JOURNAL OF APPLIED DEVELOPMENTAL PSYCHOLOGY 2, 247–265 (1981)

Training Research in Aging on the Fluid Ability of Inductive Reasoning*

ROSEMARY BLIESZNER
SHERRY L. WILLIS
PAUL B. BALTES

College of Human Development
The Pennsylvania State University

The modifiability of older adults' performance on Induction tasks was examined through cognitive training. The posttraining performance of 52 older adults ($\overline{X}$ age = 70.3 years) was assessed with regard to a transfer-of-training paradigm and maintenance of training effects at three posttests (one week, one month, six months). A pattern of differential training transfer across the posttest battery of fluid and crystallized intelligence measures was predicted, with the largest training effects expected for near transfer (Induction) measures. The predicted pattern of training transfer was obtained at one-week and one-month posttests, with significant training effects to the nearest Induction measure. Large retest (practice) effects across posttests also occurred for training and control groups. This research contributes to the position that modifiability of intellectual performance through cognitive intervention extends across the adult life span.

INTRODUCTION

Much research on adult intellectual functioning has focused on normative levels of performance (Botwinick, 1977; Horn, 1978; Matarazzo, 1972). Historically,

*This study was conducted as part of a research program entitled the Adult Development and Enrichment Project (ADEPT), supported by a grant from the National Institute on Aging (No. 5-R01-AG00403) to Paul B. Baltes and Sherry L. Willis, co-investigators. ADEPT examined the effect of cognitive training on intellectual performance of older adults. Thanks are due to several research assistants (Steven Cornelius, Margie E. Lachman, Vincent Morello, Gail Peck, Brian Hofland, and Manfred Schmitt), to its field and training staff (Carolyn Nesselroade, and Myrtle Williams), and to statistical consultants (John R. Nesselroade, and Paul A. Games). Rosemary Blieszner was supported in part by a NIA Predoctoral Traineeship (T32-AG0048).

Address requests for reprints to either Sherry L. Willis, College of Human Development, The Pennsylvania State University, University Park, PA. 16802, or Paul B. Baltes, Max-Planck Institute of Human Development and Education, 94 Lentzeallee, 1 Berlin 33, West Germany.

this descriptive approach has been related to a view of intellectual development assuming a peak in functioning in young adulthood, relative stability through middle age, and widespread decline in old age. Moreover, just as intellectual development in childhood has frequently been associated with maturational factors, the normative pattern of decline in old age has often been linked with age-related decrement in sensory and neurophysiological functioning. This perspective, emphasizing early intellectual development and physiologically-based decline has not fostered an experimental, interventive approach to the study of intellectual potential in later adulthood (Baltes & Willis, 1979).

In contrast, intelligence research in childhood had examined both normative and experimental perspectives (Glaser, 1978; Resnick, 1976; Sternberg & Detterman, 1979). The predominant approach has continued to be descriptive, defining the normative or average course of intellectual development in childhood. However, this focus has been supplemented with experimental studies examining the range of modifiability in intellectual functioning achieved through intervention (Brown, 1981; Kuhn, 1974). Training research has addressed not only the timing or rate (acceleration) of intellectual development but also developmental antecedents and cognitive processes associated with intellectual ontogeny.

Such an experimental approach reflects the growing recognition of experiential as well as maturational factors in intellectual development in childhood. Those involved in cognitive training research in childhood have suggested that a normative approach may underestimate intellectual potential in disadvantaged populations, involving environmental deficits or mild retardation (Brown & French, 1979). Traditional methods of assessing intelligence were designed primarily as predictive tools. They predicted future functioning based on prior development and learning, assessed primarily within educational contexts. In contrast, an experimental approach seeks to assess potential, rather than a current level of functioning, by examining children's performance under varying amounts of instruction or practice (Brown & French, 1979; Carlson & Wiedl, 1978). This method may be particularly useful when the level of prior learning or development does not provide an adequate index of learning potential, as with disadvantaged or retarded populations.

The recent emphasis on a life-span perspective suggests the need for a similar approach to intellectual development and assessment in adulthood (Baltes & Willis, 1979, 1981). A life-span view emphasizing continued intellectual development throughout adulthood was supported initially by longitudinal research involving cohort-sequential methodologies, as exemplified in the work of Schaie (Schaie, 1979). This research indicated that a peak in intellectual functioning for current cohorts of healthy, well educated adults may not occur until early middle age, and that decrement in old age occurred later and was less pervasive than suggested in prior cross-sectional studies. Moreover, cohort differences indicated that for many abilities, earlier cohorts functioned at a lower level than later cohorts at the same chronological age. These cohort differences may indicate experiential (educational) and environmental deficits, rather than a primary decline in ability (potential) per se. Current elderly generations may be experiencing cohort-related obsolescence, evidenced by lower intelligence test performance.

Such longitudinal findings suggest the need for experimental research that focuses on learning potential in adulthood, similar to that conducted with disadvantaged children. Examination of the range of modifiability in adult intellectual functioning provides information on intraindividual variability to complement existing knowledge about normative age changes. In addition, such research can provide a knowledge base for the development of strategies for educational intervention.

A number of earlier gerontological training studies on a variety of intellectual abilities suggested that intellectual performance can be modified even in late adulthood (Denney, 1979; Labouvie-Vief, 1976; Willis & Schaie, 1981). However, careful examination of the learning potential hypothesis suggests that training effects need to be assessed within a design provided by a theoretical framework. Such a framework defines the target ability for training, conceptual design of the training program, and criteria for assessment of training and transfer effects. Within the present study, the Cattell-Horn theory of fluid and crystallized intelligence (Cattell, 1971; Horn, 1978; Horn & Cattell, 1967) provided such a framework. The target ability domain for training was fluid intelligence. Fluid and crystallized intelligence are two broad, second-order intellectual dimensions, postulated by the theory to show differential, normative patterns of development in adulthood. Fluid intelligence exhibits a gradual decline beginning in early adulthood, while crystallized intelligence is said to increase or remain stable through much of adulthood. On the first-order level, fluid intelligence is indexed by primary abilities such as figural relations, induction, and memory span (Horn, 1970).

The present training research focused on the fluid ability of Induction. Training research on Induction has not been extensively investigated in adulthood, although it has been studied in childhood (Holzman, Glaser, & Pellegrino, 1976; Kotovsky & Simon, 1973). One study (Labouvie-Vief & Gonda, 1976) reported improvement in the older adults' performance on letter sets, which is one type of Induction problem; training effects were maintained over a two-week period. The present research provided a more extensive examination of Induction training in later adulthood, with regard to the array of Induction tasks trained, the breadth of training transfer, and temporal maintenance effects examined. Training transfer was examined for a broad array of fluid and crystallized measures. Maintenance of training effects were evaluated across three posttests over a six-month period.

METHOD

Subjects

Subjects were older adults from rural communities in central Pennsylvania. They were recruited from local organizations and paid ($2.00 per hour) for participation individually or by contribution to their organizations. Subjects reported good health, with no major auditory or visual defects.

Fifty-six subjects (Training: N = 30, Control: N = 26) completed the study through the six-month posttest. Data on four randomly selected training subjects were dropped to achieve equal numbers of subjects in training (N = 26) and control (N = 26). Mean age of these 52 subjects (M = 10; F = 42) was 70.32 years (SD = 6.09, Range = 60-85). Average educational level was 11.53 years (SD = 2.46, Range 6-17). There were no significant differences between training and control groups on age, educational level, or pretest scores.

Dropouts. Some subject attrition occurred, given the length of the study. Table 1 presents a summary of attrition across study intervals. A larger number of subjects was initially assigned at random to training, since a larger dropout was expected for training, given the greater time commitment.

Two dropout issues were examined. First, subjects completing the study (six-month posttest) were compared with the total sample pretested (N = 103). No significant differences were found between study participants and dropouts (N = 47) in age, education, or pretest scores. Also, training (N = 28) and control (N = 19) dropouts did not differ in age, education, or pretest scores.

TABLE 1
Number and Percent of Subjects Participating at Study Occasions[a]

Occasion	Group: Training		Group: Control	
	n	%	*n*	%
Pretest	58	100	45	100
Training Session 1	44	75		
Training Session 2	43	74		
Training Session 3	42	72		
Training Session 4	42	72		
Training Session 5	42	72		
Posttest 1 (1 wk.)	41	71	36	80
Posttest 2 (1 mo.)	40	70	34	76
Posttest 3 (6 mo.)	30	52	26	58

[a]Because attrition was expected to be larger in the training group, a larger sample of training participants was formed. Dropout rate of the Control group was assessed at only Pre- and Posttest occasions. Attrition occurred primarily after Pretest and between Posttests 2 and 3.

Second, due to attrition between the one-month and six-month posttests, dropouts (N = 18; T = 10; C = 8) between Posttests 2 and 3 were compared with subjects completing the study. No significant differences were found in age, education, or pretest scores. However, when Posttest 2 scores on Induction tests (ADEPT Induction, Induction Composite) of dropouts and participants were compared, the average performance of dropouts across treatment conditions was significantly lower [F (1,65) = 11.57, p = .001]. The interaction between treatment and dropout status, however, was not significant.

Design and Procedure

Subjects completed a pretest session (1.5 hours) and were randomly assigned to one of two conditions: Induction Training or Control. Training, beginning within one week of pretest, involved five one-hour sessions scheduled over two weeks. Control subjects received no training. Three posttests (3 hours each) were administered to all subjects, approximately one week, one month, and six months after training. Experimental and control subjects were pre- and posttested in groups of 8 to 16 persons. The pretest battery consisted of the ADEPT Induction test (Form A), and the ADEPT Figural Relations test (Form A; Plemons et al., 1978). The posttest battery of fluid and crystallized intelligence measures, including alternate forms of pretest measures, is described below.

Development of Induction Training Materials and Transfer Measure

The target fluid ability for training was Induction, one of the more "pure" primary abilities representing fluid intelligence (Cattell, 1971; Horn, 1970, 1978). Induction has been defined as the eduction of relationships in reasoning tasks that do not have semantic content. Induction ability is measured, for example, by tasks involving identification of the next letter or number in a series.

Content of the training program was based on a task analysis of three published tests previously shown (Cattell, 1971) to mark the Induction ability: Letter Sets (Ekstrom, French, Harman, & Derman, 1979), and Number Series and Letter Series (Thurstone, 1962). Task analyses identified relational rules (pattern descriptions) involved in problem solutions. Relational rules identified for the Letter Series test included those studied by Simon and Kotovsky (1963; Kotovsky & Simon, 1973). Additional rules (letter skips) were also identified for some items of the Thurstone Letter Series test. Relational rules identified for the Thurstone Number Series test were similar to those examined by others (Restle & Brown, 1970; Bjork, 1968; Egan & Greeno, 1974). Relational rules identified for letter sets problems were similar to those (letter repeats, alphabetical skips, and sequences) reported by Labouvie-Vief and Gonda (1976). Problems for the training program and items for the ADEPT Induction test were developed, based on the most frequently occurring relational rules identified in task analyses. In comparison to the published tests, the test items and training problems constructed were only similar in relational rules used, not in specific content.

ADEPT Induction Test. The ADEPT Induction test (Forms A & B) was developed as a direct assessment of the training program. The three published Induction tests and a pool of items generated on the basis of the task analyses were administered to 46 undergraduate students. Selection of items for inclusion in the two forms of the ADEPT Induction test were based on two criteria: Item difficulty level, ranging from 25% to 100%; and item-total correlation, ranging from .20 to .69. Items meeting these criteria were ordered by difficulty and assigned alternately to Forms A or B. Within each form, problems were then randomly ordered by difficulty level. Each form included: 30 letter sets items, 20 number series items, and 20 letter series items. Based on the college student data, the coefficient alpha for Form A is .91, and for Form B is .92. Correlation of Forms A and B is .91. In an independent study of 109 older adults aged 60–89, the correlation of Form A with a composite of the three published Induction tests was .81 (Baltes, Cornelius, Spiro, Nesselroade, & Willis, 1980).

Training Program. Five one-hour training sessions were conducted with groups of three to eight subjects by a female trainer. Each of the first four sessions included a review of the material from the previous session, individual practice on new problems, feedback from the trainer, and group discussion of solution strategies. Training tasks included letter sets, number series, and letter series problems. The fifth session involved a review of all types of problems. An adaptation of verbal rule learning shown effective in both problem solving (Wittrock, 1963) and Piagetian (Beilin, 1976) training research was the primary instructional method used. Instructional materials included a trainer's manual and individual practice booklets for each of the sessions.[1]

Assessment of Training Effects

Two major criteria were used in the assessment of training effects. The first focused on the transfer of training effects across the posttest battery of fluid and crystallized intelligence measures. The second criterion focused on maintenance of training and transfer effects across three posttests, spanning six months.

Training Transfer: Near and Far Transfer Levels. Training and transfer effects were assessed at each posttest occasion with a broad battery of fluid and crystallized intelligence measures, shown in Table 2. Transfer effects were predicted to follow a hierarchical, theory-derived pattern, based on the Cattell-Horn theory of intelligence (i.e., empirical factor loadings), and the degree of similarity of transfer measures to the training program.

[1]A sampler of training materials is available.

TABLE 2
Posttest Assessment Battery [a]

General Intelligence Dimensions	Primary Mental Ability	Predicted Transfer Pattern and Marker Measures	Source
FLUID	Induction	Nearest Transfer:	
		ADEPT Induction Test, Form B[b]	Present Study
		Near Transfer:	
		Induction Composite Test:	
		Letter Sets	Ekstrom et al., 1976
		Number Series	Thurstone, 1962
		Letter Series	Thurstone, 1962
FLUID	Figural Relations	Far Fluid Transfer	
		ADEPT Figural Relations Test, Form B[b]	Plemons et al., 1978
		Far Fluid Transfer:	
		Culture Fair Test (Scale 2, Form A) and Power Matricies [Scale 3, Form A (1963 ed.) & Form B (1961 ed.)]	Cattell & Cattell, 1957; Cattell & Cattell, 1961, 1963
		Far Fluid Transfer:	
		Raven's Advanced Progressive Matrices (Set II)	Raven, 1962
SPEED	Perceptual Speed	Far Non-Fluid Transfer:	
		Number Comparisons	Ekstrom et al., 1976
CRYSTALLIZED	Verbal Comprehension	Far Non-Fluid Transfer:	
		Vocabulary, V-2	Ekstrom et al., 1976

[a] Tests are listed in descending order of predicted magnitude of training transfer effects. All measures were included at Posttests 1, 2, 3.
[b] ADEPT Induction and Figural Relations Tests, Form A, were administered at Pretest.

Three levels of training transfer were examined. The first level focused on *near* transfer to the target fluid ability, Induction. The second and third levels assessed *far* transfer to the fluid and far nonfluid ability measures. With regard to near transfer, the ADEPT Induction test (Form B) was predicted to be the most direct measure of near transfer, since it included items representing only the relational rules presented in training. No test items were identical to those included in training. The second measure of near transfer was the Induction Composite test, consisting of the three published Induction tests: Letter Sets (Ekstrom et al., 1976), Number Series, and Letter Series (Thurstone, 1962). The Induction Composite included items representing both the trained relational rules and additional relational rules not involved in the training program. Inclusion of several tests of Induction permitted examination of a broad continuum of within-ability (Induction) transfer.

The second level of training transfer involved far-fluid transfer to Figural Relations, another relatively "pure" fluid ability factorially distinct from Induction at the primary ability level (Horn, 1970). Three Figural Relations measures were administered: the ADEPT Figural Relations test, Form B (Plemons et al., 1978); the Culture Fair test (Scale 2, and Power Matrices from Scale 3; Cattell & Cattell, 1957, 1961, 1963); and Raven's Advanced Progressive Matrices (Raven, 1962).[2]

The third level of training transfer involved far-non-fluid transfer to Crystallized Intelligence and Perceptual Speed. Within the fluid-crystallized theory, Crystallized Intelligence and Perceptual Speed are considered to have different developmental antecedents and to be factorially distinct from Fluid Intelligence. Crystallized Intelligence was represented by a Vocabulary test (V-2; Ekstrom et al., 1976), and Perceptual Speed by the Identical Pictures test (Ekstrom et al., 1976).

Maintenance of Training Effects. The second major assessment criterion examined maintenance of training and transfer effects across one-week, one-month, and six-month posttests. Temporal maintenance of such effects was critical, if training effects were interpreted to represent modification of the level of functioning on the target fluid ability. The immediate posttest, included in most previous training studies, was considered inadequate to assess stability of training effects. Thus, the assessment design involved three posttests across six months.

Retest Effects. Repeated posttesting, used in the assessment of training maintenance, was predicted to result in retest (practice) effects, particularly for a test-naive population such as the elderly. Practice effects have been demonstrated in prior training research (Plemons et al., 1978) and in research on retesting (Hofland, Willis, & Baltes, 1981) with the elderly. Retest effects, occurring as a function of practice in test taking per se, were hypothesized to be more general than the training effects and to be exhibited by *both* training and control subjects. Moreover, practice effects should occur for all or most of the posttest measures, without regard to the differential training transfer pattern. A differentiation of training and retest effects is important in an ability-specific interpretation of training.

RESULTS

Results of the study will be reported for training and transfer effects, temporal maintenance of training effects, and retest effects.

The entire data matrix (across occasions and treatments) for each of the

[2]Due to the length of the posttest battery, only odd-numbered items of Raven Advanced Progressive Matrices, Set II, were administered.

seven posttest measures was standardized, using the Control group's Posttest 1 performance as the standardization metric (X = 50; SD = 10). This procedure was employed for two reasons: It provided a common base line of performance on each measure to which all other data points for the same measure were compared. Second, it eliminated scale level differences between the measures, and thus facilitated examination of transfer across measures.

Training, Transfer, and Maintenance of Effects

Table 3 presents the means and standard deviations for the seven transfer measures at each posttest. A graphic summary of these mean scores for training and control is shown in Figure 1. The pattern of training transfer is represented by the relative *difference* between the mean of standardized scores for the training and control groups. The following outcomes are suggested by

TABLE 3
Mean Standardized Scores and Standard Deviations for Training and Control Groups by Posttest Occasion for Transfer Measures[a]

Measure	Posttest 1		Posttest 2		Posttest 3	
	Training	Control	Training	Control	Training	Control
ADEPT Induction						
$\bar{X}$	54.9	50.0	57.8	53.2	54.8	55.1
SD	14.5	10.0	13.3	12.1	14.2	11.6
Induction Composite						
$\bar{X}$	52.7	50.0	55.1	52.8	53.8	53.1
SD	14.0	10.0	14.2	12.0	14.7	13.0
ADEPT Figural Relations						
$\bar{X}$	52.0	50.0	55.7	54.4	53.9	53.1
SD	8.2	10.0	8.8	11.7	8.5	10.6
Culture Fair						
$\bar{X}$	49.2	50.0	52.6	52.4	51.9	52.6
SD	9.8	10.0	9.6	10.5	8.4	8.8
Raven						
$\bar{X}$	47.5	50.0	48.1	47.7	51.0	49.4
SD	7.4	10.0	5.6	8.5	9.1	8.0
Number Comparison						
$\bar{X}$	50.0	50.0	54.1	52.7	53.9	53.0
SD	8.1	10.0	10.3	9.5	9.6	10.0
Vocabulary						
$\bar{X}$	48.3	50.0	51.1	52.1	50.2	52.2
SD	10.1	10.0	8.7	10.0	7.5	10.0

[a]Scores on each measure were standardized across occasions and treatments with Control group's Posttest 1 performance as standardization metric ($\bar{X}$ = 50, SD = 10).
Training N = 26
Control N = 26

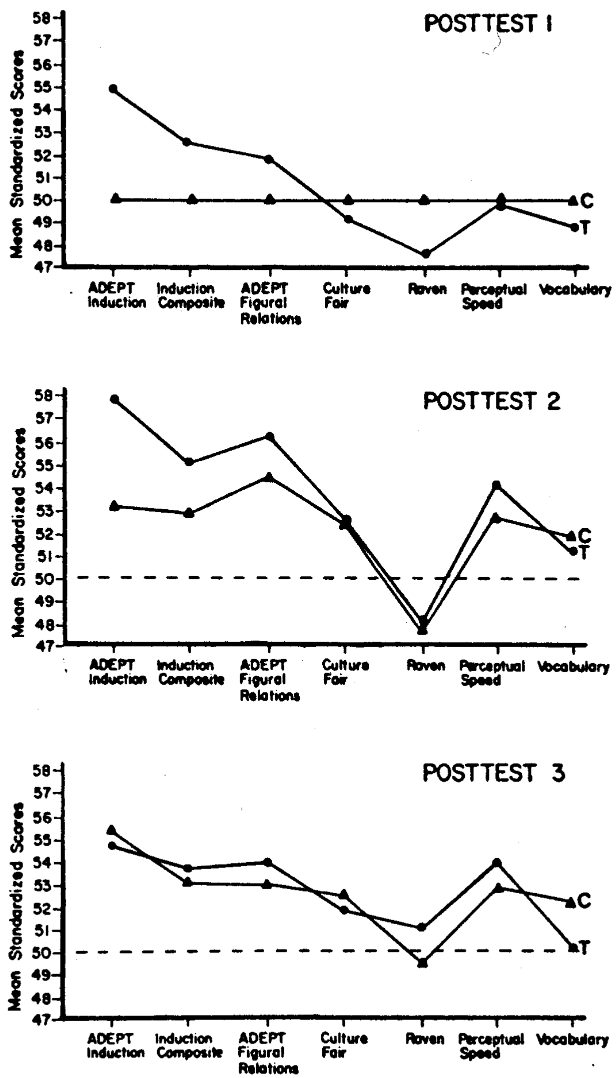

Figure 1. Mean Standardized Scores on Seven Transfer Measures for Training and Control Groups at Three Posttest Occasions. [a]

[a] All data points for each measure were standardized with reference to Control Group's Posttest 1 scores (X = 50, SD = 10).

Figure 1 and will be described in more detail in the report of statistical analyses. At Posttests 1 and 2, mean scores of the training group were larger than those of the control for the three nearest transfer measures, the two Induction measures (near transfer) and the closest far-fluid transfer measure, the ADEPT Figural Relations test. At the third posttest, however, there was no difference in the performance of the training and control groups on the transfer measures. In addition, large retest gains occurred for the control group across the three posttest occasions, for most transfer measures. At Posttest 3, there appears to be a small drop in the training group's mean score on the ADEPT Induction test.

To obtain an initial assessment of training effects, an overall 2 (Treatment: Training, Control) X 3 (Occasion: Posttests 1, 2, 3) X 7 (Measures), repeated measures analysis of covariance was conducted (Table 4). The covariate was the pretest score for the nearest transfer measure, ADEPT Induction test. Analysis of covariance was used to reduce the large individual differences within training and control groups, often found in aging research. This overall analysis resulted in a significant Occasion main effect, which was interpreted as primarily reflecting retest effects, common to both training and control groups. The significant Occasion by Measure interaction suggested differential effects across the posttest occasions resulting from gains and/or training decline. The significant Measure main effect occurred as a function of differential training and occasion effects, given the standardization procedure. In addition, there was a trend (p = .08) toward a significant Treatment X Occasion X Measure interaction.

Specific hypotheses, regarding training and transfer effects, were examined at two levels. First, the overall *pattern* of training and transfer effects was considered. Second, training effects were examined separately by *measure*.

Pattern of Training Effects. The pattern of training transfer and maintenance of this pattern was examined in a series of orthogonal planned comparisons, summarized in Table 5. In Table 4, results of these planned comparisons are shown within 2 X 3 X 7 analysis of covariance.

For the Treatment X Measure, and Treatment X Occasion X Measure interactions, differential weights were assigned to measures according to the predicted pattern of training transfer. The weights for each contrast were multiplied together and then applied to the appropriate means. First, in Comparison ψ_3 of the Treatment X Measure interaction, the two measures of near (Induction) transfer were weighted according to their conceptual relatedness to the training program, and compared with the five measures of far (fluid and non-fluid transfer, weighted equally. Comparison ψ_3 indicated a trend (p = .08) toward differential effects for the near and far transfer measures. In Comparison ψ_4, further differential transfer within the five far transfer measures was examined by comparing the three far-fluid measures with the two non-fluid measures. Comparison ψ_4 indicated no significant differences between the three far-fluid and two far non-fluid transfer measures.

TABLE 4
Summary of Analysis of Covariance and Orthogonal Planned Comparisons [a]

Source	SS	df	F	p
Pretest	44368.21	1		
Treatment (T)	110.47	1	.33	.55
Error	16285.68	49		
Occasion (O)	1498.19	2	27.61	.00
ψ_1	1492.56	1	55.00	.00
ψ_2	5.62	1	.21	.65
T × O	69.43	2	1.28	.28
Measure (M)	3022.21	6	4.50	.00
T × M	590.25	6	.88	.51
ψ_3	354.55	1	3.17	.08
ψ_4	32.60	1	.29	.59
T × M Residential	203.10	4	.45	
Error	33587.42	300		
O × M	485.59	12	2.36	.01
T × O × M	334.01	12	1.62	.08
ψ_5	22.01	1	1.28	.26
ψ_6	183.66	1	10.70	.00
T × O × M Residual	128.34	10		
Error	10302.52	600		

[a] Total N = 52; Training N = 26; Control N = 26

Given the trend toward the predicted transfer pattern shown in Comparison ψ_3, two contrasts of the Treatment X Occasion X Measure interaction were conducted to compare the transfer pattern between posttest occasions. Comparison ψ_5 was not significant ($p = .26$), indicating that the transfer pattern did not differ between Posttests 1 and 2. Comparison ψ_6 was significant ($p = .00$), indicating that the pattern differed between the average of Posttests 1 and 2, and Posttest 3. These differences in patterns across occasions are also depicted in Figure 1.

Training Effects for Separate Measures. Based on the T x O x M contrasts, a 2 (Treatment) X 2 (Occasion: Posttest 1,2) repeated measures analysis of covariance was conducted separately for each Induction measure and for the nearest measure of far-fluid transfer (ADEPT Figural Relations test). In each case, the covariate was the ADEPT Induction test.

For the nearest transfer measure, ADEPT Induction, a significant treatment effect ($F[1,49] = 7.76$, $p = .01$) was found. The treatment effect across Posttests 1 and 2 was not significant for the Induction Composite test ($F[1,49] = 2.40$, $p = .13$), nor for the ADEPT Figural Relations test ($F[1,49] = 0.97$, $p = .33$).

Retest Effects

Strong retest effects were predicted across posttests due to the test naiveté of elderly subjects. This prediction is supported by the significant Occasion main

TABLE 5
Summary of Orthogonal Planned Comparisons [a]

Comparisons	ADEPT Induc.	Induc. Comp.	ADEPT Fig. Rel. (Measures (M))	Cult. Fair	Raven	P. Speed	Voc.	Posttest Occasions (O) 1	2	3	Treatments (T) Training	Control
O												
ψ_1								2	−1	−1		
ψ_2								0	−1	1		
T × M												
ψ_3	3	2	−1	−1	−1	−1	−1				1	−1
ψ_4	0	0	2	2	1	−2.5	−2.5				1	−1
T × O × M												
ψ_5	3	2	−1	−1	−1	−1	−1	−1	1	0	1	−1
ψ_6	3	2	−1	−1	−1	−1	−1	−1	−1	2	1	−1

[a] Differential weights were assigned to measures according to predicted pattern of training transfer.

effect. Two comparisons (ψ_1, ψ_2) of the Occasion main effect were conducted. Comparison ψ_1 was significant suggesting strong retest effects across treatments and measures from Posttest 1 to Posttests 2 and 3 combined. Comparison ψ_2, examining retest effects between Posttests 2 and 3, was not significant.

The significant Occasion × Measure interaction suggested that occasion effects occurred differentially by measure. Follow-up analyses of the Occasion X Measure interaction, via the Tukey WSD method, showed significant retest effects ($p < .05$) between Posttests 1 and 2 for all measures except the Raven. A graphic summary of these effects is depicted in Figure 2. Note, however, that interpretation of this interaction involves consideration of the Treatment X Occasion X Measure interaction as well. The means shown in Figure 2 are averaged across training and control groups. They reflect differential training effects by measure and occasion (shown in the Treatment X Occasion X Measure interaction), as well as retest effects for training and control. Follow-up analyses were conducted on the Treatment X Occasion X Measure interaction for only the control group, since that group's performance across occasions reflects pure retest effects. The control group's mean scores on the two Induction measures and the ADEPT Figural Relations test were significantly ($p < .05$) greater at both Posttests 2 and 3 than at Posttest 1. In addition, the Posttest 3 mean score on the perceptual speed measure was greater ($p < .02$) than at Posttest 1. These retest effects, unconfounded with training effects, are represented in the control group's performance, shown in Figure 1. Particularly notable in interpreting training results are the significant retest effects for the control group on the three nearest transfer measures.

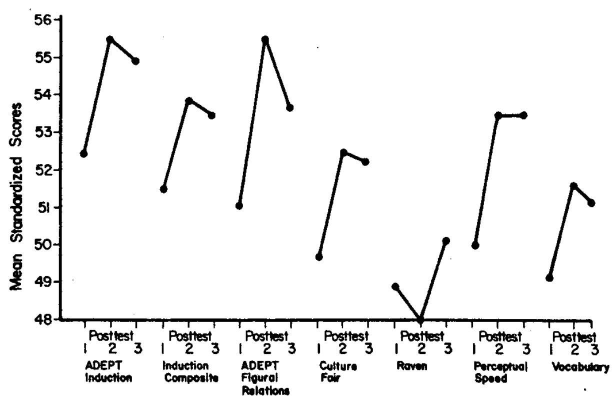

Figure 2. Standardized Mean Scores on Seven Transfer Measures at Three Posttest Occasions.[a]

[a] All data points for each measure were standardized with reference to Control Group's Posttest 1 scores (X = 50, SD = 10).

DISCUSSION

Training Transfer and Maintenance

The effectiveness of Induction training for older adults was examined. Two criteria of training effectiveness was assessed: Training and transfer effects to a broad battery of fluid and crystallized intelligence measures, and maintenance of training transfer effects across three posttests. The largest training effects were predicted to occur at early posttests (one-week, one-month) for measures of the target ability (Induction) with less or no transfer to far (fluid and non-fluid) transfer measures.

Transfer effects were examined at two levels, assessing first the overall pattern of effects across near and far transfer measures and, secondly, transfer effects separately by measure. The predicted pattern of transfer effects was found at the first two posttests. Statistically, this pattern effect is confirmed by the planned comparisons conducted on the Treatment X Measure and the Treatment X Occasion X Measure interactions. Training and transfer effects were also examined at the level of individual measures. A significant training effect to the nearest transfer measure, ADEPT Induction test, was found at Posttests 1 and 2. This test served as the criterion measure for training. Test items of ADEPT Induction were based on the same relational rules utilized in training, although none of the test items were identical to training items. The training effect for the next nearest transfer measure, Induction Composite, was not significant at the individual test level of analysis. None of the other transfer measures yielded significant treatment effects.

Maintenance of training effects was examined across three posttest occasions. The overall *pattern* of predicted transfer effects was achieved and maintained at the one-week and one-month posttest. However, no significant overall pattern differences between training and control groups' performance were found at the six-month posttest. At the individual test level, significant training effects for the ADEPT Induction measure were also maintained at the one-week and one-month posttest, but not at the six-month occasion.

Retest Effects

Retest effects were predicted, given the test naiveté of the elderly. Such retest effects were attributed to nonability-specific performance factors (anxiety reduction, test sophistication) which should affect all or most measures.

In the present design, the purest index of the direction and magnitude of retest effects is the occasion-related observations of the control group. The control group's performance (Figure 1) suggests that the magnitude of such retest effects is differential by measure. Significant retest effects for the control group were found for the ADEPT Induction, Induction Composite, and the ADEPT

Figural Relations tests from Posttest 1 to 2, and for the Perceptual Speed test from Posttest 1 to 3. Notable are the significant retest effects on the three nearest transfer measures. The control group's mean score on the ADEPT Induction reached that of the training group at Posttest 3. In contrast, little practice effect was exhibited for the Raven.

Conclusions and Implications

The present study contributes to the emerging view that older adults continue to have the capacity to benefit from cognitive training efforts (Baltes & Willis, 1981). The range of modifiability of intellectual performance through experimental intervention procedures, then, becomes an issue to be examined across the life-span. Such an experimental approach can complement the existing normative data base on adult intellectual functioning, as has occurred in cognitive research in childhood.

More specifically, the magnitude, breadth, and maintenance of training and transfer effects needs to be examined for a variety of intelligence dimensions. However, the paucity of gerontological training research limits both the comparison of this study's findings and their more general conclusions. Few studies have incorporated transfer paradigms or delayed posttests into their designs. For the fluid intelligence dimension, our prior training research on the fluid ability of Figural Relations (Plemons et al., 1978; Willis et al., 1981) indicated somewhat broader within-ability training and transfer effects than were found in the present Induction study. The more limited training effects on Induction may have resulted from a less adequately designed training program or differences in trainers. A replication of the study with a revised training program is planned. However, the Labouvie-Vief and Gonda (1976) Induction training study also found no significant transfer to the Raven Standard Progressive Matrices for the cognitive training group at immediate posttest. In addition, comparatively narrow training effects have been found for other abilities and performance factors. These include spatial egocentrism (Schultz & Hoyer, 1976), perceptual speed (Hoyer, Labouvie, & Baltes, 1973), and conservation (Hornblum & Overton, 1976).

A major issue in interpreting the present findings involves the magnitude of retest gains observed. In the present study, the significant retest effects for the control group on the three fluid nearest transfer measures makes it difficult to interpret ability-specific training effects. The retest (practice) effects found, however, are consistent with both prior gerontological training research and specific studies of retest (practice) (Hofland et al., 1981). In two Induction studies by Labouvie-Vief and associates (Labouvie-Vief & Gonda, 1976; Pannicucci & Labouvie-Vief, 1975), a practice group outperformed the training group when each was compared with a no-contract control group at a delayed posttest, suggesting the occurrences of strong retest effects such as those found in this study. Moreover, Wing (1980) reported a larger retest effect on Induction measures than for crystallized measures in a young adult sample, and suggested that fluid tasks (due to their relative novelty) may be particularly susceptible to practice effects.

Future Research. Three issues deserve particular attention in future research. A first involves the examination of training effects under a variety of training and assessment conditions. For example, the present research, as well as prior Induction training studies (Labouvie-Vief & Gonda, 1976), suggest the effectiveness of no-feedback practice, as well as an instructor-directed training condition. Future research examining the range of training effects under several different types of training conditions (including a practice condition) would be useful and could assist in delineating the nature of cognitive mechanisms and performance factors associated with training effects in the elderly. In addition, an assessment of effects under a variety of posttesting conditions is needed. In prior research, assessment has involved timed tests in standard administration. However, the slower response speed of older persons may limit the range of training effects demonstrated if assessed under speeded testing conditions. A learning potential approach would suggest training assessment under both speed and "power" testing conditions.

A second issue deals with the possible differential modifiability (trainability) of various intelligence dimensions. For example, if various fluid abilities differed in trainability and/or in optimal intervention strategies (e.g., reinforcement, direct instruction), this could have important implications for the fluid-crystallized theory, as well as the understanding of intellectual aging. Related to such comparative training research would be issues of floor and ceiling effects and measurement scale differences.

The third issue is methodological and deals with the need for additional control groups to study retest effects. The research design should allow assessment of training and maintenance effects unconfounded with retest effects. Initial performance of older persons, due to the lack of test sophistication, is likely to be below their optimal level and, thus, retest effects are strong. Such retest effects seriously contaminate posttraining assessment when training maintenance is monitored over several posttests. Because assessment of training maintenance is desirable, future studies should consider added control groups varying, for example, in number of retests given, and thereby allowing the disentanglement of retest and maintenance effects. Test sophistication and associated general performance factors are an important class of determinants of the level of intellectual functioning in older persons. Thus, a comprehensive view of the conditions for and range of intellectual modifiability in the elderly will need to include statements about the role of testing experience itself.

REFERENCES

Baltes, P. B., Cornelius, S. W., Spiro, A., Nesselroade, J. R., & Willis, S. L. Integration vs. differentiation of fluid-crystallized intelligence in old age. *Developmental Psychology*, 1980, *16*, 625–635.

Baltes, P. B., & Willis, S. L. The critical importance of appropriate methodology in the study of aging: The sample case of psychometric intelligence. In F. Hoffmeister & C. Müller (Eds.), *Brain function in old age*. Heidelberg, Germany: Springer, 1979.

Baltes, P. B., & Willis, S. L. Enhancement of intellectual functioning in old age: Penn State's Adult Development and Enrichment Project (ADEPT). To appear in F. I. M. Craik, & S. E. Trehub (Eds.), *Aging and cognitive processes*. New York: Plenum, 1981, in press.

Beilin, H. Constructing cognitive operations linguistically. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 11). New York: Academic Press, 1976.

Bjork, R. A. All-or-none subprocesses in the learning of complex sequences. *Journal of Mathematical Psychology*, 1968, *5*, 182–195.

Botwinick, J. Aging and intelligence. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand-Reinhold, 1977.

Brown, A. L. Learning and development: The problems of compatibility, access, and induction. *Human Development*, 1981, in press.

Brown, A. L., & French, L. A. The zone of potential development: Implications for intelligence testing in the year 2000. In R. J. Sternberg, & D. K. Detterman (Eds.), *Human intelligence: Perspectives on its theory and measurement*. Norwood, N.J.: Ablex, 1979.

Carlson, J. S., & Wiedl, K. H. Use of testing-the-limits procedure in the assessment of intellectual capabilities in children with learning difficulties. *American Journal of Mental Deficiency*, 1978, *82*, 559–564.

Cattell, R. B. *Abilities: Structure, growth and action*. New York: Houghton-Mifflin, 1971.

Cattell, R. B., & Cattell, A. K. S. *Test of "g": Culture Fair (Scale 2, Form A)*. Champaign, Il.: Institute for Personality and Ability Testing, 1957.

Cattell, R. B., & Cattell, A. K. S. *Test of "g": Culture Fair (Scale 3, Form A*, 1963 Edition; *Form B*, 1961 Edition, Second). Champaign, Il.: Institute for Personality and Ability Testing, 1961, 1963.

Denney, N. W. Problem solving in later adulthood: Intervention research. In P. B. Bates, & O. G. Brim, Jr. (Eds.), *Life-span development and behavior* (Vol. 2). New York: Academic Press, 1979.

Egan, D. E., & Greeno, J. G. Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning, and problem solving. In L. W. Gregg (Ed.), *Knowledge and cognition*. New York: John Wiley, 1974.

Ekstrom, R. B., French, J. W., Harman, H., & Derman, D. *Kit of factor-referenced cognitive tests, 1976 Revision*. Princeton, N.J.: ETS, 1976.

Glaser, R. (Ed.), *Advances in instructional psychology*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Hofland, B. F., Willis, S. L., & Baltes, P. B. Fluid intelligence performance in the elderly: Intraindividual variability & conditions of assessment. *Journal of Educational Psychology*, 1981, *73*, 573–586.

Holzman, T. G., Glaser, R., & Pellegrino, J. W. Process training derived from a computer simulation theory. *Memory and Cognition*, 1976, *4*, 349–356.

Horn, J. L. Human ability systems. In P. B. Baltes (Ed.), *Life-span development and behavior*, (Vol. 1). New York: Academic Press, 1978.

Horn, J. L. 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*. New York: Academic Press, 1970.

Horn, J. L., & Cattell, R. B. Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 1967, *26*, 107–129.

Hornblum, J. N., & Overton, W. F. Area and volume conservation among the elderly: Assessment and training. *Developmental Psychology*, 1976, *12*, 68–74.

Hoyer, W. F., Labouvie, G. V., & Baltes, P. B. Modification of response speed deficits and intellectual performance in the elderly. *Human Development*, 1973, *16*, 233–242.

Kotovsky, K., & Simon, H. Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology*, 1973, *4*, 399–424.

Kuhn, D. Inducing development experimentally: Comments on a research paradigm. *Developmental Psychology*, 1974, *10*, 590–600.

Labouvie-Vief, G. Toward optimizing cognitive competence. *Educational Gerontology*, 1976, *1*, 75–92.

Labouvie-Vief, G., & Gonda, J. N. Cognitive strategy training and intellectual performance in the elderly. *Journal of Gerontology*, 1976, *31*, 327–332.

Matarazzo, J. D. *Wechsler's measurement and appraisal of adult intelligence*. 5th ed. Baltimore, Md.: Williams & Wilkins, 1972.

Panicucci, C. L., & Labouvie-Vief, G. *Effect of training on inductive reasoning behavior*. Paper presented at the 28th Annual Scientific Meeting of the Gerontological Society, Louisville, October, 1975.

Plemons, J. K., Willis, S. L., & Baltes, P. B. Modifiability of fluid intelligence in aging: A short-term longitudinal training approach. *Journal of Gerontology*, 1978, *33*, 224–231.

Raven, J. C. *Advanced Progressive Matrices, Set II, 1962 Revision*. London: H. K. Lewis & Company Lts., 1962.

Resnick, L. B. (Ed.), *The nature of intelligence*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Restle, F., & Brown, E. Organization of serial pattern learning. In G. H. Bower (Ed.) *The psychology of learning and motivation, advances in research and theory*. (Vol. 4). New York: Academic Press, 1970.

Schaie, K. W. The primary mental abilities in adulthood: An exploration in the development of psychometric intelligence. In P. B. Baltes, & O. G. Brim, Jr. (Eds.), *Life-span development and behavior* (Vol. 2). New York: Academic Press, 1979.

Schultz, N. R., & Hoyer, W. J. Feedback effects on spatial egocentrism in old age. *Journal of Gerontology*, 1976, *31*, 72–75.

Simon, H., & Kotovsky, K. Human acquisition of concepts for sequential patterns. *Psychological Review*, 1963, *70*, 534–546.

Sternberg, R. J., & Detterman, D. K. (Eds.), *Human intelligence: Perspectives on its theory and measurement*. Norwood, N.J.: Ablex, 1979.

Thurstone, T. G. *Primary mental abilities, Grades 9–12, 1962 Revision*. Chicago, Ill.: Science Research Associates, 1962.

Willis, S. L., Blieszner, R., & Baltes, P. B. Training research in aging: Modification of performance on the fluid ability of figural relations. *Journal of Educational Psychology*, 1981, *73*, 41–50.

Willis, S. L., & Schaie, K. W. Maintenance and decline of adult mental abilities: II. Susceptibility to experimental manipulation. In F. Grote (Ed.), *Ninth symposium on learning: Adult learning and development*. Bellingham, Wa.: Western Washington University Press, 1981.

Wing, H. Practice effect with traditional mental test items. *Applied Psychological Measurement*, 1980, in press.

Wittrock, M. C. Verbal stimuli in concept formation: Learning by discovery. *Journal of Educational Psychology*, 1963, *54*, 183–190.