Universität Potsdam

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

On the locus of training gains in research on the plasticity of fluid intelligence in old age

On the Locus of Training Gains in Research on the Plasticity of Fluid Intelligence in Old Age

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

Cognitive training research has shown that many older adults have a substantial reserve capacity in fluid intelligence. Little is known, however, about the locus of plasticity. Two studies were conducted to examine whether training gains in fluid abilities are critically dependent on experimenter-guided training and/or whether older adults can achieve similar improvements by themselves on the basis of cognitive skills already available in their repertoire. Several comparisons were made: (a) between test performances after trainer-guided training in ability-specific cognitive skills and after self-guided retest practice (without feedback), (b) between performances under speeded and power conditions of assessment, (c) between performances on easy and difficult items, and (d) between the relative numbers of correct and wrong answers. Results suggest that a large share of the training improvement shown by the elderly can plausibly be explained as the result of the activation and practice of cognitive skills already available in their repertoire. The results also have implications for educational practice, pointing to the appropriateness of strategies of self-directed learning for many elderly adults.

A substantial body of research shows that older adults can benefit markedly from cognitive training of fluid intelligence (Baltes, Dittmann-Kohli, & Kliegl, 1986; Baltes & Willis, 1982; Dittmann-Kohli, 1983; Willis, 1985). Fluid intelligence is a cluster of intellectual abilities that declines with age (Cattell, 1971; Horn, 1982). Findings on the improvement of performance on fluid tests of intelligence are similar to results found in cognitive training research with the elderly conducted in other domains of cognitive functioning (Baltes & Kliegl, 1986; Labouvie-Vief, 1985; Salthouse, 1985). Especially in the area of psychometric intelligence, however, little is known about the mechanisms that may be responsible for training gains.

A central issue is whether performance gains from cognitive training are critically dependent on guided instruction in criterion-relevant problem-solving skills. Alternatively, it may be that older adults can demonstrate comparable gains through self-instruction without direct assistance from a trainer. Theoretically, this question may be phrased as whether the training gains reported are the result of (a) an activation of cognitive skills already available in the repertoire of older adults or (b) the learning of new cognitive skills that would not be easily acquired through self-directed learning (Dittmann-Kohli, 1981; Overton & Newman, 1982). Because previous research with the elderly has produced training gains rather easily and consistently (Baltes & Kliegl, 1986; Labou-

vie-Vief, 1985; Willis, 1985), our general assumption is that the source of such gains with fluid intelligence largely reflects the activation and use of latent, existing skills rather than the acquisition of skills introduced by the trainer. According to this theoretical interpretation, past cognitive training with tests of intelligence has been successful because elderly subjects received practice in the activation and use of preexisting cognitive skills.

In this article, two studies with comparable samples of older adults are described. In these studies, three design parameters were varied in order to obtain different levels of self-activation of skills versus practice of skills introduced by the experimenter: (a) a power (time-extended) condition of assessment, (b) retest practice (without feedback), and (c) experimenter-guided ability training with feedback as to the nature of rules and problem-solving skills associated with the ability trained. The power condition of assessment should most clearly indicate the skills available to research participants without outside instruction. The next level of self-activation without outside assistance is associated with retest practice. If older adults without outside instruction could achieve equivalent levels of test performance as ability-trained subjects, it would be likely that training gains are largely a function of self-guided activation of cognitive skills that are already available in elderly adults' repertoire. When comparing the training gains achieved, not only do we consider the number of correct responses, but, because the possible advantage of ability training may involve aspects of the "quality" of performance, we look at two additional characteristics: relative amount of error (accuracy) and the level of difficulty of the items solved.

General Method

The experiments involved training programs and a battery of intelligence tests used in previous studies of cognitive training in older adults. This is so because our interest is in understanding the locus

of training gains observed in past work and based on a well-standardized procedure of cognitive training. Thus the information on the nature of training and test assessment, given later, is restricted to a minimum. Detailed information about the training programs and the test batteries, including the rationale for test selection, is contained in earlier publications (Baltes et al., 1986; Baltes & Willis, 1982; Blieszner, Willis, & Baltes, 1981; Willis, Blieszner, & Baltes, 1981).

In this section, we describe common features of the two experiments, including subject recruitment and measures of psychometric intelligence.

Subjects

The two study samples were recruited by means of newspaper advertisements and radio announcements in the city of (West) Berlin. All subjects were volunteers and physically able to come to the laboratory. They were paid DM 25 ($15) for each individual session. When compared with the general population of older adults residing in Berlin, both samples were in relatively good health and were better educated. Such a positive bias was desired because it reduced the likelihood of having a sample with a high proportion of disease-related conditions associated with secondary aging.

The first sample ($n = 87$) consisted of 27 male and 60 female older adults (mean age = 73 years; range 63–89 years). The initial sample involved 89 subjects. Two persons had incomplete data protocols. Reported subjective health was above average ($M = 3.8$, $SD = 0.6$) on a 5-point scale, where 5 indicated perfect subjective health and 1 indicated very bad health. Average educational level was 12.4 years ($SD = 3.0$).

The second sample ($n = 60$) consisted of 21 male and 39 female older adults (mean age = 74 years; range 65–87). Reported subjective health again was above average ($M = 3.9$, $SD = 0.6$) on the 5-point scale. Average educational level was 12.9 years ($SD = 3.1$).

The two samples were not drawn randomly at the same time from a precisely defined population frame. Rather, they were drawn successively as the program of research unfolded. Thus the question of sample comparability can be assessed only indirectly. On the basis of the information reported above about educational level, age, and subjective health, the two samples seem rather similar. Moreover, in comparison with the United States samples used in previous cognitive training research with the elderly in Pennsylvania (e.g., Baltes & Willis, 1982), the German samples scored about one-half standard deviation higher on the tests of intelligence, and, in terms of their scores, there was also slightly more heterogeneity.

Measures of Psychometric Intelligence

Intellectual functioning was measured by a test battery (see Table 1) that was identical to the one used in earlier research (Baltes et al., 1986; Baltes & Willis, 1982). The battery was selected to cover a large spectrum of psychometric intelligence. Because the focus of ability-specific training is on fluid intelligence, that cluster was overrepresented in the test battery.

In these two experiments, we did not use the entire test battery shown in Table 1. However, the critical test variables were always selected from this battery and, if necessary, were modified, as described in the separate reports of the experiments.

Experiment 1

Objective

The focus of Experiment 1 was on the comparative analysis of training effects associated with experimenter-guided ability training versus self-guided retest practice.

Table 1
Transfer Assessment Battery

Primary mental ability predicted transfer pattern	Source
Intelligence dimension: Fluid	
Figural relations	
Near–near transfer: ADEPT Figural Relations Test (Form B)	Plemons, Willis, & Baltes (1978)
Near transfer: (a) Culture Fair Test (Scale 2, Form A) and (b) Power Matrices (Scale 3, Form A, 1963 ed., and Form B, 1961 ed.)	(a) Cattell & Cattell (1957) and (b) Cattell & Cattell (1961, 1963)
Near transfer: Raven's Advanced Progressive Matrices (Set II)	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) and Thurstone (1962)
Intelligence dimension: Speed	
Perceptual speed	
Far nonfluid transfer: Identical pictures, number comparison	Ekstrom et al. (1976)
Intelligence dimension: Crystallized	
Verbal comprehension	
Far nonfluid transfer: Vocabulary (V-2)	Kognitiver Fähigkeitstest (Form A; Heller, Gaedicke, & Weinläder, 1976), similar to that of Ekstrom et al. (1976).

Note. The terms *fluid, speed,* and *crystallized* refer to general intelligence dimension. More complete information on tests and their factorial structure is contained in Baltes, Cornelius, Spiro, Nesselroade, and Willis (1980). ADEPT = Adult Development and Enrichment Project.

In earlier research (Baltes et al., 1986; see also Willis, 1985), retest practice alone produced significant increases in performance, some of which were comparable in magnitude with the effects of experimenter-guided ability training. However, when compared with effects of ability training, self-guided retest gains involved more error and a smaller gain on the more difficult items. A deficiency of this earlier research was that the amount of time in practice sessions was not equivalent to time spent in the ability training programs. Therefore, Experiment 1 was designed so that practice time would be constant for retest practice and ability training.

Method

Design and procedure. Subjects ($n = 87$) were assigned randomly to three experimental conditions: (a) a no-contact control group, (b) a retest-practice group involving three sessions totaling about 10 hr, and (c) an experimenter-guided ability training group, involving 10

sessions also totaling about 10 hr. In addition, a pretest and a posttest were given to all groups. The test batteries used as pretest and posttest, and also used as the retest program for the retest group, were identical and included all tests listed in Table 1.

Training programs. Retest practice, or self-guided retesting, involved three repetitions of the same psychometric test battery of intelligence (see Table 1) under standardized instructions and without feedback. After the first session, the time taken for warm-up examples was reduced because subjects reported that further and complete repetitions of the general and subtest-specific testing instructions were not necessary. Subjects were not informed about their test results. They were also asked not to communicate with their companion research participants about the testing program.

Ability training consisted of experimenter-guided instruction and practice in the content, rules, and problem-solving skills associated with two subabilities of fluid intelligence: figural relations (Willis et al., 1981) and induction (Blieszner et al., 1981). Because of the concern with the teaching of specific rules and problem-solving strategies in the ability training program, we expected that this form of training would enhance the likelihood that the subjects would acquire "new" cognitive skills.

As described in more detail in earlier publications (Baltes & Willis, 1982; Blieszner et al., 1981; Willis et al., 1981), the ability training programs were developed through task analysis of the items and rules contained in tests of figural relations and induction (see Table 1). The rules most frequently required for test solution were identified and combined into a training program. No items used in training, however, were identical to items making up the standard tests. The training program, conducted by an instructor in small groups of 6–12 participants, focused on helping subjects to identify rules and concepts and use them in solving the types of items that constituted the psychometric ability factors of figural relations and induction. Feedback about correct and wrong applications of the rules was given in a supportive manner.

Data analysis. To obtain comparability with previous data, we transformed the data to a common metric for all tests ($M = 50$, $SD = 10$), using the standardization provided by the control group at pretest in Baltes et al.'s (1986) study. We computed difference scores by subtracting pre- from posttest scores, and they were used as dependent variables. There is a controversy in the literature over the use of such difference scores as indicators of change (Cronbach & Furby, 1970; Horn, 1963). However, Rogosa, Brandt, and Zimowski (1982) argued that the use of difference scores is acceptable if reliability (precision) of measurement and the variance of intraindividual change are sufficient. In our view, these conditions apply to our data set, as is evident from the magnitude of the stability coefficients reported in Table 2.

Difference scores were analyzed with a 3 × 8 (Experimental Condition × Subtest of Intelligence mixed-model analysis of variance (ANOVA) and with post hoc tests where indicated. The ANOVA yielded significant main effects for group, $F(2, 84) = 30.6$, $MS_E = 56.7$, and for test, $F(7, 588) = 24.1$, $MS_E = 40.1$, and a significant Group × Test interaction, $F(14, 588) = 5.7$ (all $ps < .01$). When analyses of covariance were performed (with pretest scores as covariates), the same pattern of outcomes was obtained.

Results

Overall effects of retest practice and ability training. Data of Experiment 1 are summarized in Table 2 and Figure 1. First, both experimental conditions, ability training and self-guided retest practice, produced significant increases in performance. As expected, however, effects were not equal for all tests of transfer. For the ability-training condition, training gains were restricted to the four tests of near–near transfer (see Table 1) that mark the two trained ability factors of fluid intelligence: induction and figural relations. For the retest practice condition, gains were observed not only on these four tests but also on two additional tests indexing perceptual speed. Transfer of training gains was further analyzed in a post hoc ANOVA in which we contrasted self-guided retesting and ability training on the four fluid tests; neither the main effect nor the interaction with group was significant.

In general, the ability-training findings replicated earlier results with the same training program and similar-age participants reported by Baltes et al. (1986). The scope and order of transfer of training for the ability training group were in accord with our theoretical expectations. This effect included transfer to a set of test items not included in the training materials. By contrast, for the retest practice group, the training gains were based on the same test items included in the practice sessions.

Level of difficulty and accuracy. The second question addressed in Experiment 1 was whether the performance gains due to a comparable amount of self-guided retest practice and ability training (i.e., 10 hr) show differences in the qualitative nature of performance. The comparison was aided by the fact that self-guided retest practice and ability training produced similar gains in the critical transfer tests of the battery (Subtests 1–4 in Figure 1). Level of difficulty and accuracy were selected as critical features of these analyses.

Table 2
Pretest and Posttest Means and Standard Deviations: Experiment 1

	Ability training				Retesting				No-contact control				Pretest–posttest r across group
	Pretest		Posttest		Pretest		Posttest		Pretest		Posttest		
Measure	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
ADEPT Induction	52.0	11.8	65.6	11.8	53.1	10.8	68.2	14.7	57.3	12.7	61.4	12.3	.79
ADEPT Figural Relations	51.8	9.0	62.2	10.4	52.2	9.1	63.5	8.8	52.5	10.0	57.7	9.9	.77
Induction Standard	51.9	9.8	60.7	10.1	52.5	10.4	61.5	10.9	54.5	11.4	58.4	12.0	.87
Culture Fair Test	50.9	10.3	57.5	8.5	50.8	11.5	60.2	10.3	53.6	9.7	57.7	9.0	.76
Raven's Matrices	51.4	9.4	53.3	10.1	52.9	9.1	55.6	10.3	54.1	10.5	54.5	10.8	.83
Perceptual Speed I (NC)	52.1	7.9	52.2	9.0	50.1	9.3	57.1	13.6	54.0	9.3	58.6	10.6	.74
Perceptual Speed II (IP)	50.3	9.5	52.6	9.5	50.1	13.6	63.2	9.5	56.1	9.0	57.9	10.1	.65
Vocabulary	51.4	11.3	52.1	10.1	56.0	7.0	57.8	7.6	54.6	8.3	56.1	8.9	.76

Note. ADEPT = Adult Development and Enrichment Project; NC = Number Comparison Test P-2; IP = Identical Pictures Test P-3.
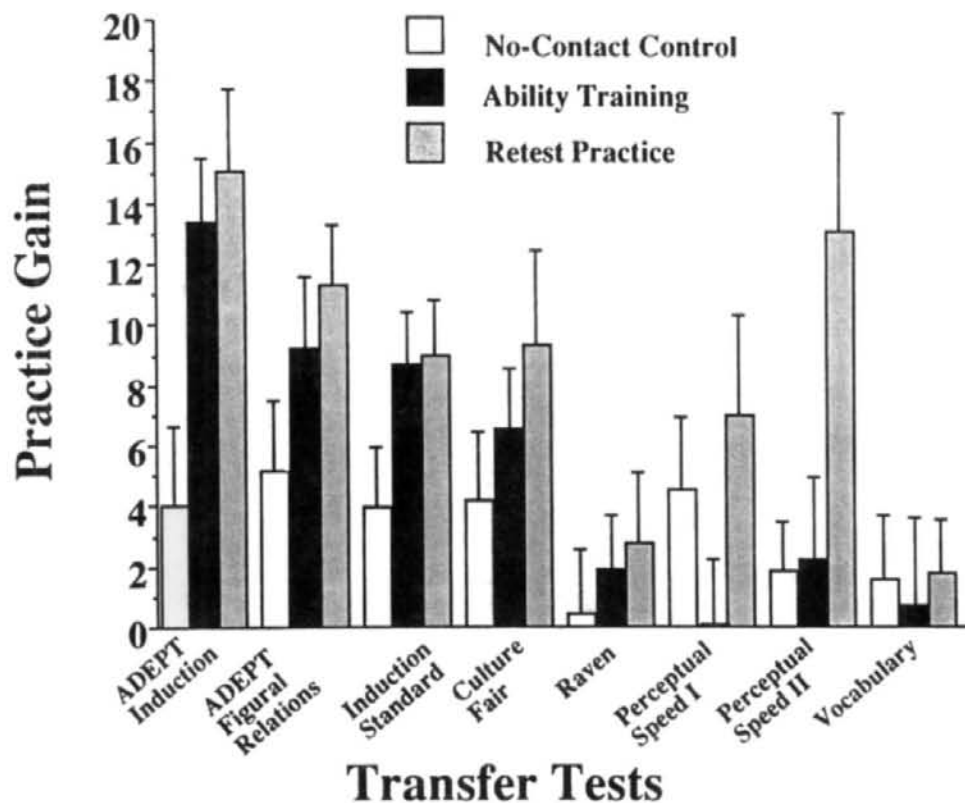
Figure 1. Pattern of transfer of training gains in standardized scores. (Bars indicate the 95% confidence interval. ADEPT = Adult Development and Enrichment Project.)

The difficulty level of items was estimated from the performance of an independent sample of comparable older adults who were asked to solve the items of Subtests 1–4 under a power condition of assessment (see Experiment 2 for further details). On the basis of subjects' performance in this power condition of assessment, items in each subtest were classified in three levels of difficulty: easy (67% or more correct solutions), medium (34–66%), and difficult (33% or fewer). This item classification by level of difficulty was then applied to the data set of Experiment 1.

Accuracy was defined as the proportion of correct answers within the total of attempted answers. For this purpose, we classified all marked responses on the test sheets as correct or wrong answers (the typical scoring of the psychometric tests used is based on the number of correct answers only). The third possible category of scoring, unanswered items, was not considered in the statistical analysis because of its ipsative relation to the other two categories. Table 3 (which also includes data from Experiment 2) displays pretest and posttest means for the number of correct answers, the number of wrong answers, and the proportional index of accuracy for the retesting and ability training groups.

Two ANOVAs, one for correct answers and one for accuracy, were conducted on difference scores from pretest to posttest. In these ANOVAs, factors included group, test, and difficulty. The three levels of difficulty were specified as nested within the four tests because levels of absolute difficulty and number

of items were not comparable across the four tests. The two ANOVAs, of course, did not yield independent information. Nevertheless, we hoped for more detailed information about the locus of training gains and the effect of training on response strategies.

In the analysis of correct answers, there were significant effects of type of test, $F(3, 168) = 22.3$, $MS_E = 10.0$, and difficulty within test, $F(8, 448) = 7.0$, $MS_E = 6.0$, and there was an interaction between group and difficulty within test, $F(8, 448) = 2.70$, $MS_E = 6.0$ (all $p < .05$). Four post hoc ANOVAs, one for each test, were conducted to locate the source of variance of the interaction (α was set at .0125). Significant interactions between group and difficulty were obtained for two of the four tests: Adult Development and Enrichment Project (ADEPT) Induction and Culture Fair. These interactions are displayed in Figure 2. In both tests, self-guided retest practice produced an added improvement in correct answers for items with medium difficulty. For the remaining two of the four near-transfer tests (see Table 1) the group effect did not vary by level of item difficulty.

In the analysis of the accuracy index, there were significant effects for the interaction between group and test, $F(3, 168) = 3.6$, $p < .02$, $MS_E = .03$, and difficulty within test, $F(8, 448) = 1.9$ $p < .01$, $MS_E = .02$. The interaction between group and difficulty within test approached significance ($p < .07$). To identify the source of variance for the Group $\times$ Test interaction, we conducted again four post hoc ANOVAs, one

Table 3

Comparison of Performances on Five Subtests: Training Versus Self-Guided Retesting Versus Power Condition of Assessment Without Training

Condition	ADEPT Induction (70)			Induction Standard (70)			ADEPT Figural Relations (42)			Culture Fair Test (72)			Raven's Matrices (18)		
	Correct	Wrong	Accurate	Correct	Wrong	Accurate	Correct	Wrong	Accurate	Correct	Wrong	Accurate	Correct	Wrong	Accurate
						Experiment 1: Ability training									
Pretest															
M	23.4	14.9	.62	29.9	12.1	.70	23.8	11.4	.68	36.7	24.9	.59	7.4	8.5	.47
SD	11.9	11.5	.22	12.4	10.1	.22	6.2	5.9	.16	9.8	8.9	.15	3.8	4.1	.22
Posttest															
M	37.1	7.1	.83	41.1	6.6	.85	30.3	8.1	.79	43.0	20.3	.68	8.1	8.3	.50
SD	11.9	4.8	.13	12.8	5.7	.15	7.2	6.0	.16	8.1	7.2	.11	4.1	4.3	.24
						Experiment 1: Retest practice									
Pretest															
M	24.6	13.6	.65	30.7	11.1	.74	23.4	11.0	.68	36.6	25.8	.59	8.0	8.3	.49
SD	10.8	11.6	.21	13.1	10.7	.21	6.3	5.8	.16	11.0	10.9	.17	3.7	3.4	.21
Posttest															
M	39.7	11.0	.78	42.0	10.0	.81	31.3	8.5	.78	45.5	22.3	.67	9.1	8.6	.51
SD	14.7	10.8	.19	13.8	10.2	.19	6.1	5.5	.14	9.8	8.6	.14	4.2	4.0	.23
						Experiment 2: Power									
M	43.0	17.7	.71	45.6	16.8	.73	27.3	6.4	.81	47.5	23.2	.67	10.9	6.8	.61
SD	14.6	12.2	.19	14.5	12.2	.20	4.2	4.0	.12	9.7	9.4	.13	4.3	4.1	.24

Note. ADEPT = Adult Development and Enrichment Project. Numbers in parentheses are the number of items in each subtest.
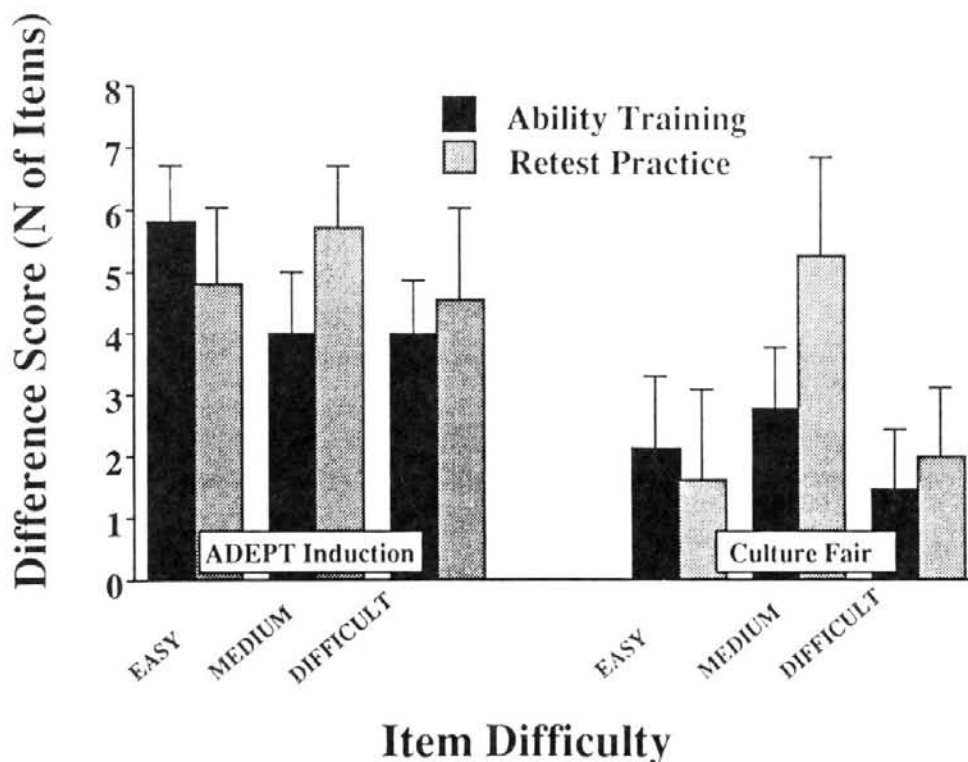
Figure 2. Differential performance gains in correct responses by level of difficulty for two of the near-transfer tests. (Bars indicate the 95% confidence interval.)

for each test, (α = .0125). As shown in Figure 3, significant differences between experimental groups were obtained only for ADEPT Induction and Induction Standard tests. In both tests, ability training resulted in a significant reduction in the degree of error (or increase in accuracy) of the problem-solving process in relation to that of the retest practice group.

Discussion

There are two important aspects to the findings of Experiment 1. First, the pattern of findings obtained for the ability-trained subjects in this experiment is highly similar to that of past work (Baltes et al., 1986; Baltes & Willis, 1982). Therefore, it provides a replication and baseline of comparison. The consistency in findings suggests that our additional analyses can be generalized to earlier training research.

Second, the results of Experiment 1 relate to the question of whether the unequal amount of self-guided retest practice and ability training administered in past research accounts for the superiority of ability training involving level of difficulty and accuracy obtained. The dominant finding of Experiment 1 is that a similar degree of training gain was achieved when amount of time (10 hr) for ability training and retest practice was held constant. There continued to be some, though moderate, advantage involving accuracy for the ability-trained group. In addition to the training gains in correct answers observed for both the ability-training and the retest-practice groups, the ability-training group made fewer errors

when solving items from two of the four transfer tests (Induction). However, when level of item difficulty is considered, the ability-trained group did not display an added advantage.

Are the cognitive skills that subjects learned critically dependent on external instruction, or is the evidence primarily one of activation of skills in the repertoire of the participating subjects? Because of the substantial degree of similarity in the results obtained for the retest-practice and the ability-trained groups, it seems most likely that the gains are a function of self-guided activation of already available skills. However, the finding that ability training produced some increase in accuracy on two of the four transfer tests (modest as it was) leaves open the possibility that an added skill associated with experimenter-guided training was also involved. Such a skill would have allowed the ability-trained subjects to solve items with higher accuracy (or fewer errors).

Experiment 2

Objective

Experiment 2 was conducted to examine further the level of performance that could be reached by older adults when they used their own cognitive repertoire in a self-directed manner. The aim was to identify a procedure that would permit an assessment of self-activated cognitive skills with as little practice and experimenter-guided feedback in the nature of problem solving as possible. The strategy chosen for Ex-
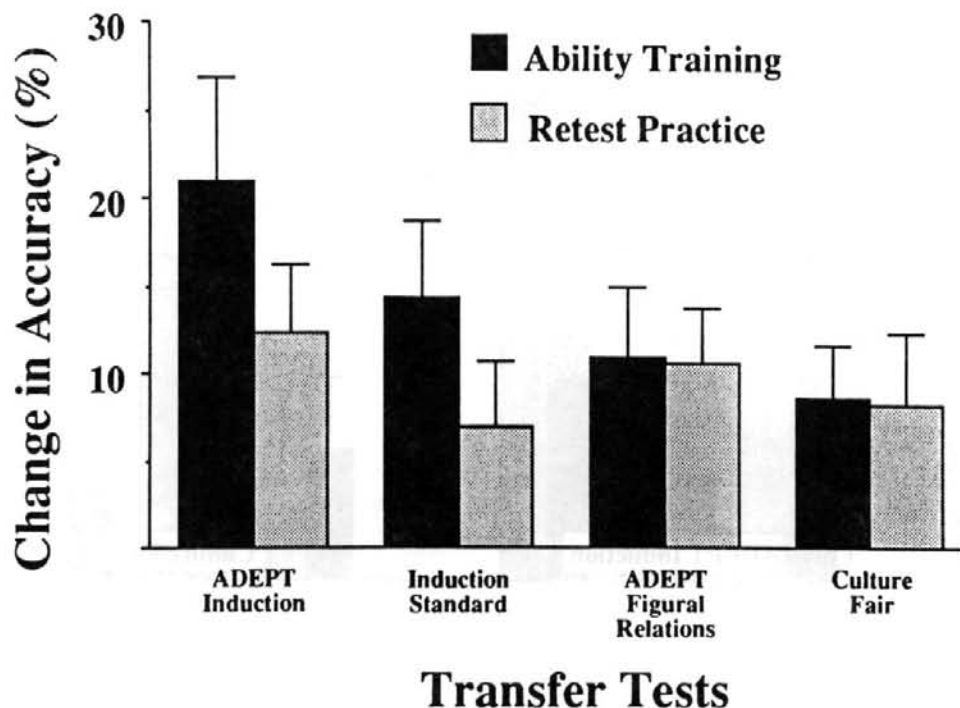
Figure 3. Significant changes in accuracy after ability training for two of the four near-transfer tests. (Bars indicate the 95% confidence interval. ADEPT = Adult Development and Enrichment Project.)

periment 2 was to construct a power condition of assessment involving shorter test length and more time for completion. Our prediction was that if untrained elderly adults had the cognitive skills available in their repertoire, they would show marked gains associated with shorter test length and extended time. If, on the other hand, there were limits to their cognitive repertoire, their performance under optimized conditions of assessment would not approach that of subjects who were trained by an instructor in test-relevant skills and rules of problem solving.

Method

The five near-transfer tests of the battery shown in Table 1 were used. These tests included the four tests shown to be affected by ability training: ADEPT Figural Relations, Standard Figural Relations, ADEPT Induction, Induction Standard. For these tests, significant gains due to ability-specific training had been reliably demonstrated in Experiment 1 and in previous research (e.g., Baltes et al., 1986; Willis, 1985). In addition, because of the gain observed in one earlier training study (Blieszner et al., 1981), the Raven's Matrices were also included.

Testing time was the standard time (prorated by number of items) multiplied by 4. In addition, test length was reduced by half, and items were given to separate subsamples in ascending and descending order to control for position effects. Specifically, the sample of 60 elderly persons described in the General Method section was assigned randomly to one of four conditions of time-extended assessment: (a) ascending versus descending order of items and (b) even versus odd halves.

Results

As in Experiment 1, we examined the number of correct answers and accuracy. Because each subject was administered

only half of the items (odd or even, respectively), their scores were multiplied by 2 to obtain data comparable with those of Experiment 1. We could not investigate level of difficulty because data from Experiments 1 and 2 were not comparable as a result of variations in the position of items within subtests.

Table 3 provides comparative information involving Experiments 1 and 2. The comparison permits contrasting non-trained elderly (under a quadruple-time-length, half-test-length power condition of assessment) with subjects who experienced 10 hr of either experimenter-guided ability training or self-guided retest practice. Because the samples involved in the two experiments were not drawn at the same time from the same parent population, it is difficult to conduct a decisive statistical test. The following ANOVAS, therefore, must be considered in view of the fact that the samples may not have been equivalent at the outset.

We conducted two ANOVAS. In the first we used the number of correct answers as dependent variables, and in the second we used the proportional index of accuracy. Each analysis was a 3 × 4 (Groups × Tests) ANOVA with two orthogonal contrasts specified on the group factor (Experiment 2 vs. Experiment 1 groups, and ability-training vs. retest-practice groups). Because the contrast between ability training and retesting is essentially a repeat of the analyses described in Experiment 1, we will discuss only the contrast in which we compared Experiment 1 and Experiment 2 groups. We applied univariate post hoc tests when necessary. The data for the Raven's Matrices were analyzed separately.

For correct answers (see Table 3), there was a significant interaction between group and test, $F(3, 345) = 9.1$, $MS_E = 45.4$, $p < .01$. The main effect of group was not significant.

The interaction indicated trends for ADEPT Induction ($p <$ .08), Induction Standard ($p <$ .12), and Culture Fair Figural Relations ($p <$ .06) in favor of the power condition, and a significant disadvantage of the power condition for ADEPT Figural Relations ($p <$ .01). A significant difference between Experiments 1 and 2 resulted for the Raven's Matrices, $F(1, 115) = 8.8$, $MS_E = 17.7$, $p < .01$, showing that subjects under a power condition of assessment achieved a larger number of correct answers than did subjects who participated in the two 10-hr practice programs.

In the analysis of accuracy scores (see Table 3), we modulated the findings on correct answers and showed an advantage of the two practice groups that were part of Experiment 1. The statistical contrast between Experiments 1 and 2 produced a significant main effect of group, $F(1, 115) = 2.9$, $MS_E = .08$, $p < .05$. The interaction between group and test, $F(3, 345) = 14.9$, $MS_E = .01$, $p < .01$, suggested that the power condition produced less accuracy on ADEPT Induction, $t(116) = 3.3$, $p < .01$, and on the Induction Standard test, $t(116) = 3.3$, $p < .01$. A significant difference between Experiments 1 and 2 was also obtained for the Raven test, $F(1, 115) = 6.3$ $MS_E = .05$, $p < .03$. In contrast to the results for the induction subtests, subjects had higher accuracy on the Raven test under the power condition of assessment than after ability training or retest practice.

Discussion

The results of Experiment 2 suggest that simply extending the time available for solution of items and reducing test length by half make it possible for subjects to show the same amount of performance enhancement as the two practice groups when the number of correct answers is the focus of analysis. In fact, the level of performance produced by elderly subjects in the quadruple-time-length, half-test-length power condition tends to be somewhat higher (with the exception of the ADEPT Figural Relations test) than the performance achieved after 10 hr of ability training and self-guided retest practice. This finding suggests—as was true for Experiment 1—that the cognitive skills necessary for finding correct solutions to problems are in the repertoire of older subjects. The information on accuracy is complex. For two of the five subtests (the two induction tests), there was a parallel increase in the number of wrong answers for the power condition. The opposite, however, was true for the Raven test.

General Discussion

Are cognitive training effects in research on fluid intelligence with older adults primarily the result of self-directed activation (or learning) of existing cognitive skills, or are they the result of experimenter-guided learning of new problem-solving skills? Our primary approach was to speak of the activation of skills available in the subjects' preexperimental repertoire, if subjects without experimenter-guided feedback were able to show levels of performance comparable with that of subjects trained in test-relevant cognitive skills and rules of problem solving. Two conditions without experimenter-provided feedback about the required skills and rules, and about the correctness of performance during practice, were studied:

self-guided retest practice and a power condition of assessment (fewer items and quadruple time).

Our interpretation of the outcome pattern is that performance gains obtained in past cognitive training research with psychometric measures of fluid intelligence in the elderly were possible, in principle, because of the activation and practice of existing cognitive skills. Conversely, the gains do not appear to be critically dependent on experimenter-based instruction in specific skills and rules of problem solution. Activation of available skills is demonstrated by the marked performance gains after self-guided retest practice and in power conditions of assessment, and by the fact that ability-trained subjects did not solve more difficult items than retest-practice subjects. The high levels of performance reached after 10 hours of self-guided retest practice under standard time limits of test administration, in terms of correct answers for all levels of difficulty, were surprising to us (see also Horn, Donaldson, & Engstrom, 1981).

There is one exception to this general interpretation. When accuracy of performance is considered, ability-trained subjects made fewer errors on the two induction tests of the transfer battery. This was true for the comparison between ability training and retest practice, as well as for the comparison between ability training and a power condition of assessment. The latter, however, was based on the post hoc analysis of a marginally significant trend. Within the overall spectrum of performance gains, however, this advantage of the ability-trained subjects is fairly small. Moreover, because this advantage of an ability-trained group has not been obtained in a subsequent study (Baltes, Sowarka, & Kliegl, in press), we tend to view this finding with caution.

We do not know, of course, whether the same cognitive processes, associated with having either more time available or with retest practice, were operative during experimenter-guided ability training. Relevant observations of the cognitive skills used by subjects during ability training and retest practice are unavailable. All we do know is that elderly subjects without external instruction reached similar performance levels as a function of self-guided retest practice and under a power condition of assessment. These findings suggest that appropriate cognitive skills are available in their repertoire.

What are the implications of our findings, and what is the interpretative posture for past cognitive training research with tests of fluid intelligence? The results lend support to the general position that training gains are relatively easy to obtain because elderly subjects have available in their repertoire a good base of required skills. What seems primarily important is to offer an opportunity for practice. This finding is in accord with the results of Schaie and Willis (1986), who reported success in training older adults on tests of fluid intelligence to the same level of performance that they had shown earlier in their adult lives.

In addition, the findings have significant implications for educational questions. Educators are much concerned about whether older adults are able to participate in programs that are based on self-directed learning (Dittmann-Kohli, 1981, 1983). On the basis of our results, it seems clear that many elderly adults are quite capable of reaching high levels of cognitive functioning in educational contexts entailing the use of self-directed learning strategies. This inference is partic-

ularly true if one accepts the conclusion (Hooper, Hooper, & Colbert, 1985) that the difficulty of the measures of fluid intelligence that we used display a fair degree of similarity with other cognitive measures (such as Piagetian ones) designed to index high levels of abstract reasoning.

To prevent possible misinterpretation of our research on the plasticity of fluid intelligence in old age, some cautionary comments are necessary. The results do *not* imply the following two conclusions. The first is that our findings bear directly on the issue of whether there is age-related loss in the latent abilities underlying fluid intelligence. Our demonstration of a substantial reserve capacity within the elderly population does not imply that there is no such loss. Such a conclusion would require age-comparative research showing that older adults could reach the same final level of performance as young adults (Baltes & Kliegl, 1986; Denney, 1984; Horn, 1979; Horn et al., 1981; Kliegl & Baltes, 1987). Because age-comparative studies on training gain and the upper range of possible levels of performance are unavailable, the extent of reserve capacity available in different age groups is unknown. The second possible misinterpretation is that all aging individuals display the reserve capacity reported here. We do not know the extent to which our findings are generalizable to other elderly adults. Although we believe that many elderly individuals up to the age studied are capable of major training benefits, as reported here, we need to emphasize that our samples were clearly healthy. Most likely, the results might not generalize to persons who suffer from manifest brain-related diseases such as senile dementia of the Alzheimer type.

References

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., & 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., & Kliegl, R. (1986). On the dynamics between growth and decline in the aging of intelligence and memory. In K. Poeck, H. J. Freund, & H. Gänshirt (Eds.), *Neurology* (pp. 1–17). Heidelberg: Springer Verlag.

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.*

Baltes, P. B., & Willis, S. L. (1982). Plasticity and enhancement of intellectual functioning in old age: Penn State's Adult Development and Enrichment Project (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. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

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). *Test of "g": Culture Fair (Scale 3, Form B).* Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R. B., & Cattell, A. K. S. (1963). *Test of "g": Culture Fair (Scale 3, Form A).* Champaign, IL: Institute for Personality and Ability Testing.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74*, 68–80.

Denney, N. W. (1984). A model of cognitive development across the life span. *Developmental Review, 4*, 171–191.

Dittmann-Kohli, F. (1981). Learning how to learn: A psychological approach to self-directed learning. *Education, 24*, 23–33.

Dittmann-Kohli, F. (1983). Intelligenzforschung im höheren Erwachsenenalter [Cognitive training research in old age]. *Unterrichtswissenschaft, 4*, 361–369.

Ekstrom, R. B., French, J. W., Harman, J., & Derman, D. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Heller, K., Gaedicke, A. L., & Weinläder, H. (1976). *Kognitiver Fähigkeitstest (Form A).* Weinheim, FRG: Beltz Test GmbH.

Hooper, F. H., Hooper, J. O., & Colbert, K. K. (1985). Personality and memory correlates of intellectual functioning in adulthood: Piagetian and psychometric assessments. *Human Development, 28*, 101–107.

Horn, J. L. (1963). Equations representing combinations of components in scoring psychological variables. *Acta Psychologica, 21*, 184–217.

Horn, J. L. (1979). Some correctable defects in research on intelligence. *Intelligence, 3*, 307–322.

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.

Horn, J. L., Donaldson, G., & Engstrom, R. (1981). Apprehension, memory, and fluid intelligence decline in adulthood. *Research on Aging, 3*, 33–84.

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. 95–117). 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.

Overton, W. F., & Newman, J. L. (1982). Cognitive development: A competence-activation/utilization approach. In T. Field, A. Houston, H. Quay, L. Troll, & G. Finley (Eds.), *Review of human development* (pp. 217–241). 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: Lewis & Company Ltd.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*, 726–748.

Salthouse, T. A. (1985). *A theory of cognitive aging.* Amsterdam: North Holland.

Schaie, K. W., & Willis, S. L. (1986). Can adult intellectual decline be reversed? *Developmental Psychology, 22*, 223–232.

Thurstone, T. G. (1962). *Primary Mental Abilities, Grades 9–12, 1962 revision.* Chicago: Science Research Associates.

Willis, S. L. (1985). Towards an educational psychology of the older adult learner: Intellectual and cognitive bases. 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.