



Universität Potsdam

Paul B. Baltes, Freya Dittmann-Kohli, Reinhold Kliegl

## Reserve capacity of the elderly in aging-sensitive tests of fluid intelligence : replication and extension

first published in:  
Psychology and aging, 1 (1986) 2, S. 172-177, ISSN 0882-7974,  
DOI 10.1037/0882-7974.1.2.172

Postprint published at the Institutional Repository of the Potsdam University:  
In: Postprints der Universität Potsdam  
Humanwissenschaftliche Reihe ; 143  
<http://opus.kobv.de/ubp/volltexte/2009/3993/>  
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-39939>

Postprints der Universität Potsdam  
Humanwissenschaftliche Reihe ; 143

# Reserve Capacity of the Elderly in Aging-Sensitive Tests of Fluid Intelligence: Replication and Extension

Paul B. Baltes, Freya Dittmann-Kohli, and Reinhold Kliegl  
Max Planck Institute for Human Development and Education, Berlin, FRG

Fluid intelligence belongs to that cluster of intellectual abilities evincing aging loss. To examine further the range of intellectual reserve available to aging individuals and the question of replicability in a new cultural and laboratory setting, 204 healthy older adults (mean age = 72 years; range = 60–86) participated in a short-term longitudinal training study. For experimental subjects, 10 sessions consisted of cognitive training involving two subability tests (Figural Relations, Induction) of fluid intelligence. The pattern of outcomes replicates and expands on earlier studies. Older adults have the reserve to evince substantial increases in levels of performance in fluid intelligence tests. Transfer of training, however, is narrow in scope. Training also increases accuracy of performance and the ability to solve more difficult test items. Difficulty level was estimated in a separate study, with a comparable sample of  $N = 112$  elderly adults. Future research is suggested to examine whether intellectual reserve extends to near-maximum levels of performance.

A theoretically and practically significant question regarding the often-reported aging loss in fluid intelligence (Cattell, 1971; Horn, 1970, 1982) is its ability to be modified, or its plasticity (Baltes, Dittmann-Kohli, & Dixon, 1984; Baltes & Willis, 1982; Willis, 1985). To what degree and under what conditions do older adults maintain a capacity to raise their level of functioning in fluid intelligence? Knowledge about this topic is important, irrespective of one's position on the question of whether aging loss in fluid intelligence is empirically substantial.

In the past few years, a number of cognitive training studies involving aging-sensitive abilities of fluid intelligence have been performed with healthy older adults, primarily in one laboratory at the Pennsylvania State University (Baltes & Willis, 1982; Dittmann-Kohli, 1983; Willis, 1985; for cognitive training research dealing with other forms of cognition, see Denney, 1982; Labouvie-Vief, 1985; Overton & Newman, 1982). In terms of measures and subabilities, the cluster of fluid intelligence is indicated by mental abilities such as figural relations and induction. For each of these subabilities, educationally based training programs were developed. Findings from several single-subability training studies (Blieszner, Willis, & Baltes, 1981; Hofland, Willis, & Baltes, 1981; Willis, Blieszner, & Baltes, 1981) indicated that many older adults in the age range from 60 to 80 have the capacity

to raise significantly their level of performance in tests of fluid intelligence.

In the present article we report a study that, through replication and extension, contributes to enlarging our data base about the topic of intellectual reserve capacity in old age. The replication part addresses the question of whether performing training research in a different laboratory and cultural context (West Germany) and with a larger sample, results in similar outcomes. Extension consisted of two aspects. Training was expanded to include, in the form of dual-ability training, more than one subability of fluid intelligence in the training module. Thereby we explored further the scope of intellectual reserve available to older adults. In addition, we considered the question of level of difficulty and accuracy of performance. This permitted us not only to examine average level of performance, but also to explore some of the processes associated with changes in performance level.

## Method

Detailed information about rationale, test battery, and training programs is contained in earlier publications (Baltes & Willis, 1982; Blieszner et al., 1981, Willis et al., 1981).

## Subjects

The study sample ( $N = 204$ ), on which the main analyses reported here are based, consisted of 155 female and 49 male older adults (average age = 72 years; range = 60–86 years) from Berlin, FRG. All of the subjects were volunteers, physically able to come to the laboratory. Reported subjective health was above average ( $M = 3.7$ ,  $SD = .7$ , on a 5-point self-report scale). Average educational level (roughly comparable to U.S. information on educational history) was 11 years ( $SD = 2.7$ ). When comparing level of intellectual performance on equivalent tests (figural relations, induction) with the U.S. samples from rural Pennsylvania (e.g., Hofland et al., 1981), the Berlin sample scored about 0.5  $SD$  higher. In terms of interindividual variability, the Berlin sample was also slightly more heterogeneous.

---

In addition to expressing heartfelt appreciation to our research participants, who contributed much of their time and energy to this project, we thank Anita Günther, Annette Rentz, Peter Grund, Eva Derge, Eva Grün, and Margot Sperling for their assistance in data collection and data analysis. Collegial consultation by Sherry L. Willis and Carolyn Nesselroade in staff training involving the testing and training modules is also acknowledged with much gratitude. Roger A. Dixon, Daniel P. Keating, Jacqui Smith, and Sherry L. Willis provided helpful comments on an earlier draft of this article.

Correspondence concerning this article should be addressed to Paul B. Baltes, Max Planck Institute for Human Development and Education, Lentzeallee 94 D-1000, Berlin 33, FRG.

## Procedure and Design

The experimental-control group design involved three main parts: a pretest, cognitive training, and three posttests administered 1 week, 1 month, and 6 months following training. The test battery given at pretest and at all posttests was identical and consisted of eight subtests (cited later in the article). Participants were randomly assigned to one of two experimental groups or to a no-contact control group. Earlier research (Blieszner et al., 1981) had shown that a contact control group was not necessary.

The cognitive training program consisted of 10 sessions conducted in 1 month. Five consecutive sessions each dealt with the fluid ability figural relations, and the other 5 with induction. The sequence of presentation of the two training modules was counterbalanced. One half of all of the subjects (experimental and control) participated in an additional intermediate posttest after experimental subjects had completed the first 5 training sessions, which dealt with one or the other of the two training modules. At this intermediate posttest, the same test battery was given as at the other occasions of measurement. The intermediate test given to a subsample after 5 sessions permitted a replication of the single-ability training studies, which dealt, however, with only 1 posttest occasion.

Thus, the main study design followed a  $3 \times 2 \times 8 \times 3$  (Groups [two training constellations, one control]  $\times$  Intermediate Test  $\times$  Tests  $\times$  Posttests) arrangement, with groups and intermediate as between-subject, and tests and posttests as within-subject factors. Analyses dealing with the counterbalanced presentation of the two training modules are not presented here. Their findings do not alter the results reported.

## Measures

At pretest and at each of the three posttests, a 3.5-hr battery of intelligence tests was administered. The battery was chosen to measure the domain of fluid and crystallized intelligence such that a fine-grained analysis on the level of subabilities of fluid intelligence was possible. In addition, because training may result in increase of speed of information processing, two measures of the intelligence factor, perceptual speed, were included (Baltes & Willis, 1982). Depending on their similarity to the training program, the eight tests of the battery can be ordered from near-near through near to far transfer (as identified in Table 1 and graphically arranged in Figure 1). Information about the factorial composition is reported by Baltes, Cornelius, Spiro, Nesselroade, and Willis (1980).

## Dropout Analyses

Eleven percent of the participants dropped out during pretest, before being assigned to design conditions. Of those completing the pretests, 67% of the experimental and 72% of the control subjects had complete data protocols for the entire 8-month period of the study. Subjects with complete data protocols at pretest and at all posttests constitute the study sample ( $N = 204$ ). The possible selection effect of dropout was analyzed using pretest scores on several control variables (age, reported health, education, and average test performance). Dropouts differed in two of the variables from the continuing participants. They showed lower initial test performance ( $0.7 SD$ ) and had a lower average level of education ( $0.8 SD$ ). The continuing experimental and control subjects, however, did not differ in dropout characteristics. Thus, the dropout selection effect affected only the overall characteristics (external validity) of the sample and not the internal validity of the experimental-control comparison.

## Training Program

The two experimental groups were trained in two subabilities of fluid intelligence: figural relations (Willis et al., 1981) and induction (Blieszner

et al., 1981). Each of the two fluid subabilities was trained for five consecutive sessions of 1 hr each. The two experimental groups differed only in the order in which they received the training.

The training programs were developed through task analysis of the items and rules contained in established tests of figural relations and induction. None of the items used in training was identical in content to the ones constituting the standardized test forms used to assess training effectiveness. The training program, conducted in small groups of 6 to 12 participants, focused on helping subjects to identify rules and concepts and to use them in solving the types of problems associated with figural relations and induction. The tutor and participating subjects modeled how to identify and use the rules; subjects practiced employing these strategies. Feedback and discussion about these strategies, as well as about the possibility of using alternate strategies, followed. Further information about the training program is contained in Blieszner et al. (1981) and Willis et al. (1981).

## Results and Discussion

The data were transformed into a common metric for all tests ( $M = 50$ ,  $SD = 10$ ), using the performance of all of the control subjects at pretest as the measurement space. Individual changes from pretest scores were used as dependent variables and analyzed with a multivariate analog of a  $3 \times 2 \times 8 \times 3$  mixed-model analysis of variance (ANOVA) and a priori specified orthogonal contrasts.

There is controversy in the literature over the use of change or gain scores (Cronbach & Furby, 1970). However, several methodologists (Nesselroade, Stigler, & Baltes, 1980; Nunnally, 1982) have strongly supported the use of gain scores whenever (a) the substantive target is the direct representation of change and (b) the measures involved do not exhibit unsatisfactory reliability. Unreliability is the major culprit in change score analysis. In this study, both criteria are fulfilled. The critical question is the range of change associated with training, and gain scores are the most direct measure of change. Furthermore, the reliability of pre- and posttest measures are good, and thus error of measurement problems (e.g., regression toward the mean) would be expected to affect interpretation of results only minimally. To double check, we have also conducted analyses of covariance on posttest scores, with pretest scores as covariates. The key outcomes were the same as those reported later in the article.

The ANOVA yielded significant main effects on all four design variables: Group,  $F(2, 198) = 20.6$ ; intermediate test,  $F(1, 198) = 19.8$ ; type of test,  $F(7, 1386) = 65.4$ ; and time of assessment,  $F(2, 369) = 41.0$  (all  $ps < .01$ ). In addition, there were significant interactions between group and test,  $F(14, 1386) = 4.3$ ,  $p < .01$ , and between intermediate test and time,  $F(2, 369) = 5.1$ ,  $p < .01$ .

## Training Effects and Pattern of Transfer

Of central concern for the present study were differences between groups and their interactions with tests indexing different degrees of potential transfer. The group main effect and its interactions with performances on different tests were entirely restricted to differences between experimental and control subjects (contrast control vs. experimental subjects,  $F(1, 198) = 40.5$ ,  $p < .01$ ; contrast between two experimental groups,  $F(1,$

Table 1  
Transfer Assessment Battery

General intelligence dimension	Primary mental ability	Predicted transfer pattern of marker measures	Source
Fluid	Figural relations	Near-near transfer ADEPT Figural Relations Test (Form B)	Plemons, Willis, & Baltes, 1978
		Near transfer Culture Fair Test (Scale 2, Form A) Power Matrices (Scale 3, Form A, 1963 ed., and Form B, 1961 ed.) Raven's Advanced Progressive Matrices (Set II)	Cattell & Cattell, 1957 Cattell & Cattell, 1961, 1963 Raven, 1962
	Induction	Near-near transfer ADEPT Induction Test (Form B): Letter Sets, Number Series, Letter Series	Blieszner, Willis, & Baltes, 1981
		Near transfer Induction Standard Tests: Letter Sets, Number Series, Letter Series	Ekstrom, French, Harman, & Derman, 1976; Thurstone, 1962
Speed	Perceptual speed	Far nonfluid transfer Identical pictures Number comparison	Ekstrom et al., 1976 Ekstrom et al., 1976
Crystallized	Verbal comprehension	Far nonfluid transfer Vocabulary (V-2)	German form of KFT (Beltz); similar to Ekstrom et al., 1976

Note. Bibliographic references to sources of tests are contained in Willis, Blieszner, and Baltes (1981). The terms *fluid*, *speed*, and *crystallized* refer to general intelligence dimensions. All measures were included at all occasions of assessment. More complete information on tests and their factorial structure is contained in Baltes, Cornelius, Spiro, Nesselroade, and Willis (1980).

198) = .3, *ns*). Thus, it did not matter in which order subjects were instructed in the two subabilities.

What about the pattern of transfer involving tests differing in degree of relatedness to the training? The critical Group × Test interaction is displayed in Figure 1. This interaction also was

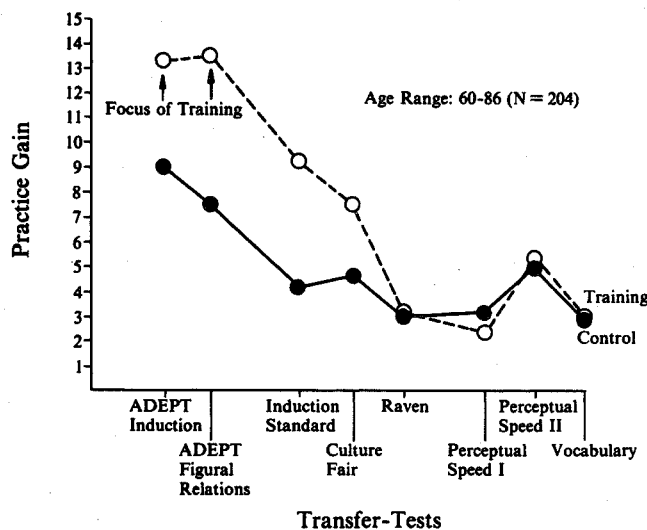


Figure 1. Retest and training gains of training and control groups in transfer tests of intelligence averaged across three posttests (1 week, 1 month, and 6 months following training). (Tests are ordered on a continuum of decreasing similarity to training focus: fluid, 1-5; perceptual speed, 6-7; crystallized, 8.)

analyzed further by orthogonal contrasts. These orthogonal contrasts showed that the training and control groups differed in favor of the training group on the two near-near and two of the three near-transfer tests only. This statistical conclusion squares well with the graphic representation contained in Figure 1. For the four closest transfer measures, the training gain in terms of the pretest metric amounts to 1.35 SD for the near-near tests and 0.85 SD for the next level of transfer tests. These gains reflect retesting plus training.

The spectrum of transfer obtained suggests that there is some generalization to other tests of induction and figural relations abilities. However, generalization within the fluid ability cluster was not comprehensive because there was no significant training effect on the Raven Advanced Progressive Matrices, which is also a measure of fluid intelligence. This finding differs from previous research in which moderate transfer to the Raven was obtained in a study dealing with training of figural relations (Willis et al., 1981).

Intermediate Test

Note first that participation in the intermediate test did not interact significantly with the test battery and the experimental/control-group distinction. The Intermediate Test × Time of Measurement interaction indicated that although the advantage of intermediate test participation was maintained over all three posttests, the benefit was especially pronounced at the first posttest (contrast for interaction between intermediate test and difference between Posttest 1 and the average of Posttests 2 and 3,  $F(1, 198) = 8.9, p < .01$ ; contrast for interaction between inter-

mediate test and difference between Posttests 2 and 3,  $F(1, 198) = 1.8, ns$ ).

### *Maintenance of Training and Retest Effect*

The significant effect of time of posttraining assessment (three posttests) was further analyzed by orthogonal contrasts. They indicated that there still was significant improvement from the first to the average of the second and third posttest,  $F(1, 198) = 58.5, p < .01$ . Performance dropped significantly, however, from the second to the third posttest,  $F(1, 198) = 20.4, p < .01$ . Nevertheless, mean performance after 6 months was still slightly above the performance at the first posttest.

### *Retesting*

The effect of testing practice, measured most directly in the four test repetitions of the control group, was substantial (see Figure 1). The effect of testing affected all tests of the battery. The magnitude of average retest gains from pretest to the average of the three posttests was 0.85 *SD* for the four near–near and near–transfer tests when measured in the initial metric of pretest.

### *Subject Differences*

We analyzed also whether amount of gain varied with subject characteristics (initial performance level, age, and education). The design permitted an age variation along lines of a young–old versus old–old distinction. Using an unweighted average of the four tests that showed significant training effects, a significant main effect of age at the 1% level of significance (persons under 70 years of age exhibited a larger gain than did older persons, amounting to 0.1 *SD*) and a marginally significant Initial Performance Level  $\times$  Treatment interaction (0.1 *SD*,  $p < .1$ ) were obtained. Persons with lower initial ability or level of performance profited less from training. Compared to the magnitude of the main effects of training and testing, however, these age effects are rather small. The dominant statistical finding is that both the young–old and the old–old exhibited substantial gains.

### *Level of Difficulty and Accuracy*

A new feature of this study is the attempt to examine more broadly than in past research whether changes in level of test performance are accompanied by changes in the qualitative nature of performance. Level of difficulty and accuracy were considered for this purpose. For example, it could be that trained subjects may have improved exclusively on items that were relatively easy, or that they simply tried to answer more items than did control subjects, thereby benefitting from guessing (Hofland et al., 1981).

*Level of difficulty.* Because test order of items is fixed in standardized psychometric tests of intelligence and because many tests are given under speeded conditions, serial position and difficulty level of test items are easily confounded (often subjects do not reach the last items of a test). Difficulty level, therefore, needs to be estimated from independent evidence.

For this purpose, a parallel study with a comparable sample ( $N = 112$ ) was conducted in which level of difficulty of the items

considered in the near–transfer tests was estimated without the confounding produced by serial position and speeded instruction. Induction and Figural Relations tests were given to random subsamples in test halves and counterbalanced order (odd: ascending, descending; even: ascending, descending). In addition, test time was extended by a factor of four in order to approximate a power condition of assessment. Based on these results, average probability of correct answers was computed as an estimate of difficulty level for each test item. The difficulty level per item obtained was then used to categorize test items of the main study into thirds: easy, medium, and difficult.

When analyzing performance of subjects in the main study at the first posttest by level of difficulty, we obtained the following results. All of the subjects solved more items correctly in each of the three difficulty levels compared to their pretest performance. Furthermore, experimental (ability-trained) subjects solved significantly more items at all levels of difficulty than did control subjects. Specifically, averaged across the four tests exhibiting training gains, experimental subjects correctly solved 8%, 9%, and 8% more items, respectively, in the easy, medium and difficult conditions. Thus, the training effect was not restricted to items of low-level difficulty.

*Accuracy.* When computing the total number of items attempted in the four near–transfer tests at the first posttest, the experimental and control subjects did not differ. The average percentage of unattempted answers was 22.5% for experimental and 23% for control subjects. Level of performance differences between experimental and control subjects, therefore, are likely to be associated with differences in accuracy.

Accuracy was analyzed by considering percentage of correct and wrong answers (commission errors).<sup>1</sup> As to correct answers, experimental and control subjects significantly improved their accuracy at all levels of difficulty. What about wrong answers (commission errors)? Here, control subjects showed no change, whereas experimental subjects significantly decreased their percentage of error at all levels of difficulty (5%, 11%, and 9%, respectively, for difficult, medium, and easy items). The counterpart nonsignificant data for control subjects were an increase of 4% for difficult, and a decrease of 3% and 1%, respectively, for medium and easy items. Increased accuracy, then, seems to be an added benefit of ability-specific training when compared to retesting experience alone.

### Conclusions

We interpret the present study as demonstrating replicability of findings collected in another laboratory, on a larger and different (city vs. rural) sample, and in a different cultural setting. The fact of replication gains in significance for two reasons. First, educationally based training research such as the present study involves a fair degree of uncertainty, as the training program is not fully anchored in fixed steps of implementation. In educa-

<sup>1</sup> The third possible category of scoring deals with unanswered items (omission errors). Because of its ipsative relation to the other two, it was not considered further in statistical analysis. Note, however, that experimental and control subjects did not differ in level of unanswered items.

tional intervention research, positive replication is a somewhat rare though much aspired outcome. Second, sample selectivity in Pennsylvania and in Berlin differs considerably, with the Berlin sample representing a more educated population. Together, the two samples cover a fairly broad spectrum of demographic variation.

The major exception to replicability involves lack of transfer to the Raven, a test which in our view is to be considered as a part of the near-transfer spectrum. On this basis, we conclude that the transfer pattern obtained (to the four nearest transfer tests only) is indicative of fairly narrow limits of transfer.

The extension part of the study showed that reserve capacity of many older individuals in fluid intelligence is even more extensive than indicated in previous research. Our subjects exhibited sizeable training gains following experience with two abilities covering a broader range of fluid intelligence. In addition, we have shown (a) that subjects were able to solve more difficult items following practice and training and (b) that ability-specific training resulted in reduced error of performance (or increased accuracy) at all levels of difficulty. The benefits accruing from training, therefore, are not due only to an increase in response speed. Added cognitive skills, permitting more accurate problem solution and the solving of more difficult items, seem to be present as well.

To prevent possible misinterpretation of our research on reserve capacity, some cautionary comments are necessary. The results do *not* imply the following three conclusions. The first would be that older adults profit more from intelligence training than do young adults. As there are no good age-comparative studies on training gain, knowledge about the extent of reserve capacity available in different age groups is not available. The second possible misinterpretation deals with the question of scope of the training effect. The training program did not raise level of intelligence as a whole but is restricted to those test domains that were part of training. We have, however, no evidence that a more broadly based increase of intellectual performance might not, in principle, be possible if training were wider in scope. The third possible misinterpretation would be that all aging individuals display the reserve capacity reported here. We do not know the subject generality of our findings. Whereas we do believe that many elderly individuals up to the age studied are capable of major training benefits, as reported here, we need to reemphasize that our samples are clearly biased toward the healthy elderly and most likely do not include persons who suffer from manifest brain-related diseases such as senile dementia of the Alzheimer type (Hoffmeister & Müller, 1979; La Rue & Jarvik, 1982).

In related conceptual papers (Baltes, 1984; Baltes et al., 1984; Kliegl & Baltes, in press), we have expanded theoretically on several lines of research emanating from the existence of substantial intellectual reserve capacity in older adults. Among others, one focus stressed has been to extend cognitive training research to the study of limits of functioning. We have argued that the reserve capacity available to older adults may be restricted to the "normal" range of intellectual functioning in terms of level of difficulty (Fries & Crapo, 1981; Salthouse & Somberg, 1982). At near-maximum levels of performances, we expect to obtain, possibly, a different window for the study of intellectual reserve capacity.

## References

- Baltes, P. B. (1984). Intelligenz im Alter [The aging of intelligence]. *Spektrum der Wissenschaft*, 5, 46-60.
- Baltes, P. B., Cornelius, S. W., Spiro, A., Nesselroade, J. R., & Willis, S. L. (1980). Integration versus differentiation of fluid/crystallized intelligence in old age. *Developmental Psychology*, 16, 625-635.
- 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. 33-76). New York: Academic Press.
- Baltes, P. B., & Willis, S. L. (1982). Plasticity and enhancement of intellectual functioning in old age: Penn State's Adult Development and Enrichment Program (ADEPT). In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and cognitive processes* (pp. 353-389). New York: Plenum Press.
- Blieszner, R., Willis, S. L., & Baltes, P. B. (1981). Training research in aging on the fluid ability of inductive reasoning. *Journal of Applied Developmental Psychology*, 2, 247-265.
- Cattell, R. B., & Cattell, A. K. S. (1957). *Test of "g": Culture Fair (Scale 2, Form A)*. Champaign, IL: Institute for Personality and Ability Testing.
- Cattell, R. B., & Cattell, A. K. S. (1961, 1963). *Test of "g": Culture Fair (Scale 3, Form A, 1963 Ed.; Form B, 1961 Ed., 2nd)*. Champaign, IL: Institute for Personality and Ability Testing.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin*, 74, 68-80.
- Denney, N. W. (1982). Aging and cognitive changes. In B. B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 807-827). Englewood Cliffs, NJ: Prentice-Hall.
- Dittmann-Kohli, F. (1983). Intelligenzforschung im höheren Erwachsenenalter [Research in intelligence in old age]. *Unterrichtswissenschaft*, 4, 361-369.
- Ekstrom, R. B., French, J. W., Harman, J., & Derman, D. (1976). *Kit of factor-referenced cognitive tests, 1976 revision*. Princeton, NJ: ETS.
- Fries, J. F., & Crapo, L. M. (1981). *Vitality and aging*. San Francisco: Freeman.
- Hoffmeister, F., & Müller, C. (1979). *Brain functioning in old age*. Heidelberg: Springer.
- Hofland, B. F., Willis, S. L., & Baltes, P. B. (1981). Fluid intelligence performance in the elderly: Intraindividual variability and conditions of assessment. *Journal of Educational Psychology*, 73, 573-586.
- Horn, J. L. (1970). Organization of data on life-span development of human abilities. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 423-466). New York: Academic Press.
- Horn, J. L. (1982). The aging of human abilities. In B. B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 847-870). Englewood Cliffs, NJ: Prentice-Hall.
- Kliegl, R., & Baltes, P. B. (in press). 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*. Norwood, NJ: Ablex.
- Labouvie-Vief, G. (1985). Intelligence and cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 500-530). New York: Van Nostrand Reinhold.
- La Rue, A., & Jarvik, L. F. (1982). Old age and biobehavioral changes. In B. B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 791-806). Englewood Cliffs, NJ: Prentice-Hall.
- Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin*, 88, 622-637.

- Nunnally, J. C. (1982). The study of human change: Measurement, research strategies, and methods of analysis. In B. B. Wolman (Ed.), *Handbook of developmental psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Overton, W. F., & Newman, J. L. (1982). Cognitive development: A competence-activation/utilization approach. In T. Field, A. Houston, H. Quay, L. Trull, & G. Finley (Eds.), *Review of human development*. New York: Wiley.
- Plemons, J. K., Willis, S. L., & Baltes, P. B. (1978). Modifiability of fluid intelligence in aging: A short-term longitudinal training approach. *Journal of Gerontology*, 33, 224-231.
- Raven, J. C. (1962). *Advanced progressive matrices, Set II, 1962 Revision*. London: H. K. Lewis.
- Salthouse, T. A., & Somberg, B. L. (1982). Skilled performance: Effects of adult age and experience on elementary processes. *Journal of Experimental Psychology: General*, 111, 176-207.
- Thurstone, T. G. (1962). *Primary mental abilities, Grades 9-12, 1962 Revision*. Chicago, IL: Science Research Associates.
- 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.
- Willis, S. L., Blieszner, R., & Baltes, P. B. (1981). Intellectual training research in aging: Modification of performance on the fluid ability of figural relations. *Journal of Educational Psychology*, 73, 41-50.

Received January 30, 1985

Revision received November 13, 1985 ■