

Further Testing of Limits of Cognitive Plasticity: Negative Age Differences in a Mnemonic Skill Are Robust

Paul B. Baltes and Reinhold Kliegl

Max Planck Institute for Human Development and Education
Berlin, Federal Republic of Germany

Earlier testing-the-limits research on age differences in cognitive plasticity of a memory skill was extended by 18 additional assessment and training sessions to explore whether older adults were able to catch up with additional practice and improved training conditions. The focus was on the method of loci, which requires mental imagination to encode and retrieve lists of words from memory in serial order. Of the original 37 subjects, 35 (16 young, ranging from 20 to 30 years of age, and 19 older adults, ranging from 66 to 80 years of age) participated in the follow-up study. Older adults showed sizable performance deficits when compared with young adults and tested for limits of reserve capacity. The negative age difference was substantial, resistant to extensive practice, and applied to all subjects studied. The primary origin for this negative age difference may be a loss in the production and use of mental imagination for operations of the mind.

In past writings on cognitive development across the life span, we have argued that the assessment of changes is facilitated by a testing-the-limits methodology that is explicitly oriented toward the estimation of latent potential, or what we have termed *developmental reserve capacity* (Baltes, 1987; Kliegl & Baltes, 1987). Two recent articles published in this journal demonstrated the usefulness of the approach (Kliegl, Smith, & Baltes, 1989, 1990). The specific focus chosen for the investigation was a specific mnemonic strategy, the method of loci, and its use in serial recall of words (Bower, 1970). The key component of the method of loci is the forging of mental images or thoughts linking words to be remembered in order of appearance to an invariant series of mental landmarks. At recall, one mentally revisits the mental locations in order, retrieves the associated mental image or thought, and decodes from these mental images the words to be remembered.¹

In Kliegl et al. (1989, 1990), young and old adults participated in a total of 20 sessions aimed at the practice and assessment of the method of loci and its use in serial recall of words. On the one hand, it was shown that adults, including the older ones, were able to acquire the method of loci (see also Robertson-Tchabo, Hausman, & Arenberg, 1976; Yesavage & Rose, 1984). After relatively few training sessions, subjects were able

to reproduce in correct order more than 20 words of a long list (such as 30 words) after a single presentation, as long as the words were presented at about 20-s intervals during encoding. Without application of the method of loci, the usual performance in recalling lists of words in serial order is below 10 words.

On the other hand, however, the Kliegl et al. (1989, 1990) research suggested major negative age differences near limits of functioning. Despite definite evidence for reserve capacity, negative age differences in serial word recall were substantial and magnified after 20 sessions of practice. Furthermore, regression analyses showed that chronological age attracted more variance following testing-the-limits training than at pretest or in the early stages of acquisition.

This study was implemented because the data reported in Kliegl et al. (1989, 1990) were insufficient in two respects. First, there remained the question if after more extensive practice, would the age differences continue to persist? Second, there was also a question about the degree of effectiveness of the training program used. Specifically, the original practice program for the method of loci did not contain the option of backtracking when performance was below criterion and, therefore, was somewhat insensitive to individual differences in rates of learning. To answer these questions, the original study was thus extended by an additional 18 sessions.

This research was conducted in the context of the Expertise and Cognitive Aging Project codirected by Reinhold Kliegl and Paul B. Baltes. Although we cannot mention all of them, we would like to express our gratitude to some of the colleagues who made significant contributions to our thinking on this topic: Ulman Lindenberger, Ulrich Mayr, Timothy Salthouse, and Jacqui Smith. We also thank the study participants for their high level of commitment, Annette Rentz and Werner Scholtysik for research coordination, and Gregor Caregnato and Amoura Youssef for assistance in data collection.

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

¹ This article does not address the question of a componential analysis of performance in the method of loci. Specifically, we do not address the issue of whether, or to what degree, the generation of visual analogues is critical for successful retention of mental images or whether accounts in terms of integrative relations are sufficient (e.g., Marschark & Hunt, 1989). However, instructions strongly emphasized the generation of visual images depicting landmarks and words to be remembered in interactive relations. There is also a posteriori evidence from verbal interviews that participants focused on the generation of visual images (see also Lindenberger, 1991).

Method

More detailed information on subjects, procedure, and assessment is contained in Kliegl et al. (1989, 1990). This article presents specific information only concerning the extension features of the study.

Subjects

Thirty-five of the 37 original research subjects in the Kliegl et al. (1989) study continued into this extension experiment (Phase 2). All subjects were above average in education and intelligence (young IQ: $M = 118$, $SD = 6$; old IQ: $M = 125$, $SD = 8$) and were in good health, as rated with a criterion of self-reported health. Young and older subjects' raw scores did not differ on the German version of the Wechsler Vocabulary Test (HAWIE) or on their ratings of subjective health status. As expected, they did differ, however, on a digit symbol substitution test (an age-sensitive test of fluid-perceptual intelligence), with young subjects scoring significantly higher than the older subjects.

Design

The entire experimental schedule consisted of thirty-eight 1-hr sessions distributed over a period of 1 year and 4 months. The sessions can be divided into two major categories: training sessions and assessment sessions. In addition, there were two debriefing sessions (Sessions 20 and 38).

Material. Lists of 30 words were constructed from a pool of 2,100 concrete nouns. Words used for the assessment lists appeared only once in each study phase. Words used for the training sessions were repeated once in each study phase but were paired with a different location of the mental map.

Training conditions. Individual training in the method of loci was conducted in 17 of the 38 sessions (i.e., Sessions 5, 6, 8, 9, 11, 12, 14, 15, 21, 22, 23, 28, 29, 30, 33, 34, and 35). In each training session, scheduled at approximately 2 sessions per week, four lists of 30 words each were administered. Across all sessions of training and assessment, each subject performed 4,380 trials of trying to generate a mental image linking a location to a word to be remembered.

The training program consisted of three successive parts. The first part (Sessions 5 and 6) was devoted to explanation and instruction in the method of loci. During this part, words were presented at time intervals comfortable to each subject, that is, without time pressure. The second part of the training program involved nine sessions (Sessions 8, 9, 11, 12, 14, 15, 21, 22, and 23) and was oriented toward maintaining a criterion level of performance with shorter and shorter presentation times. In this part, subjects started with a presentation time of 20 s per word. Presentation time was decreased to 15 s, 10 s, 5 s, 3 s, and 1 s per word for the next list whenever 50% of the list (15 or more of the 30 words) had been recalled in correct serial position.

In the third part of the training program, the extension part of the study (Sessions 28, 29, 30, 33, 34, and 35), the presumed effectiveness of the program was enhanced in two ways. First, to achieve smaller steps in training and assessment, the number of presentation times was increased from 6 to 12 levels within the same general range (20 s to 0.84 s). Specifically, beginning with 15 s, subsequent times were decreased each by 25% up to the shortest time interval of 0.84 s per word. Second, the possibility of backtracking was introduced to arrange for better, that is, more individualized learning conditions. Subjects were moved to shorter presentation times when they passed a criterion (50% or 80% correct).² When performance was below criterion on two successive lists, a subject was moved back to the previous longer presentation time.

Performance assessment. For this article, assessment of progress in the method of loci at 9 occasions (7, 10, 13, 16, 17, 25, 26/27, 31/32, and

36/37) was the focus of our analyses. There were two assessment formats reflective of the change in rates of presentation toward a more differentiated scale from the first to the second phase of the study. At the first 6 occasions (Phase 1), six lists of 30 words were administered with presentation times of 20 s, 15 s, 10 s, 5 s, 3 s, and 1 s per word. At the final three occasions (Phase 2), assessment extended over two sessions. In the first assessment session, four lists were administered, with presentation times of 20 s, 15 s, 11.84 s, and 8.44 s; in the second assessment session, eight lists, with presentation times of 6.33 s, 4.75 s, 3.56 s, 2.67 s, 2.00 s, 1.5 s, 1.14 s, and 0.84 s, were administered in this order. To achieve comparability with the Phase 1 assessment format, the mean of the 11.84-s and 8.44-s conditions was taken as an estimate for the original 10-s condition; similarly, means for 6.33 s and 4.75 s, and means for 3.56 s and 2.67 s, were used as estimates of the original 5-s and 3-s presentation times, respectively.

For the four shortest presentation times, it was unclear whether the older adults still used the mnemonic. Therefore, only scores based on the five presentation times of 20 s, 15 s, 10 s, 5 s, and 3 s were used for the analyses. A maximum of 10 min was allowed for recalling the 30 words of a list. Spearman-Brown split-half reliabilities of the aggregated scores were greater than .90 within each age group.

Results

The data were analyzed with regard to two main questions: (a) Did the age difference persist across training? (b) What can we infer about the degree of robustness of age-related limits or constraints in the developmental reserve capacity associated with the method of loci?

Each of these questions was addressed by analyzing a relevant subset of the data displayed in Figure 1. Figure 1 (left panel) shows the average number of words recalled in the correct position by the two age groups at baseline and criterion assessment sessions. Note that scores for the final three assessment were collected in two sessions each.

Robust Age Differences Despite Substantial Reserve Capacity

As already presented in Kliegl et al. (1989, 1990) for the first phase of this study, both young and older subjects had the reserve capacity to acquire the memory strategy trained. After two sessions of instruction (Sessions 5 and 6) and averaging across presentation times of 20 s and 15 s per word, both age groups were able to recall 23.6 ($SD = 5.3$) and 13.3 ($SD = 6.9$) words, respectively. Statistical analysis of this set of data supported this conclusion, yielding significant effects of age, $F(1, 33) = 26.6$, $MS_e = 31.4$, $p < .01$, and of time of assessment, $F(1, 33) = 142.1$, $MS_e = 20.3$, $p < .01$, and a significant interaction between age and time of assessment, $F(1, 33) = 9.4$, $MS_e = 20.3$, $p < .01$.

Did age differences persist across training? The answer seems clear. Age differences in favor of the young were marked, continuing beyond the eight training sessions reported in Kliegl et al. (1989), and were not eliminated by further practice. For example, in a repeated measures analysis of variance on the

² For another purpose, study participants were randomly assigned to two levels of accuracy (50% vs. 80%). However, these two conditions had no differential effect on performance assessments for this study.

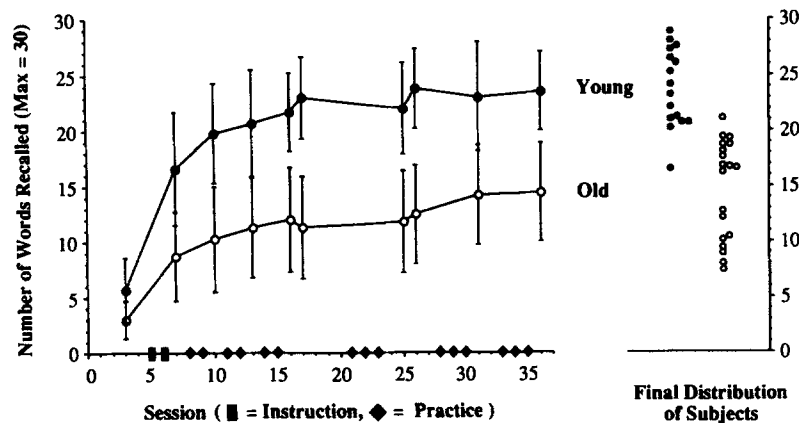


Figure 1. Performance by young and old adults in serial recall of lists of words as a function of training in the method of loci (left panel). (The bars indicate standard deviations. In the right panel, individual scores are given for the last assessment sessions [36/37]. Max = maximum.)

last eight assessments (beginning with Session 10), age group, $F(1, 33) = 59.4$, $MS_e = 115.9$, $p < .01$, and time of assessment, $F(7, 231) = 11.7$, $MS_e = 5.2$, $p < .01$, were highly significant. The interaction between age and time of assessment, however, was not significant. $F(7, 231) = 1.8$, $MS_e = 5.2$, $p < .09$. Thus, although both age groups showed improvement across the various training phases, the higher level of recall of the young adults when compared with the old adults was maintained.

Degree of Robustness of Age Differences in Reserve Capacity

Would the older adults be able to catch up to the level of the young adults with further practice? Within the confines of the present training schedule, there was little evidence for major progress during the second half of training. Two aspects of the magnitude of the negative-age difference are highlighted to illustrate this: level of performance and within- versus between-age variability.

Regarding the average level of performance reached: In the final two criterion assessments at the end of the entire training program (see Figure 1), the old adults still performed at a lower level (14.3 words) than the young adults did in the first 2 assessments after the initial two instructional training sessions (17.0 words), $t(33) = -1.93$, $p < .065$. There was also almost no overlap in the performance of young and old for within- and between-age variability. As shown in the right panel of Figure 1, there was only 1 person from each age group who fell into the distribution range of the other age group. None of the older adults were above the mean of the young adults.

The conclusion of a general age-related reduction in developmental reserve capacity for method of loci performance was further supported by the following two outcomes. The first concerns the temporal stability of individual differences across sessions. Temporal stability coefficients continued to be very high into the last sessions (e.g., young: $r = .90$; old: $r = .83$). Because subjects maintained their positions relative to each other, there was no evidence that some of the older adults might have possessed a higher level of developmental reserve capacity

that, with further training, might have resulted in sudden gains for specific individuals.

The second outcome concerns the role of chronological age in regression analyses in which the effect of age on serial word recall is assessed after statistically controlling for other criterion-relevant predictors. Again in agreement with the results described in more detail in Kliegl et al. (1990) for Phase 1, chronological age turned out to carry unique and important variance as the program of testing the limits unfolded and subjects approached their limits of reserve capacity. Even after criterion-relevant cognitive predictors (serial word recall, cued recall, digit symbol substitution test, card rotation test, and vividness of visual imagery test) collected at pretest were partialled, chronological age accounted for an additional 11% of the variance in the method of loci performance at the last sessions of assessment; for ΔR^2 , $F(1, 28) = 10.0$, $p < .01$, overall $R^2 = .70$. For early sessions of the study, chronological age had not resulted in significant predictive power of performance in serial word recall after criterion-relevant predictors had been controlled for (Kliegl et al., 1990). We submit, therefore, that the testing-the-limits strategy used indeed resulted in an outcome where the age differences observed became increasingly associated with age per se rather than with other markers of cognitive functioning.

Discussion and Conclusion

It is in the nature of studying developmental reserve capacity that the evidence is never final in an absolute sense. Continued practice, new treatments, and other response formats (such as extending the time available for retrieval beyond 10 min) may produce novel information about what is and is not possible in principle. Despite this general limitation, we argue that this extension study provides a strong plausibility case for the existence of robust (something akin to irreversible) negative-age differences in certain factors and processes of the mind associated with the use of the method of loci. Although there are other results in the cognitive aging literature consistent with

this finding (e.g., Craik, 1983; Craik, Byrd, & Swanson, 1987; Salthouse, 1988; Salthouse, Babcock, Skovronek, Mitchell, & Palmon, 1990), none are based on as extensive an exploration of latent potential as reported in this study. We are particularly impressed by two of our findings. The first is that the older adults, even after extensive training, performed, on average, below the level reached by the young adults after only two instructional sessions. The second is that none of the older adults reached a level of performance approaching the average of the young adults. To our knowledge, there is no comparable empirical evidence available in the present literature on cognitive aging.

The results of this study raise three important issues: (a) Are the negative age differences observed indicative of a true aging loss, or are they potentially due to cohort (generational) rather than age effects? (b) What are the origins of the observed negative age differences? and (c) Are there possible exceptions to the negative age trend for persons or cognitive domains of functioning not included in this study?

A final answer to the issue of age versus cohort effects cannot be obtained until longitudinal studies using cohort-sequential designs have been implemented. In our view, however, it is not likely that cohort differences will be of the same magnitude as the negative age differences reported here. This is because previous work demonstrating cohort or generational differences in cognitive performance has focused on the *normal* range and not on peak (testing-the-limits) levels of intellectual performance. Within the normal range of functioning, we know from past cognitive training research that it is possible for the current cohorts of elderly persons to benefit from training to such a degree that their level of performance is similar to that of young adults or of their own earlier adult life stage (Baltes & Lindenberger, 1988; Willis, 1987). However, because *none* of the elderly subjects after training in this study were found to function above the mean of IQ-comparable young subjects, we think it unlikely that an entire cohort of future elderly adults could be moved easily to the level of performance displayed by young adults.

Concerning the question of the likely origins of the observed negative age difference in method of loci performance, we cannot specify the neurophysiological or experiential factors and processes that might be responsible. Performance in the method of loci involves a complex ensemble of components that is not yet sufficiently understood. However, because of the robustness of the findings, we believe that the negative age difference is not easily explainable by the often invoked "disuse-based accounts" of cognitive aging. Rather, neurobiological constraints seem to be a more likely determinant. Such neurobiological constraints can be due, of course, to both a genetically determined program of biological aging as well as to the neurophysiological substrate resulting from a lifetime of experience and cognitive activity. One possibility concerns the various components involved in the method of loci (see also Footnote 1); it might be the use of mental imagination for the production of novel relations that is critical (see also Thompson & Kliegl, 1991).

Although our ability to specify the origin of the observed negative age differences is still limited, we submit, however, that the high degree of experimental control possible with the

cognitive engineering of the method of loci offers promise for the future study of the specific mechanisms involved. In the past, such analysis of the origins of phenomena of cognitive aging was seriously hampered because cognitive age differences obtained were easily modified through cognitive practice, and there were always some older adults who functioned in the same range as young adults. Whenever there is as much fluidity or plasticity in age-comparative data, it is difficult to pinpoint factors or processes that are intrinsic to the biological nature of aging.

Concerning the question of exceptions from the negative aging difference observed, we first need to emphasize the content specificity of this study's finding. It involves a component of what in psychological intelligence theory is called *fluid intelligence* or the *mechanical* aspects of cognitive functioning. As argued by others and by us (e.g., Baltes & Smith, 1990; Denney, 1990; Perlmutter, 1990), there are other facets of the mind (e.g., knowledge-based domains of functioning such as language comprehension, professional expertise, or wisdom) in which old adults may be equivalent to or better than young adults, even in situations that require maximum use of reserve capacity. Therefore, it is important not to generalize this study's outcome to all domains of the mind. This most certainly would be inappropriate.

Second, there is the question of individual exceptions from the aging loss observed. In general, this study's sample of healthy and well-functioning older adults seems to represent a good first effort at finding such persons. At the same time, it is important to continue the search for aging individuals who might not be affected by an apparent reduction in developmental reserve capacity associated with the method of loci. For example, certain older adults with special talents or a lifelong practice in mental imagination, such as graphic designers or memory specialists, might not display the performance deficit obtained in this study. However, a recent study conducted in our laboratory with well-functioning older adult graphic designers did not corroborate this possibility (Lindenberger, Kliegl, & Baltes, 1991). Although older adult graphic designers showed less of a performance deficit in the method of loci than normal older control subjects, they did not attain the level of performance of normal younger adults.

In summary, we contend that this extension of the Kliegl et al. (1989, 1990) work from 20 to 38 experimental sessions (or from 8 to 17 training sessions) has further strengthened the original findings. Despite sizable developmental reserve capacity in old age, there is a robust, if not irreversible, negative age difference in some basic components of the mind relevant for the use of the method of loci in achieving superior levels of memory performance. This fact was demonstrated by using a testing-the-limits paradigm that is explicitly geared toward the laboratory study of the upper limits of performance (Baltes, 1987; Kliegl & Baltes, 1987).

References

- 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., & Lindenberger, U. (1988). On the range of cognitive

- plasticity in old age as a function of experience: 15 years of intervention research. *Behavior Therapy*, 19, 283–300.
- Baltes, P. B., & Smith, J. (1990). The psychology of wisdom and its ontogenesis. In R. J. Sternberg (Ed.), *Wisdom: Its nature, origins, and development* (pp. 87–120). Cambridge, England: Cambridge University Press.
- Bower, G. H. (1970). Analysis of a mnemonic device. *American Scientist*, 58, 496–510.
- 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.
- Craik, F. I. M., Byrd, M., & Swanson, J. M. (1987). Patterns of memory loss in three elderly samples. *Psychology and Aging*, 2, 79–86.
- Denney, N. W. (1990). Adult age differences in traditional and practical problem solving. In E. A. Lovelace (Ed.), *Aging and cognition* (pp. 329–349). Amsterdam: Elsevier.
- 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. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology*, 25, 247–256.
- Kliegl, R., Smith, J., & Baltes, P. B. (1990). On the locus and process of magnification of age differences during mnemonic training. *Developmental Psychology*, 26, 894–904.
- Lindenberger, U. (1991). *Aging, professional expertise, and cognitive plasticity*. Stuttgart, Germany: Klett-Cotta.
- Lindenberger, U., Kliegl, R., & Baltes, P. B. (1991). *Professional expertise does not eliminate negative age differences in imagery-based memory performance during adulthood*. Unpublished manuscript, Max Planck Institute for Human Development and Education, Berlin, Federal Republic of Germany.
- Marschark, M., & Hunt, R. R. (1989). A reexamination of the role of imagery in learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 710–720.
- Perlmutter, M. (Ed.). (1990). *Late-life potential*. Washington, DC: Gerontological Society of America.
- Robertson-Tchabo, E. A., Hausman, C. P., & Arenberg, D. (1976). A classical mnemonic device for older learners: A trip that works! *Educational Gerontology*, 1, 215–226.
- Salthouse, T. A. (1988). Resource-reduction interpretations of cognitive aging. *Developmental Review*, 8, 238–272.
- Salthouse, T. A., Babcock, R. L., Skovronek, E., Mitchell, D. R. D., & Palmon, R. (1990). Age and experience effects in spatial visualization. *Developmental Psychology*, 26, 128–136.
- Thompson, L. A., & Kliegl, R. (1991). Adult age-differential effects of plausibility on memory: The role of time constraints during encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 542–545.
- Willis, S. L. (1987). Cognitive training and everyday competence. In K. W. Schaie (Ed.), *Annual review of gerontology and geriatrics* (Vol. 7, pp. 159–188). New York: Springer.
- 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.

Received October 22, 1990

Revision received April 26, 1991

Accepted May 7, 1991 ■