Method of Loci Prior to Posttest on Serial Word Recall

Group	Training sessions (N)	Lists (N)	Images (N)	Total time spent encoding (min)
Young				
M	8.8	19.3	436.3	188.0
SD	1.0	1.5	59.4	9.7
Old				
M	17.0	30.8	698.6	381.8
SD	4.5	6.9	190.6	127.7

differences the more difficult the task, that is, in this case, the faster the presentation rate. Counter to this expectation, age differences at posttest were not larger for the 4-s compared with the 10-s presentation rate. We had expected that reducing the time available for encoding would lead to a more pronounced decline in recall performance of the elderly compared with the young adults. The data, however, (tentative as they are, especially for the young adult group) suggest that both groups suffered equally from the cutback in presentation time.

There are three possible weaknesses in the present study. The first concerns the small sample of young adults ($N = 4$). The second is that the amount of practice was not held constant for the two age groups. The third is that the training schedule involved additional task components (e.g., learning of historical events) that may have distracted from an effective acquisition of the specific mnemonic skill (e.g., because of a lack of task focus or cognitive overload). Experiment 2 was designed to correct these weaknesses by studying a larger number of young adults and by improving the training program.

Experiment 2: Replication and Extension

The primary purpose of Experiment 2 was to replicate the results pertaining to the range and age-related limits in the range of developmental reserve capacity using a newly designed training program. Aside from increasing the sample of young adults, three additional design conditions were introduced. The first involved focusing training only on serial recall of word lists. The second concerned increasing the number of presentation rates. An unexpected result of Experiment 1 was the lack of an age differential effect that was due to different presentation rates at posttest. We had expected that the performance difference between young and old adults would be enlarged as a function of faster presentation rates (i.e., 4 s vs. 10 s). To investigate this more precisely, six instead of two different presentation rates were used at pretest and posttest in Experiment 2. Specifically, lists of words were presented at fixed rates of 20 s, 15 s, 10 s, 5 s, 3 s, and 1 s per word.

The third modification concerned achieving a better match between participants' performance potential and the training schedule. The training procedure of Experiment 1 was aimed primarily at assisting the elderly. Young adults performed close to ceiling throughout most of the training sessions and, thus, reached the stated criterion rather early. Therefore, it is likely that they did not profit as much from the training as did the elderly. In Experiment 2, the training program was modified to give young and old adults comparable opportunities to actualize their developmental reserve capacity. Thus, all participants received the same number of test and practice lists for recall. When a criterion of 50% accuracy was reached on training lists, presentation rate was increased so that all participants practiced at an individually determined level of medium difficulty. This adaptive format of training was assumed to maximize the demands on reserve capacity because it avoids both cognitive overload and nonchallenging task conditions.³

Method

Participants

The final study sample consisted of 18 (7 women and 11 men) young adults ($M = 23.9$ years; range = 19 to 29 years) and 19 (11 female and

8 men) old adults ($M = 71.7$ years; range = 65 to 80 years). Of an initial sample of 42, 3 young and 2 old adults had dropped out for lack of interest (2 young) and for health-related reasons (1 young and 2 old). All of the participants were volunteers and physically able to come to the laboratory. Reported subjective health was rated as above average (for the young adults, $M = 4.2$, $SD = .7$; for the old adults, $M = 3.9$, $SD = .7$, on a 5-point self-report scale). This difference was not statistically significant.

The young adults were university students specializing in various fields of study. They all had completed 13 years of school education plus 1 to 6 years of university studies. Their average Hamburg-Wechsler IQ (HAWIE), assessed in the first session, was 117.5 ($SD = 6.8$).

On average, the old adults had completed 12.5 years of education ($SD = 3.0$). Four months prior to this experiment, the old adults had participated in a cognitive training study involving items defining the fluid ability factor figural relations (Baltes, Sowarka, & Kliegl, in press). On the pretest measures of intelligence given in that study, they had scored above the 60th percentile of the sample. Their average HAWIE IQ prior to the present experiment was 125.4 ($SD = 8.0$).

In terms of IQ, the present samples were comparable with those in Experiment 1. Although old adults had a significantly higher age-corrected IQ than young adults, $t(35) = 3.2$, $p < .01$, the age groups were equal in verbal raw scores ($M = 63.3$, $SD = 3.8$, for young, and $M = 63.1$, $SD = 6.1$, for old adults). Moreover, the older adults scored significantly lower than the young on performance scores (for the young, $M = 62.2$, $SD = 6.4$; for the old, $M = 54.0$, $SD = 7.1$), $t(35) = 3.7$, $p < .01$. Thus, the older sample's intelligence profile was consistent with the usual pattern of healthy aging.

Apparatus and Materials

Memory tests and training were administered on APPLE-IIe computers. For pretest and posttest, two sets of six lists with 30 concrete nouns each were used for assessment of serial word recall. Presentation order of sets was counterbalanced across test occasions. Lists within sets and words within lists were administered in an invariant order.

For training, two random orders of a single set of 360 concrete nouns were constructed. Participants were exposed to both forms in a counterbalanced order (i.e., 24 lists with 30 concrete nouns each). Thirty Berlin landmarks (well-known buildings and places) constituted the mental map for the Method of Loci mnemonic.

Overview of Test and Training Program

The test and training program consisted of a total of 20 sessions per participant. In the first session, the Hamburg-Wechsler Intelligence Scales (HAWIE), a test of mental rotation, the vividness of visual imagery questionnaire (VVIQ), and a short biographical questionnaire were administered. Serial word recall was tested in the 2nd (pretest) and 17th (posttest) sessions at six presentation rates. Initial instruction in the Method of Loci occurred in the 5th and 6th sessions. Further training in the use of the Method of Loci with an adaptive format (i.e., dynamically adjusted at an individual level) was scheduled in Sessions 8, 9, 11, 12, 14, and 15. The remaining sessions involved the collection of additional measures not covered in the present report (see Footnote 3). Five research assistants conducted the experimental sessions.

³ Experiment 2 also included additional memory tests prior to and after the training, charted the acquisition of mnemonic skill at four occasions during the training, and assessed the impact of mnemonic training on memory-related and general control beliefs. Results pertaining to these questions will be reported in separate articles currently in preparation.

Instruction in the Method of Loci

Instruction in the Method of Loci was similar to Experiment 1 except that participants were not physically taken on a "sightseeing" trip. In the initial instruction sessions (5 and 6), participants "overlearned" the 30 locations until they were able to recite them, in order, within 90 s without error. They also received instruction in imagery techniques and practiced using the technique to recall three 10-word lists. The first instruction session was held in a group setting involving 3 to 5 persons; the second session was individualized.

Procedure for Serial Word Recall (Pretest and Posttest)

Prior to and after adaptive training in the Method of Loci, participants recalled six lists of 30 concrete nouns each. Presentation rate for the lists was in this order: 20 s, 15 s, 10 s, 5 s, 3 s, and 1 s per word, respectively. Words were presented in the center of the monitor in 40-column Apple-font, in green letters on a dark background. After presentation, participants wrote down the words on a sheet of paper containing 30 numbered lines. They were instructed to attempt to recall both the word and, as close as possible, the absolute position of the word in the list. Recall time per list was limited to a maximum of 10 min. The tests were conducted in group settings of 3 to 5 persons.

Procedure for Adaptive Training of Method of Loci

In each of six adaptive training sessions, four lists of 30 words were presented at a rate that depended on the participant's prior serial recall performance. Starting with a 20-s rate, presentation rate was increased when 15 or more of the 30 words were recalled in correct serial position on two consecutive lists. The 50% criterion was adopted to ensure progress for all participants within the 20-session program and to avoid floor and ceiling effects for both age groups. To provide for similarity with pretest and posttest assessment, the incremental rates used during adaptive training were as follows: 20 s, 15 s, 10 s, 5 s, 3 s, and 1 s per word.

For all practice lists, encoding occurred without explicitly presented cues (landmarks); during recall the computer prompted with a landmark, and the participant provided the appropriate word if possible. The tutor entered this response on the computer, and latencies were collected. A response had to be given within 20 s of the prompt.

After each list, participants received feedback about their performance. The computer displayed the landmarks, the to-be-remembered nouns, and the participant's responses. Between lists, the tutor probed participants about their images or thoughts, offered evaluative comments, and reminded them of the mnemonic principles.

Results and Discussion

Pretest and Posttest Assessment of Serial Word Recall

Means and standard deviations of serial word recall are shown in Table 3 broken down by age, time of assessment, and presentation rate. These data were analyzed using an Age (2) \times Time of Assessment (2) \times Presentation Rate (6) ANOVA, with repeated measures on the second and third factors.⁴

Overall, young adults recalled more words than old adults, $F(1, 35) = 73.0, p < .01$. There was a significant effect of time of assessment, $F(1, 35) = 263.9, p < .01$, and, most important, the interaction between age and time of assessment was also significant, $F(1, 35) = 29.6, p < 0.1$. The corresponding pattern of means is displayed in Figure 2. Each data point in Figure 2 represents the mean of six lists with different presentation rates. Unlike Experiment 1, there was a significant difference between

Table 3
Means and Standard Deviations of Serial Word Recall

Presentation rate	Young adults		Old adults	
	Pretest	Posttest	Pretest	Posttest
20 s				
<i>M</i>	7.6	28.7	4.0	17.8
<i>SD</i>	6.2	1.9	3.9	8.7
15 s				
<i>M</i>	8.6	27.2	3.5	15.4
<i>SD</i>	5.6	3.3	3.2	8.2
10 s				
<i>M</i>	7.3	23.4	3.0	12.1
<i>SD</i>	7.8	5.6	2.0	6.1
5 s				
<i>M</i>	4.3	19.3	2.6	7.1
<i>SD</i>	2.4	6.2	2.0	3.1
3 s				
<i>M</i>	4.8	16.4	1.9	4.1
<i>SD</i>	3.7	8.1	1.2	3.2
1 s				
<i>M</i>	2.0	4.4	1.3	2.3
<i>SD</i>	1.6	2.4	1.6	1.6
Overall				
<i>M</i>	5.8	19.9	2.7	9.8
<i>SD</i>	3.1	3.3	1.6	3.9

Note. Maximum score was 30. Only words that were recalled in the original serial position were scored as correct answers. Presentation rate was confounded with list order (i.e., lists were always administered in the order indicated in the table).

age groups at pretest, $t(35) = 3.8, p < .01$. This finding most likely reflects the statistical benefit of having a larger number of young adults in the study. As in Experiment 1 (see Figure 1), young adults improved more than old adults. Thus, the magnification of age difference was replicated with a new training program.⁵

The repeated measures ANOVA indicated also that all effects involving presentation rate were significant: for presentation rate, $F(5, 175) = 117.9, p < .01$, for the Age \times Presentation Rate interaction, $F(5, 175) = 9.3, p < .01$, for the Time of Assessment \times Presentation Rate interaction, $F(5, 175) = 31.9, p < .01$, and for the Age \times Time of Assessment \times Presentation Rate interaction, $F(5, 175) = 2.4, p < .05$. Figure 3 is a graphic rendition of the three-way interaction.

Interactions involving age and presentation rate derived their

⁴ Presentation rate was confounded with order of lists within the experimental sessions (i.e., 20-s rate was first and 1-s rate was last). Thus, proactive interference may have exacerbated differences between presentation rates. Although interpretations must be qualified in this respect, the dominant source of variance in all likelihood was the encoding times.

⁵ The critical age by time of assessment interaction was also significant (a) with a lenient scoring criterion (disregarding order information) and (b) in a repeated measures analysis of variance on ranks (i.e., assuming only ordinal level of measurement for the dependent variable). The second analysis of variance is important, because with significant age differences at pretest the interaction could have been due to a scaling artifact (e.g., Embretson, 1987).

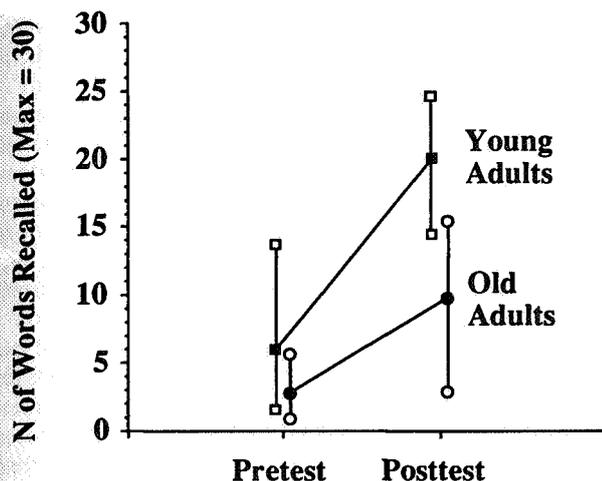


Figure 2. Age × Time of Assessment interaction for serial recall of words in Experiment 2 (averaging across 20-s, 15-s, 10-s, 5-s, 3-s, and 1-s presentation rates). (Bar lines indicate range of scores.)

source of variance mainly from the small group difference at the 1-s presentation rate: Training had very little impact on performance for the fastest presentation rate. When the 1-s rate was left out of the ANOVA, F ratios for the interaction between age and presentation rate and for the triple interaction dropped below 1.0. Thus, both groups benefited equally for rates between 20 s and 3 s per word. Again, this replicates the results of Experiment 1, where 10-s and 4-s rates led to equivalent group differences (see Figure 2). Across both age groups, however, training gains were larger when more time per word was available at encoding.

Progress Through the Training Program

In contrast to Experiment 1, all participants recalled the same number (i.e., 24) of 30-word lists (although at individually adjusted presentation rates) in the six training sessions. Presentation rate was increased whenever a participant had recalled 50% or more of the words on two consecutive lists. Thus, the 1-s rate could have been reached after a minimum of 10 of the total 24 lists. Progress through the training program can be visualized by plotting the percentage of participants who passed the thresholds (i.e., two times with 50% correct) of the various presentation rates.

All young and old participants reached criterion at the 20-s and the 15-s rate; all but 1 old adult also passed the 10-s threshold. Age differences became clearly visible at the 5-s and 3-s rates: All except 1 young adult passed the 5-s threshold; in contrast, only 58% ($n = 11$) of the old adults did so. Whereas 72% of the young adults succeeded at the 3-s rate, none of the old adults did so. Finally, only 1 young person also met the 1-s criterion.

With presentation rates adjusted according to individual ability, recall performance should eventually be equal in both age groups. An Age (2) × Training Session (6) ANOVA, with repeated measures on the second factor, yielded significant effects resulting from age, $F(1, 35) = 8.1, p < .01$, training session, $F(5,$

175) = 86.1, $p < .01$, and a significant Age × Training Session interaction, $F(5, 175) = 6.5, p < .01$. The interaction is displayed in the left panel of Figure 4.

Of course, the age-differential decline in recall performance across training sessions reflects the fact that young adults worked with faster presentation rates in later sessions (see the right panel of Figure 4). Young adults recalled more words than old adults in the first three sessions, but there were no significant differences between groups in the final three sessions (post hoc t tests with α adjusted to $.05/6 = .008$). The second half of the training program, then, was of comparable difficulty for both age groups, with most young adults working at a 1-s rate and most old adults, at a 5-s or 3-s rate.

General Discussion

The two main results of the experiments reported here were the demonstration of substantial cognitive plasticity or reserve capacity and the magnification of age differences when limits of reserve capacity are approximated by means of a training-based strategy of testing-the-limits. The results are consistent with our theoretical concept of developmental reserve capacity: Although both young and old adults have cognitive capacity available for performance gains, "estimated" limits of developmental reserve are lower for old than for young adults.

Testing-the-limits in combination with a cognitive training program entails a high degree of control over the cognitive processes involved. We were able, therefore, to specify, more precisely than is the case with research using single-session assessments of age differences, the "type" of capacity that has been taxed by conditions of testing-the-limits. In the present study, for example, processes associated with the continuous and fast

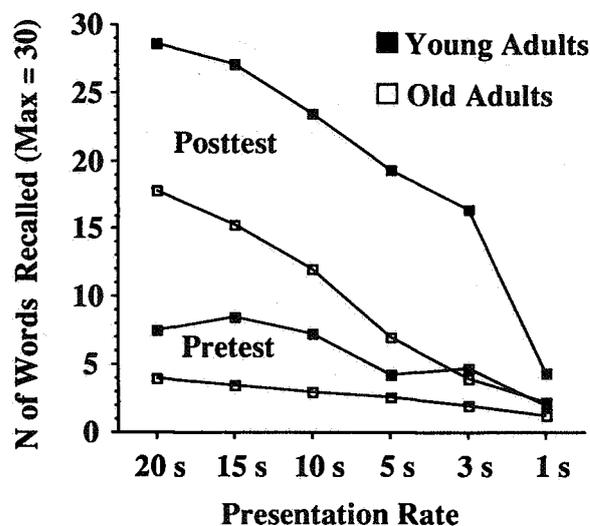


Figure 3. Age × Time of Assessment × Presentation Rate interaction for serial recall of words in Experiment 2. (The three-way interaction was due to the 1-s presentation rate only. Differences between groups did not vary as a function of presentation rate at pretest and at posttest for the other rates. Training effects covaried positively with presentation rate—that is, there was larger improvement with slow than with fast rates in both groups.)

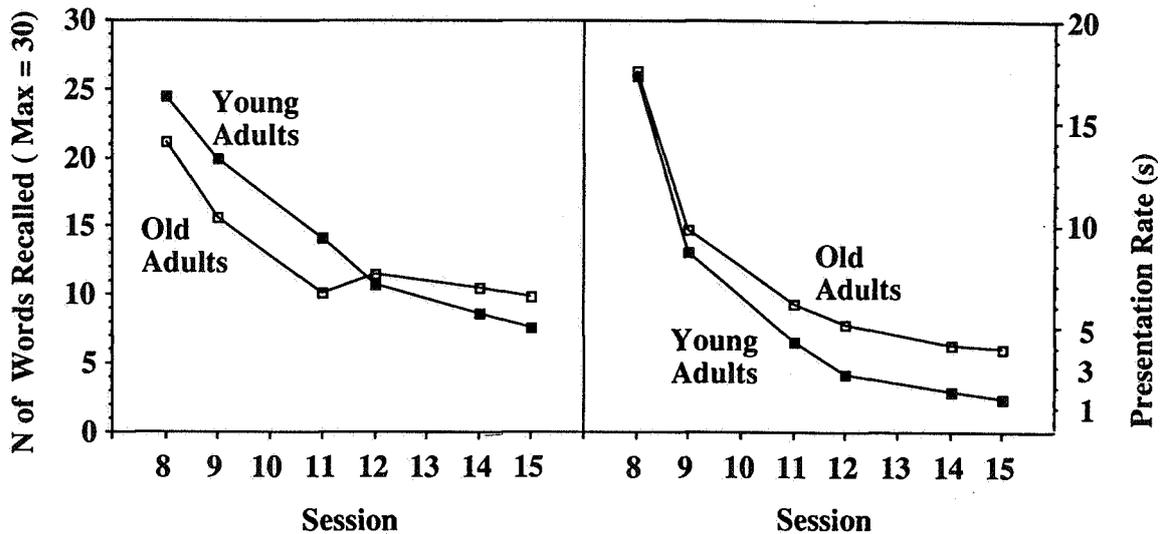


Figure 4. Number of words recalled (left panel) and average presentation rate (right panel) as a function of age group and adaptive training session. (Presentation rates were individually determined and depended on the recall level in preceding lists. Only the presentation rates specified in the ordinate were used. Each data point represents the average of four lists.)

generation of images or thoughts were indicated as a possible source of differences between young and old adults.

Information about approximate limits of developmental reserve capacity is expected to become available not only following extensive training, but also under more demanding performance conditions. The less time that is available for encoding operations, the larger should be the difference between groups. This interaction was also thought to increase as a function of training. Counter to expectations, however, old and young adults showed comparable training gains across a wide range of presentation rates (i.e., 20 s to 3 s per word in Experiment 2). Moreover, age differences were actually smallest at the fastest (1-s) presentation rate. Most likely, we did not obtain training-related age differences at the 1-s rate because neither young nor old adults could effectively deploy their mnemonic skill at this speed. This could have occurred for two reasons. First, "true" limits were not reached by the current training program. With more sessions, magnification may also have occurred at the 1-s rate. Second, the 1-s rate may be beyond "absolute" cognitive processing limits. From both accounts, it follows that increases in task difficulty will lead to a magnification of age differences only if the level of difficulty is not beyond limits associated with the necessary cognitive processes. The results also leave open the possibility that old adults following extensive training may be able to reach levels of performance comparable with those of young adults.

Are the results of such training-related magnification of age differences compatible with current accounts of cognitive aging? One general proposal is that age differences covary with task difficulty and with the degree of self-initiated processing required to perform effectively (i.e., complexity); the more difficult the task or the more self-initiated processing required, the larger the differences between young and old adults (Baltes, Dittmann-Kohli, & Dixon, 1984; Craik, 1983; Hasher & Zacks,

1979; Salthouse, 1985). Our results are consistent with this account on the assumption that instructing participants in the mnemonic technique increases their degree of self-initiated processing (e.g., Craik, 1983). Preliminary evidence for a training-related increase in self-initiated processing was indicated in a training-related increase in self-paced encoding rates: From ongoing research we know that untrained adults spend an average of 5 s per word at self-paced encoding, whereas trained adults use an average of 24 s per word (cf. Experiment 1).

As a corollary to this interpretation, it should be mentioned that, unlike other research, the hypothetical training-related increase in processing complexity was not accompanied by an increase in "objective" task difficulty. To the contrary, instruction in the mnemonic device made the task easier in terms of the number of words that participants were able to recall. In this sense, training research involving the theory-guided engineering of a cognitive skill may be one way to resolve the often-lamented confound of task difficulty and complexity (e.g., Cella, Poon, & Williams, 1980; Salthouse, 1985; Welford, 1958).

Although we argue that the results that we obtained are clear in demonstrating plasticity and suggestive of age-related limits of plasticity, there are several limitations concerning the interpretation. First, we do not know the temporal course of the emergence and magnification of age differences nor do we know the final "limits." For example, magnification of age differences may have been produced immediately following initial instruction in the mnemonic technique (Sessions 5 and 6), and there could have been very little or no magnification occurring thereafter. Alternatively, age differences could have emerged gradually and continually in the course of the training program.

Second, it would be desirable to know whether further continuation of the training program would permit old adults to reach levels of performance comparable with young adults. Ideally, one would want to compare age groups after they reached an

asymptotic performance level to rule out explanations in terms of differences in rate of acquisition rather than final level.

Third, we did not examine whether there are other tests of memory that, if given before training, would have yielded similar predictive validity and magnification of age differences as was true for our posttest assessment. Variations in memory-task formats (i.e., free recall, cued recall, recognition) could be used to check on the dependence of age-differential effects of training on the degree of environmental support (i.e., the presence or absence of encoding and retrieval cues).

Fourth, the interpretation of age differences is limited by the fact that they also represent cohort and other subject selection effects (Baltes, Reese, & Nesselroade, 1977). Only additional research with cohort-sequential designs and more broadly based samples can examine the degree to which the present findings are generalizable, for example, to other historical cohorts and sample characteristics.

Despite these limitations, the two chief purposes of the study were achieved. First, we wanted to demonstrate in the context of memory functioning the fact of sizeable reserve capacity (plasticity) in young and old adults. Second, we were interested in moving cognitive training research beyond the sheer demonstration of plasticity to the study of limits of reserve capacity (Baltes, 1987; Kliegl & Baltes, 1987). The fact that age differences were magnified as a consequence of training and at quasi-expert levels of performance, suggests that effects of aging may be more clearly identifiable at performance conditions near the upper limit of reserve than at baseline performance conditions. These findings on "limits" of reserve capacity also suggest that the aging effects that we obtained are not easily explainable by experience- or disuse-based accounts (i.e., by accounts that assume a practice deficit on the part of old adults). Rather, we submit that the effects are more likely the results of neurophysiological limits of the aging brain.

References

- Anschutz, L., Camp, C. J., Markley, R. P., & Kramer, J. J. (1985). Maintenance and generalization of mnemonics for grocery shopping by older adults. *Experimental Aging Research*, *11*, 157-160.
- Anschutz, L., Camp, C. J., Markley, R. P., & Kramer, J. J. (1987). Remembering mnemonics: A three-year follow-up on the effects of mnemonics training in elderly adults. *Experimental Aging Research*, *13*, 141-143.
- Baltes, M. M., & Kindermann, T. (1985). Die Bedeutung der Plastizität für die klinische Beurteilung des Leistungsverhaltens im Alter [The relevance of plasticity for clinical evaluation for achievement behavior]. In D. Bente, H. Coper, & F. Kanowski (Eds.), *Hirnorganische Psychosyndrome im Alter: Methoden zur Objektivierung pharmakotherapeutischer Wirkung* [Impairment of brain functions in old age: Methods for Objectification of pharmacotherapeutic effects] (Vol. 20, pp. 171-184). Berlin: Springer.
- Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, *23*, 611-626.
- Baltes, P. B., Dittmann-Kohli, F., & Dixon, R. A. (1984). New perspectives on the development of intelligence in adulthood: Toward a dual-process conception and a model of selective optimization with compensation. In P. B. Baltes & O. G. Brim, Jr. (Eds.), *Life-span development and behavior* (Vol. 6, pp. 34-76). New York: Academic Press.
- Baltes, P. B., Dittmann-Kohli, F., & Kliegl, R. (1986). Reserve capacity of the elderly in aging-sensitive tests of fluid intelligence: Replication and extension. *Psychology and Aging*, *1*, 172-177.
- Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods*. Monterey, CA: Brooks/Cole.
- Baltes, P. B., Sowarka, D., & Kliegl, R. (in press). Cognitive training research on fluid intelligence in old age: What can older adults achieve by themselves? *Psychology and Aging*.
- Baschek, I.-L., Bredenkamp, J., Oehrle, B., & Wippich, W. (1977). Bestimmung der Bildhaftigkeit (I), Konkretheit (C), und der Bedeutungshaftigkeit (m') von 800 Substantiven [Determining imagery, concreteness, and meaningfulness values of 800 nouns]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *32*, 3-20.
- Bower, G. H. (1970). Analysis of a mnemonic device. *American Scientist*, *58*, 496-510.
- Brown, A. L. (1982). Learning and development: The problem of compatibility, access, and induction. *Human Development*, *25*, 89-115.
- Cerella, J., Poon, L. W., & Williams, D. M. (1980). Age and the complexity hypothesis. In L. W. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 332-340). Washington, DC: American Psychological Association.
- Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London*, *B 302*, 341-359.
- Denney, N. W. (1984). A model of cognitive development across the life span. *Developmental Review*, *4*, 171-191.
- Embretson, S. E. (1978). Improving the measurement of spatial aptitude by dynamic testing. *Intelligence*, *11*, 333-358.
- Engeström, Y. (1986). The zone of proximal development as a basic category of educational psychology. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, *8*, 23-42.
- Gollin, E. S. (1981). Development and plasticity. In E. S. Gollin (Ed.), *Developmental plasticity: Behavioral and biological aspects of variations in development* (pp. 231-251). New York: Academic Press.
- Guthke, J. (1982). The learning test concept—An alternative to the traditional static intelligence test. *The German Journal of Psychology*, *6*, 306-324.
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*, 356-388.
- Heineken, E., & Gekeler, C. (1985). Gedächtnisleistung und allgemeine Leistungsbereitschaft im Alter [Memory performance and general willingness to achieve in old age]. *Zeitschrift für Gerontologie*, *18*, 292-297.
- Kliegl, R., & Baltes, P. B. (1987). Theory-guided analysis of mechanisms of development and aging through testing-the-limits and research on expertise. In C. Schooler & K. W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 95-119). Norwood, NJ: Ablex.
- Kliegl, R., Smith, J., & Baltes, P. B. (1986). Testing-the-limits, expertise, and memory in adulthood and old age. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 395-407). Amsterdam: North Holland.
- Kliegl, R., Smith, J., Heckhausen, J., & Baltes, P. B. (1987). Mnemonic training for the acquisition of skilled digit memory. *Cognition and Instruction*, *4*, 203-223.
- Lerner, R. M. (1984). *On the nature of human plasticity*. New York: Cambridge University Press.
- Pavio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology (Monograph and Supplement)*, *76* (Part 2), 1-25.
- Robertson-Tschabo, E. A., Hausmann, C. P., & Arenberg, D. (1976). A classical mnemonic for older learners: A trip that works! *Educational Gerontology*, *1*, 215-226.

- Rose, T. L., & Yesavage, J. A. (1983). Differential effects of a list-learning mnemonic in three age groups. *Gerontology, 29*, 293-298.
- Salthouse, T. A. (1985). *A theory of cognitive aging*. Amsterdam: North Holland.
- Schmidt, L. R. (1971). Testing-the-limits im Leistungsverhalten: Möglichkeiten und Grenzen [Testing-the-limits in achievement behavior: Perspectives and limitations]. In E. Duhm (Ed.), *Praxis der klinischen Psychologie* (Vol. 2, pp. 9-29). Göttingen, Federal Republic of Germany: Hogrefe.
- Volkman, L. (1929). Ars memorativa. *Jahrbuch der Kunsthistorischen Sammlungen in Wien, 3*, 111-200.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Welford, A. T. (1958). *Ageing and human skill*. London: Methuen Press.
- Wiedl, K. H. (1984). Lerntests: Nur Forschungsmittel und Forschungsgegenstand? [Learning tests: Only a research method and a research topic?] *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 16*, 245-281.
- Willis, S. L. (1985). Towards an educational psychology of the adult learner. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 818-847). New York: Van Nostrand Reinhold.
- Yates, F. A. (1966). *The art of memory*. London: Routledge & Kegan Paul.
- Yesavage, J. A., & Rose, T. L. (1984). Semantic elaboration and the method of loci. A new trip for older learners. *Experimental Aging Research, 10*, 155-159.