Developmental Psychology
1986, Vol. 22, No. 2, 223–232


0012-1649/86/$00.75

# Can Decline in Adult Intellectual Functioning be Reversed?

K. Warner Schaie and Sherry L. Willis
Department of Individual and Family Studies, The Pennsylvania State University

This study examines whether or not cognitive decline in the elderly that has been reliably demonstrated over a 14-year period of time can be reversed. Participants in a long-term longitudinal study ($N$ = 229; aged 64 to 95 years) were classified into those who had declined ($n$ = 122) and those who had remained stable ($n$ = 107) on the abilities of inductive reasoning and spatial orientation. Subjects were assigned to 5-hour training programs on either ability in a pretest–posttest design that employs subjects trained on one target ability as controls for those trained on the other. Results show that: (a) Cognitive training techniques can reverse reliably documented decline over a 14-year period in a substantial number of older adults; (b) such reversal can be documented for two abilities; and (c) training procedures also enhance the performance of many older persons who have remained stable. In addition, training improvement on spatial orientation was found to be greater for decliners than for those who remained stable and greater for women than for men.

A longitudinal sequential research program that has now extended over nearly 3 decades has permitted the collection of substantial descriptive data on the adult life course of some of the principal dimensions of intelligence as identified from a psychometric perspective (Schaie, 1983). This data base informs us that the onset and rate of normative decline differs by ability. Statistically significant decline can be observed for some abilities in individuals in their mid-50s; effect sizes for such decline, however, remain small until the 60s are reached (Schaie & Hertzog, 1983). Thereafter, age-related decline reaches substantial magnitudes, and the proportion of individuals who show reliable decline increases markedly for every 7-year age interval beyond age 60. Nevertheless, it needs to be stressed that there are wide individual differences in the onset of cognitive decline, with many older individuals demonstrating stable levels of intellectual performance into the 70s and some few not showing any decline until the early eighties (Schaie, 1984).

Given the substantial individual differences in decline, we next began to examine possible sources that might help explain these differences. Persons with cardiovascular disease were found to be at excess risk with respect to ability decline. However, it soon became apparent that cardiovascular disease may only be a mediator of decline. That is, adverse life styles that contribute to the incidence of disease may also directly contribute to the occurrence of mental decline (cf. Hertzog, Schaie, & Gribbin, 1978). Detailed analyses of personal characteristics of our study participants have shown that life styles that include the pursuit of high levels of environmental stimulation, particularly those that include continuing formal and informal education, tend to be related to the maintenance of high levels of intellectual functioning (Gribbin, Schaie, & Parham, 1980; Schaie, 1984).

The relationship between intellectually stimulating pursuits and maintenance of abilities thus far has been shown to be reciprocal rather than causal (Stone, 1980). On the one hand, persons who maintain their ability levels into old age may be more likely to seek out experiences of an educational nature. Alternately, exposure to educational experiences in adulthood may be instrumental in maintaining stable levels of intellectual functioning into old age. The question of the causality of naturalistic phenomena can rarely be tested directly in the laboratory. Nevertheless, it may be possible to demonstrate under controlled conditions that intellectual performance in later adulthood can be successfully improved by means of educational training procedures. An adequate answer to this question, however, must take into account the fact that the nature of performance improvement associated with training is far more complicated in later adulthood than at earlier life stages. In childhood, for example, performance improvement associated with training represents a higher level of functioning than previously demonstrated by the child; hence training gains imply the acquisition of new skills or knowledge (Detterman & Sternberg, 1982). Improvement through training in later adulthood, however, could reflect either remediation of performance decrements from prior levels of functioning or the acquisition of new levels of performance not previously demonstrated.

During the past decade and a half there has been an increased focus in the study of adult intelligence on examining the modifiability of intellectual performance in later adulthood through cognitive training procedures (Giambra & Arenberg, 1980; Willis & Baltes, 1980). There is recurrent interest in the question of whether and how well "old dogs can be taught new tricks." From the point of view of both developmental theory and societal benefit, however, it is even more important to ask the question whether such intervention can result in the remediation of reliable

The research reported in this article was supported by Grant AG03544 from the National Institute of Aging. We gratefully acknowledge the enthusiastic cooperation of members and staff of the Group Health Cooperative of Puget Sound. Thanks are due also to Dianne Backschies, Jean Day, Jackie Levine, Cherril Perera, Alejandra Suarez, and Richard Vigesaa who ably assisted in the testing and training, and to Anne O'-Hanlon and Theresa Cooney who assisted in the data analysis.

Correspondence concerning this article should be sent to K. Warner Schaie, Department of Individual and Family Studies, S-110 Human Development Building, The Pennsylvania State University, University Park, Pennsylvania 16802.

age-related decline. The theoretical importance of this question lies in the fact that if it can be shown that reliably demonstrated decline can indeed be reversed then grave doubts would arise as to the universality of irreversible decrement models that assume normative patterns of intellectual decline. These models, of course, also imply that the observed behaviorial deficit should be isomorphic with adverse physiological age changes (cf. Botwinick, 1977; Salthouse, 1982). Instead, support would be offered to the greater plausibility of behavioral deficit being occasioned by specific patterns of disuse that are amenable to remediation for at least some individuals and would strengthen contentions as to the plasticity of behavior throughout life (cf. Baltes & Willis, 1977). In addition, findings of successful remediation efforts at the ability level would have important implications for the development of educational intervention programs that might be helpful in restoring the intellectual competence of many older individuals to levels that would maintain or at least prolong their ability to function independently.

Previous cognitive training research has strongly suggested the modifiability of older adults' performance on a number of intelligence dimensions (e.g., memory span, inductive reasoning, cognitive problem solving, spatial egocentrism; for reviews see Denney, 1982; Poon, Walsh-Sweeney, & Fozard, 1980; Sterns & Sanders, 1980; Willis, 1985). Training effects have been shown to transfer to multiple measures of the ability trained, and performance improvement has been maintained at 6-month posttests (Baltes & Willis, 1982). However, to our knowledge, all of the previous training studies in later adulthood have been conducted within a cross-sectional design, and thus it was not possible to determine whether training improvement represented remediation of prior cognitive decline or the acquisition of new performance levels for subjects suffering no decline. Examination of this question requires a longitudinal subject population, such that training improvement can be compared with subjects' prior level of functioning. This article represents a first report of a study that focuses on the above question.

## Method

### Subjects

Subjects were 229 older adults (male = 97; female = 132) from the Seattle metropolitan area, who had been participants in the Seattle Longitudinal Study (SLS) since 1970 or earlier (Schaie, 1983). All subjects are, or had been, members of the Group Health Cooperative of Puget Sound, a health maintenance organization. Mean age of the total sample was 72.8 years (range = 64–95; $SD$ = 6.41). Mean educational level was 13.9 years (range = 6–20; $SD$ = 2.98). There were no sex differences in age or educational level. Mean income level was $19,879 (range = $1,000–$33,000; $SD$ = $8,520). All subjects were community dwelling. Most of the subjects were Caucasian. Prior to initiation of the study, each subject's physician was contacted and asked to indicate whether the subject suffered any known physical or mental disabilities that would interfere with participation in the study; subjects so identified were not included in the study.

### Design and Procedure

*Classification of participants.* Subjects' test performances on the Thurstone (1948) Primary Mental Ability (PMA) Reasoning and Spatial Orientation measures were classified as having remained stable or having declined over the prior 14-year interval (1970–1984). Subjects entered the study at different points in time (from 1956 through 1970); thus, performance in 1970 was used as a common baseline in order to have a uniform criterion for all subjects. The statistical criterion for the definition of decline was 1 standard error (SE) of measurement or greater (Reasoning = 4 raw score points; Space = 6 raw score points). Subjects were first classified by placing a 1 SE confidence interval about their observed 1970 score (cf. Dudek, 1979). If their 1977 score fell below this interval they were provisionally considered to have declined, otherwise to be stable. Next their 1984 pretest score was considered. Decline subjects who in 1984 returned to within the confidence interval about their 1970 score were then reclassified as stables. Stable subjects whose 1984 score dropped below a 1 SE interval about their 1970 score were reclassified as decliners. Thus, any subject whose score had cumulatively declined by more than 1 SE below the 1970 base was considered to have declined significantly.

There were 107 subjects (46.7% of sample) classified as having remained stable on both ability measures; 35 subjects (15%) had declined on Reasoning but not on Space; 37 subjects (16%) had declined on Space but not on Reasoning; and 50 subjects (21.8%) had declined on both measures. As would be expected, stable subjects ($M$ = 70.9 years; $SD$ = 5.35) were somewhat younger than decline subjects ($M$ = 74.4 years; $SD$ = 6.84), $p < .001$). Although the mean age differed, it should be noted that a wide range occurred for both stables (range = 64–85 years) and decliners (64–95 years). Decline and stable subjects did not differ significantly on educational level or income.

*Effects of regression on classification.* In 1970, prior to the onset of decline, there was no significant difference between subjects who were classified as stables or decliners on Space performance; the decline subjects, at that time, however, performed significantly better ($p < .02$) than the stable subjects on the Reasoning measure in 1970. The possibility then arises that statistical regression effects could have led to the erroneous classification of 1970 high scorers to the decline group and of low scorers to the stable group. Several lines of evidence argue against this possibility. First of all, both tests used have fairly high reliabilities. Although the split-half coefficients in the .90s reported by Thurstone (1948) may be overstated due to the slightly speeded nature of the tests, our own data on long-term stability have consistently run to .70 or higher for Space and .80 or higher for Reasoning (Schaie, 1983, 1985). As an independent check on the plausibility of regression effects, however, we conducted a time-reversed control analysis (cf. Baltes, Nesselroade, Schaie, & Labouvie, 1972; Campbell & Stanley, 1966).

In a time-reversed control analysis one partitions scores at the point of origin (1970 in our case) into the highest, middle, and lowest thirds, and computes means for these levels at both origin and final measurement point (1984 pretest in our case). The traces from the first to the second time of measurement are then examined to see whether they are parallel or show apparent regression effects. In the latter case the mean score for the highest third should have decreased, whereas the mean for the lowest third should have increased. The same strategy is then used to partition scores at the final point (1984 pretest in our case) into three ability levels and to compute means for these ability groupings at the point of origin. If regression effects have occurred, then the traces for the "backward" (1984–1970) analyses should actually be the mirror images of the traces for the "forward" (1970–1984) analyses. Figure 1 shows the results of the time-reversed control analyses for Space and Reasoning both for the total group and separately for the decliners. The traces shown in this figure are clearly incompatible with the presence of substantial regression effect. We can conclude, then, that our classification error rate (as specified by our confidence limits) has not been significantly increased by regression effects.

*Assignment of subjects.* Subjects were assigned to either Reasoning or Space training programs based on their performance status. Subjects who had declined on Reasoning, but not on Space, or vice versa were assigned to the training program for the ability exhibiting decline. Subjects who had remained stable on both abilities or had shown decline on both abilities were randomly assigned to one of the training programs. Space training subjects included 51 stables (male = 23; female = 28) and 67 decliners (male = 29; female = 38). Reasoning training subjects included 56 stables (male = 25; female = 31) and 55 decliners (male = 20; female = 35). Stable subjects in Space training were significantly younger ($p < .03$) and better educated ($p < .05$) than decliners. Stable subjects in Reasoning training were also significantly younger ($p < .001$) than decliners, but decliners were significantly better educated ($p < .01$) than stables. There was no difference in income for either program.

*Procedure.* The study involved a pretest-treatment-posttest control group design. The Reasoning training group served as a treatment control for the Space training group and vice versa. All subjects had previously participated in the SLS and were informed via a series of letters that a new phase of the study was beginning. Subjects indicating an interest in participation were visited in their homes by a staff member. The purpose of the home visit was to discuss details of the study and to answer questions, to assess sensory handicaps that might interfere with participation, and to determine whether the home was a suitable place for conducting the training sessions. Subjects were administered a broad psychometric ability battery in two pretest sessions (2½ hours per session). Based on their prior longitudinal performance plus their pretest scores, subjects were assigned to either the Reasoning or Space training program. Training involved five 1-hour individually conducted training sessions. The majority of subjects were trained in their homes. Two middle-aged trainers, with prior educational experience in working with adults, served as trainers. Subjects were randomly assigned to the trainers within pragmatic constraints, such that each trainer trained approximately equal numbers of stable and decline subjects in each training program. Following training, subjects were assessed on a posttest battery involving the same measures administered at pretest. Subjects were paid $100 for participation in the study.

### *Measures*

The pretest–posttest battery involved psychometric measures representing five primary mental abilities, including the Thurstone PMA measures (Thurstone, 1948) administered at previous SLS assessments. Each ability was represented by two to four marker measures. The data to be presented focus on the measures representing the Spatial Orientation and Reasoning abilities; these abilities were the target of training.

Spatial Orientation was assessed by four measures. Three of the tests (PMA Space, Object Rotation, Alphanumeric Rotation) are multiple-response measures of two-dimensional mental rotation ability. The subject is shown a model line drawing and asked to identify which of six choices shows the model drawn in different spatial orientations. There are two or three correct responses possible for each test item. The Object Rotation test (Schaie, 1985) and the Alphanumeric test were constructed such that the angle of rotation in each answer choice is identical with the angle used in the PMA Spatial Orientation test (Thurstone, 1948). The three tests vary in item content. Stimuli for the PMA test are abstract figures; the Object Rotation test involves drawings of familiar objects; and the Alphanumeric test contains letters and numbers. The fourth test, Cube Comparisons (Ekstrom et al., 1976), assesses mental rotation in three-dimensional space.

Reasoning ability was assessed by four measures. The PMA Reasoning measure (Thurstone, 1948) assesses inductive reasoning ability via letter series problems. The subject is shown a series of letters and must select the next letter in the series from five letter choices. The ADEPT Letter

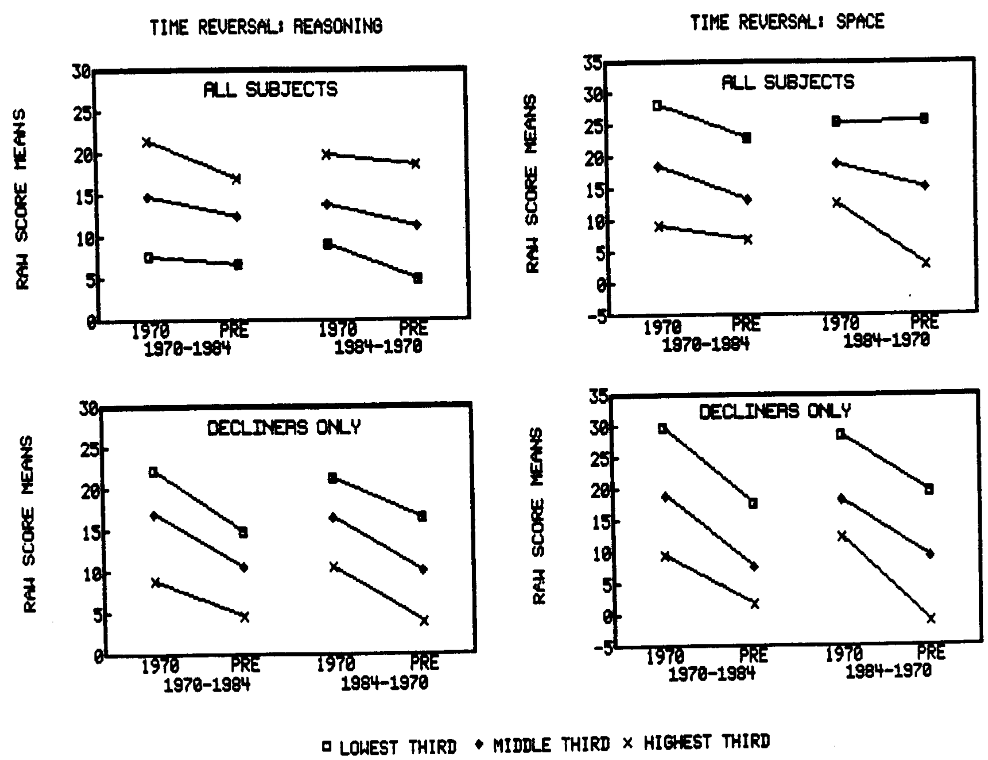


*Figure 1.* Time-reversed control analyses for regression effects from 1970 to 1984 for the Primary Mental Ability Space and Reasoning tests for all subjects and for decline subjects only.

Series test (Blieszner, Willis, & Baltes, 1981) also contains letter series problems; however, some of the problems involve pattern description rules other than those found on the PMA measure. The Word Series test (Schaie, 1985) parallels the PMA measure in that the same pattern description rule is used for each item; however, the test stimuli are days of the week or months of the year, rather than letters. The Number Series test (Ekstrom et al., 1976) involves series of numbers rather than letters and involves different types of pattern description rules involving mathematical computations. The PMA Spatial Orientation and Reasoning tests were administered at previous SLS measurement points and thus provide the most direct assessment of training improvement and remediation.

### Training Programs

The focus of the training was on facilitating the subject's use of effective cognitive strategies identified in previous research on the respective abilities. A content task analysis was conducted on the two PMA measures.

*Reasoning.* For each item of the Reasoning test, the pattern description rule(s) used in problem solution were identified. Four major types of pattern description rules (identity, next, skips, and backward next) were identified, similar to those discussed previously in the literature (Holzman, Pellegrino, & Glaser, 1982; Kotovsky & Simon, 1973). Practice problems and exercises were developed, based on these pattern description rules. Practice problems often involved content other than letters, so that the applicability of these rules to other content areas could be explored. For example, patterns of musical notes and travel schedules were devised based on these rules, and subjects were to identify the next note or destination in the series. No training problems were identical in content to test items. Subjects were taught through modeling, feedback, and practice procedures to identify these pattern description rules. These multiple procedures were employed in order to optimize training effects, rather than to assess their differential effectiveness. Three strategies for identifying the patterns were emphasized in training: visual scanning of the series, saying the series aloud in order to hear the letter pattern, and underlining repeated letters occurring throughout the series. Once a hypothesis regarding the pattern type was generated, subjects were then taught to mark repetitions of the pattern within the series, and thus to determine the next letter required to fit the pattern rule.

*Space.* A content task analysis of the PMA Spatial Orientation test was conducted to identify the angle of rotation for each answer choice. Practice problems were developed to represent the angle rotations identified in the task analysis (45°, 90°, 135°, 180°). Cognitive strategies to facilitate mental rotation that were focused on in training included: (a) development of concrete terms for various angles; (b) practice with manual rotation of figures prior to mental rotation; (c) practice with rotation of drawings of concrete, familiar objects prior to introduction of abstract figures; (d) subject-generated names for abstract figures; and (e) having the subject focus on two or more features of the figure during rotation. These cognitive strategies had been identified in prior descriptive research on mental rotation ability (Cooper & Shepard, 1973; Egan, 1981; Kail, Pellegrino, & Carter, 1980).

## Results

Results of the study are reported in three parts. First, analyses are provided for training improvement at the level of ability factor scores. Second, analyses of training effects at the level of raw scores are reported for the PMA Reasoning and Space tests because these are the measures for which longitudinal data are available. Finally, the proportion of subjects demonstrating remediation of decline is reported.

### Analyses of Factor Scores

The factor structure of the pretest ability battery was examined via confirmatory factor analyses. An acceptable five-factor model $\chi^2(243, N = 1{,}401) = 463.17, p < .001$, representing the hypothesized primary mental ability factors (Reasoning, Verbal, Space, Number, and Perceptual Speed), was obtained. Marker measures of the Reasoning and Space abilities had significant loadings, as predicted, on their respective factors. Factor regression weights for tests loading on the Reasoning factor were: PMA Reasoning = .378; ADEPT Letter Series = .213; Word Series = .298; Number Series = .111. Factor weights of tests loading on the Space factor were: PMA Space = .260; Object Rotation = .393; Alphanumeric = .287; Cube Comparison = .060. Although Number Series and Cube Comparison contribute relatively little variance to their respective factors, they were retained because they helped obtain better definition of the factors within the broader ability space in which they were embedded. Factor scores were computed for the Reasoning and Space abilities by standardizing ($M = 50$; $SD = 10$) the raw scores to the pretest base and then multiplying the standardized scores by their respective regression weights.

To examine training effects at the factor level, two Training (Reasoning, Space) × Status (stable, decline) × Sex × Occasion (pretest, posttest) analyses of variance (ANOVAs) with repeated measures were performed separately on the Reasoning and Space factor scores (Table 1). For Reasoning, there were significant main effects for status ($p < .001$), sex ($p < .03$), and occasion ($p < .001$). The status and sex main effects reflected the lower scores of decliners and men, respectively. The occasion main effect represents retest effects occurring for both groups. More importantly with regard to training effects, there was a significant Training × Occasion interaction ($p < .001$), indicating a training effect at posttest. No status, sex, or sex × status comparisons within the Reasoning training group were significant; thus, the training effect was general and not specific to status and/or sex. Figure 2 shows the pretest–posttest gain computed from the standardized factor scores for the Reasoning and Space abilities for four training subgroups (stable/decline on Reasoning, stable/decline on Space). Each set of four bars in Figure 2 compares the two subgroups trained on Reasoning with the two trained on Space (see legend for Figure 2). That is, each training group serves as a control for the other training condition.

For Space, there were significant main effects for status ($p < .001$), sex ($p < .01$), and occasion ($p < .001$; Table 1). The status and sex main effects reflect the lower scores of the decliners and women, respectively, across occasions. The occasion main effect indicates retest effects occurring for both Reasoning and Space groups. There were significant interactions for Training × Occasion ($p < .02$) and Training × Sex ($p < .04$). The Training × Occasion interaction indicated a significant training effect at posttest. The Training × Sex interaction indicates that the overall performance on Space was higher for the target training group than for the controls for women but not for men.

### Raw Score Analyses: PMA Reasoning and Space

To examine training effects for the two measures with longitudinal data, two Training (Reasoning, Space) × Status × Sex ×

Occasion ANOVAs with repeated measures were performed separately for the PMA Reasoning and Space tests (Table 2). For PMA Reasoning, there were significant main effects for status ($p < .001$), sex ($p < .01$), and occasion ($p < .001$). The status and sex main effects again reflect the lower scores on the target measure for decliners and men, respectively. The occasion main effect represents retest effects occurring for both groups. With respect to training effects of central concern, there was a significant Training × Occasion interaction ($p < .001$), indicating higher performance of those trained on Reasoning at posttest. There was a trend toward a significant fourfold interaction ($p < .09$). Post hoc tests on PMA Reasoning gain scores indicated that decliners showed greater gain than did stables. Sex and Sex × Status effects were not significant. When the Reasoning and Space training groups were compared, there were significant Reasoning training effects for stables ($p < .001$), decliners ($p < .001$), stable women ($p < .002$), and male and female decliners ($p < .001$).

For PMA Space, there were significant main effects for status ($p < .001$), sex ($p < .02$), and occasion ($p < .001$; Table 2). The status and sex main effects reflect the lower scores of the decliners and women, respectively, across occasions. The occasion main effect indicates retest effects occurring for both the Reasoning and Space training groups. As for the crucial results with respect to the training paradigm, there were significant interactions for Training × Occasion ($p < .004$) and Training × Status × Occasion ($p < .05$). The Training × Occasion interaction indicated a significantly higher performance for the Space training group at posttest. The triple interaction with Status reflects greater training gain for the decliners at posttest. A significant Sex × Occasion interaction ($p < .05$) suggests the occurrence of larger retest effects for women. Post hoc tests on PMA Space gain scores indicated that there were significantly greater ($p < .01$) gains for decliners than for stables.

Table 1
*Summary of Analyses of Variance: Factor Scores*

| Source | *MS* | *dfs* | *F* |
|---|---|---|---|
| | Reasoning | | |
| Training | 336.58 | 1 | 2.18 |
| Status | 3038.35 | 1 | 19.66*** |
| Training × Status | 25.71 | 1 | 0.17 |
| Sex | 733.10 | 1 | 4.74* |
| Training × Sex | 362.67 | 1 | 2.35 |
| Status × Sex | 30.31 | 1 | 0.20 |
| Training × Status × Sex | 0.08 | 1 | 0.00 |
| Error | 154.59 | 221 | |
| Occasion | 1649.15 | 1 | 303.15*** |
| Training × Occasion | 205.99 | 1 | 37.86*** |
| Status × Occasion | 1.76 | 1 | 0.34 |
| Training × Status × Occasion | 7.76 | 1 | 1.43 |
| Sex × Occasion | 0.58 | 1 | 0.11 |
| Training × Sex × Occasion | 1.18 | 1 | 0.22 |
| Status × Sex × Occasion | 0.07 | 1 | 0.01 |
| Training × Status × Sex × Occasion | 1.08 | 1 | 0.20 |
| Error | 5.44 | 221 | |
| | Space | | |
| Training | 84.19 | 1 | 0.56 |
| Status | 3884.11 | 1 | 30.10*** |
| Training × Status | 26.33 | 1 | 0.20 |
| Sex | 852.04 | 1 | 6.60** |
| Training × Sex | 521.31 | 1 | 4.04* |
| Status × Sex | 423.18 | 1 | 3.28 |
| Training × Status × Sex | 79.17 | 1 | 0.61 |
| Error | 129.06 | 221 | |
| Occasion | 1556.41 | 1 | 195.48*** |
| Training × Occasion | 41.82 | 1 | 5.25* |
| Status × Occasion | 9.49 | 1 | 1.19 |
| Training × Status × Occasion | 18.14 | 1 | 2.28 |
| Sex × Occasion | 4.14 | 1 | 0.52 |
| Training × Sex × Occasion | 16.27 | 1 | 2.04 |
| Status × Sex × Occasion | 0.75 | 1 | 0.09 |
| Training × Status × Sex × Occasion | 4.83 | 1 | 0.61 |
| Error | 7.96 | 221 | |

* $p < .05$. ** $p < .01$. *** $p < .001$.

### *Distinguishing Between Regression and Training Effects*

In order to exclude the possibility that regression effects might confound the results of the training study, we first of all examined the stability of our instruments over the interval between pretest and posttest by administering the measures over the same interval to a group 172 subjects of comparable age and socioeconomic status who did not receive any training. Stability coefficients obtained in this study were found to be .917 for the Space factor score and .939 for the Reasoning factor score. Stabilities for the two PMA measures were found to be .838 for Space and .886 for Reasoning. These estimates were next used to compute regressed deviation scores for our experimental subjects (cf. Nunnally, 1978). The ANOVAs described above were then repeated on the adjusted scores. As would be expected in light of the high stabilities, resulting *F* ratios differed only trivially, and none of the previously reported findings was significantly affected.

### *Effects of Age, Education, and Income*

Because of slight differences between subgroups in terms of demographic characteristics, we also repeated the ANOVAs covarying on age, education, and income. Again, effects of the covariance adjustments were trivial, and none of the findings reported above was changed significantly.

### *Pretest–Posttest Training Improvement: Proportion of Subjects*

The proportion of subjects showing statistically reliable pretest–posttest training improvement on the PMA Reasoning or Space measure was computed. The statistical criterion for significant improvement was defined as a gain ≥ 1 SE from pretest to posttest. The proportion of subjects at the individual level with reliable training gain is shown in Table 3. Approximately one half of the subjects in each training group showed significant pretest–posttest improvement. Although there was a trend for a

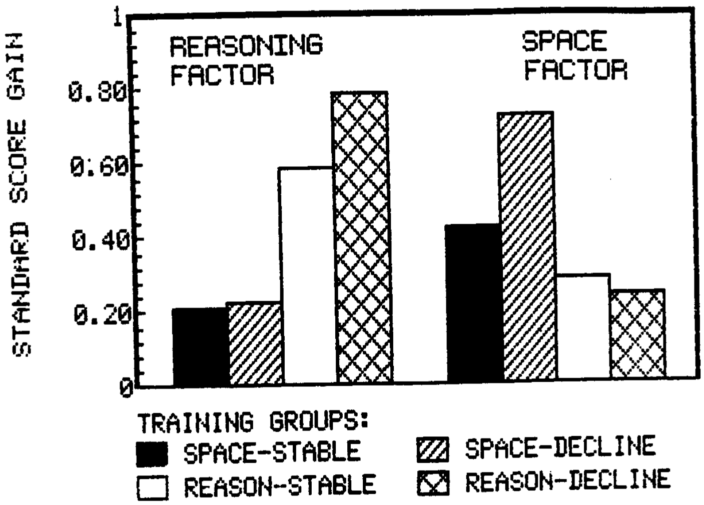


*Figure 2.* Mean factor score gains by training program and status (in standard score units).

greater proportion of decline subjects to show improvement in both training conditions, the difference between proportions is statistically significant only for the Space group ($p < .01$).

### *Remediation of Decline*

The final set of analyses examined the percentage of decline subjects whose posttest scores were equivalent to their 1970 performance. That is, we asked for what percentage of subjects did training result in a 14-year remediation to their 1970 performance level? Because longitudinal data are available only for the PMA tests, remediation was assessed with regard to the PMA Reasoning and Space measures. Two criterion levels were used to define remediation. The first level defined remediation as having occurred when the difference between the subjects' PMA posttest score and their 1970 score was ≤ 1 SE. This criterion employed the same statistical definition (1 SE) as that used in classifying subjects with regard to 1970–1984 decline. The second criterion level was even more conservative and defined remediation as the attainment of a PMA posttest score that was equal to or greater than the 1970 score. Table 4 and Figure 3 present the proportion of decline subjects reaching these remediation criteria. Sixty-two percent of the decline subjects were remediated to their predecline level if the 1 SE criterion is used. In both training groups more women were returned to their 1970 score level than were men. Using the more conservative criterion of return to the 1970 base level, 40% of the decline subjects' posttest scores were equal to or greater than their 1970 score. By this more stringent criterion, more women than men were returned to base level for Space, whereas more men than women were returned to their base level for Reasoning.

## Discussion

### *Differential Ability Decline*

This experimental study is the most recent phase of a descriptive, longitudinal program of research that has examined intellectual change in adulthood over the past 28 years. In this study we focused on the modifiability of change occurring from late midlife into old age. A major finding of the Seattle Longitudinal Study has been the observation of wide individual differences in the onset and magnitude of intellectual decline. The initial phase of this study, which involved the classification of individuals with regard to decline status on two primary abilities, provides further support for the variability of cognitive functioning in later adulthood. Almost one half (46%) of the subjects exhibited no statistically reliable decline on either of the primary abilities studied over the previous 14-year period. The finding that less than one quarter of the subjects showed decline on both Space and Reasoning argues further that the onset of decline varies significantly across various abilities. For most individuals the pattern of decline appears to be selective, perhaps even ability specific, rather than global and catastrophic. The considerable stability in intellectual functioning is noteworthy, given that both of the abilities studied (Inductive Reasoning, Spatial Orientation) involve abstract reasoning on speeded measures and would thus be expected to exhibit normative patterns of decline if one were to extrapolate from the widely accepted classical pattern of cognitive aging (Botwinick, 1977).

Some might argue, however, that regression effects could mask decline in the less able or exaggerate decline for those at the top ability levels. This issue has been previously examined (Baltes et al., 1972) and it was shown that such effects are likely only when measures of low reliability are used. That is not the case here, and we demonstrated further by means of a time-reversed control analysis that our decline criterion appears reasonably linked to ontogenetic effects that generalize across ability groupings and are minimally impacted by statistical regression.

Whereas subjects showing no decline were on average 2 years younger than those having declined, there is a wide age range (64–85 years) for those classified as stable, suggesting that for some individuals cognitive decline may be a condition of old-old age, rather than occurring in middle or young-old age. We do not construe these data to imply that there is no cognitive decline in old age, although some have in the past accused us of this position (Donaldson, 1981; Horn & Donaldson, 1976). Rather, we believe these data to argue for large individual differences in the onset and pattern of intellectual aging.

### *Training Effects at the Factorial Level*

Significant training effects were demonstrated for both Reasoning and Space abilities. These effects were demonstrated at the ability factor level as well as at the individual measure level. The observation of significant effects at the factorial level is important in that it extends the findings of prior cross-sectional training research, which has examined training improvement at the level of individual measures. Some (Birren, Cunningham, & Yamamoto, 1982; Donaldson, 1981) have questioned the breadth of training effects when assessed with respect to individual measures. The current data indicate that the effects of training are not restricted to performance on one specific test but rather reflect training improvement at the level of the primary ability.

Table 2
*Summary of Analyses of Variance: Primary Mental Ability Raw Scores*

| Source | *MS* | *df*s | *F* |
|---|---|---|---|
| | Reasoning | | |
| Training | 41.03 | 1 | 0.61 |
| Status | 1471.25 | 1 | 21.89*** |
| Training × Status | 1.34 | 1 | 0.02 |
| Sex | 483.58 | 1 | 7.20** |
| Training × Sex | 113.37 | 1 | 1.69 |
| Status × Sex | 54.51 | 1 | 0.81 |
| Training × Status × Sex | 5.77 | 1 | 0.09 |
| Error | 67.20 | 221 | |
| Occasion | 857.75 | 1 | 150.24*** |
| Training × Occasion | 231.02 | 1 | 40.47*** |
| Status × Occasion | 12.24 | 1 | 2.14 |
| Training × Status × Occasion | 12.71 | 1 | 2.23 |
| Sex × Occasion | 2.77 | 1 | 0.48 |
| Training × Sex × Occasion | 0.75 | 1 | 0.13 |
| Status × Sex × Occasion | 7.24 | 1 | 1.27 |
| Training × Status × Sex × Occasion | 16.16 | 1 | 2.83 |
| Error | 5.71 | 221 | |
| | Space | | |
| Training | 470.12 | 1 | 2.91 |
| Status | 4228.33 | 1 | 26.14*** |
| Training × Status | 177.70 | 1 | 1.10 |
| Sex | 922.39 | 1 | 5.70* |
| Training × Sex | 359.65 | 1 | 2.22 |
| Status × Sex | 388.56 | 1 | 2.40 |
| Training × Status × Sex | 0.66 | 1 | 0.00 |
| Error | 161.74 | 221 | |
| Occasion | 2044.16 | 1 | 90.68*** |
| Training × Occasion | 301.64 | 1 | 13.38*** |
| Status × Occasion | 49.23 | 1 | 2.18 |
| Training × Status × Occasion | 106.56 | 1 | 4.73* |
| Sex × Occasion | 85.12 | 1 | 3.78* |
| Training × Sex × Occasion | 55.47 | 1 | 2.46 |
| Status × Sex × Occasion | 46.55 | 1 | 2.06 |
| Training × Status × Sex × Occasion | 0.53 | 1 | 0.02 |
| Error | 22.54 | 221 | |

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3
*Proportion of Subjects Achieving Significant Pretest to Posttest Training Gain*

| | Reasoning | | | Space | | |
|---|---|---|---|---|---|---|
| Status | Men | Women | Total | Men | Women | Total |
| Stable | 52.0 | 54.8 | 53.6 | 34.8 | 42.9 | 39.2 |
| Decline | 60.0 | 60.0 | 60.0 | 51.7 | 57.9 | 55.2 |
| Total | 55.6 | 57.6 | 56.8 | 44.2 | 51.5 | 48.3 |

*Note.* Significant training improvement was defined as a pretest-posttest gain of ≥ 1 standard error of measurement on the Primary Mental Ability Reasoning or Space test.

### *Decline Status and Training Effects*

Two major questions were addressed in this study. The first question focused on the differential effectiveness of training for remediating cognitive decline versus the development of new performance levels in the elderly showing no previous decline. The second question asks for how large a proportion of subjects exhibiting decline can there be remediation through training to a previously observed (1970) level of performance. Longitudinal data are required to address these questions because the prior intellectual history of the subjects must be known in order to examine the differential effects of training (remediation vs. new learning) and to determine the extent to which remediation has occurred. Prior training studies have been cross-sectional in design, and whereas many have found significant treatment effects, the nature of the observed training gain could not be specified.

With regard to the first question, no statistically significant interaction effects involving decline status were found at the ability level. However, decliners did improve significantly more than did the stable subjects on the PMA Space test (see Table 2). The greater effectiveness of training for decliners is further supported by an examination of the proportion of subjects attaining significant training gain from pretest to posttest. Fifty-five percent ($n = 37$) of the decliners in the Space training group achieved significant pretest–posttest training improvement, compared with 39% ($n = 20$) of the stables; the difference in the proportion of decline versus stable subjects showing improvement was statistically significant ($p < .01$). This difference is less pronounced in Reasoning training where 60% ($n = 33$) of the decliners showed significant pretest–posttest training gains, compared with 53.6% ($n = 30$) of the stables. Although age, education, and income are associated with ability decline, covarying on these demographic variables does not significantly alter any of our findings.

Whereas a greater proportion of decliners showed pretest–posttest training improvement, it is impressive to note that marked improvement was also seen for the stable subjects. More than half of the stables in the Reasoning training showed significant pretest–posttest gain, as did approximately 40% of the sta-

Table 4
*Remediation of Ability Decline: Proportion of Subjects*

| | Reasoning | | Space | | |
|---|---|---|---|---|---|
| Criterion | Men ($n$ = 20) | Women ($n$ = 35) | Men ($n$ = 29) | Women ($n$ = 38) | Grand total |
| 1 SEM[a] | 55.0 | 65.7 | 55.2 | 68.4 | 62.3 |
| To 1970 score[b] | 50.0 | 40.0 | 31.0 | 42.1 | 40.2 |

[a] Proportion of decline subjects whose 1970 Primary Mental Ability (PMA) scores and 1984 PMA posttest scores differed by ≤ 1 standard error of measurement.
[b] Proportion of decline subjects whose 1984 PMA posttest score was ≥ their 1970 PMA score.

bles in Space training (Table 3). Many elderly persons who remain at stable levels of functioning are still disadvantaged by limited early educational and experiential opportunities. Intervention techniques such as those used in this study may therefore have utility in improving specific cognitive skills in these otherwise well-functioning persons.

## *Remediation of Ability Decline*

The second major question focused on the extent to which training is effective in remediating decline to a previous performance level. Remediation of decline was examined over a 14-year period (1970–1984). For most of the subjects in the study the 1970 basepoint represents a performance level prior to the onset of statistically reliable decline. Most subjects in 1970 had not reached the age at which normative patterns of statistically significant decline have been noted for their cohorts (Schaie, 1983). Remediation of decline was defined as occurring when a subject's 1984 posttest score was equivalent to his or her 1970 score on the relevant PMA measure. Forty percent of subjects exhibited remediation of decline on the ability trained, when the more conservative estimate of remediation was used, and 62% when a more lenient (confidence interval) criterion was employed (Table 4). There was no significant difference in the proportion of subjects for whom remediation occurred for Reasoning versus Space training. Thus, the data suggest that decline on both Space and Reasoning ability is responsive to training efforts.

The finding that observed cognitive decline can be remediated for large proportions of subjects as the result of the application of quite modest experiential intervention procedures suggests strongly that at least a portion of the previously observed decline may be attributable to disuse. What our intervention procedures seem to accomplish is to reactivate behaviors and skills that have remained in the subjects' behavioral repertory but that have not been actively employed (Bearison, 1974; Overton & Newman, 1982).

## *Training and Sex*

The sex main effects (Tables 1 and 2) indicate sex differences for both abilities, but in the opposite direction. There is a significant difference in favor of women on Inductive Reasoning and in favor of men on Space. These patterns of sex difference have been noted in previous research (Maccoby & Jacklin, 1974; Schaie, 1983). A number of trends in the data suggest that the training may have been particularly effective for women. Sex effects are most salient for Space training. Across both training programs, 67% of the women decliners were remediated within 1 SE of their 1970 performance level, compared with 55% of the male decliners (Table 4).

## *Conclusions*

The major objectives of this study were to examine whether or not cognitive decline that had been reliably demonstrated over a substantial period of time could be reversed by educational training techniques, and whether stable performance in old age could be significantly improved. In order to answer these questions it was necessary to classify older members of a long-term longitudinal study into those who had declined and those who had remained stable and to develop a paradigm that would provide appropriate controls by training both decliners and stables on two different target abilities.

The results of the study clearly show that: (a) Cognitive training techniques can reverse reliably documented decline over a 14-year period in a substantial number of older adults; (b) such reversal can be documented for at least two abilities; (c) training procedures enhance the performance of those older persons who have remained stable; and (d) magnitudes of training effects are unrelated to age, education, and income.

Highly significant retest effects were found for all subgroups and can be attributed in part to the extensive test batteries that provided practice on several measures of each target ability. The training paradigm employing two training programs, however, provides the necessary controls to permit the conclusion that training results are specific to the objectives of each training program and that subjects trained in each program on average show significantly greater gains than do the controls receiving the alternate training program.

It should be kept in mind that all our subjects were community dwelling adults in fair to excellent health condition who lived in reasonably supportive environments. We consequently would not wish to claim that we have discovered methods that are likely to reverse cognitive changes of a neuropathological nature that are found in some of the elderly. Our approaches, however, are directly applicable to the remediation of the cognitive decline seen in many older adults without psychopathology that may be the result of experiential changes commonly occurring in the later stages of life.

Although the improvement and reversal of decline demonstrated in this study are impressive, it should be understood that such results may be a rather conservative estimate of what could be achieved by more extensive programs of this kind. The training procedures employed in this study were relatively brief (five 1-

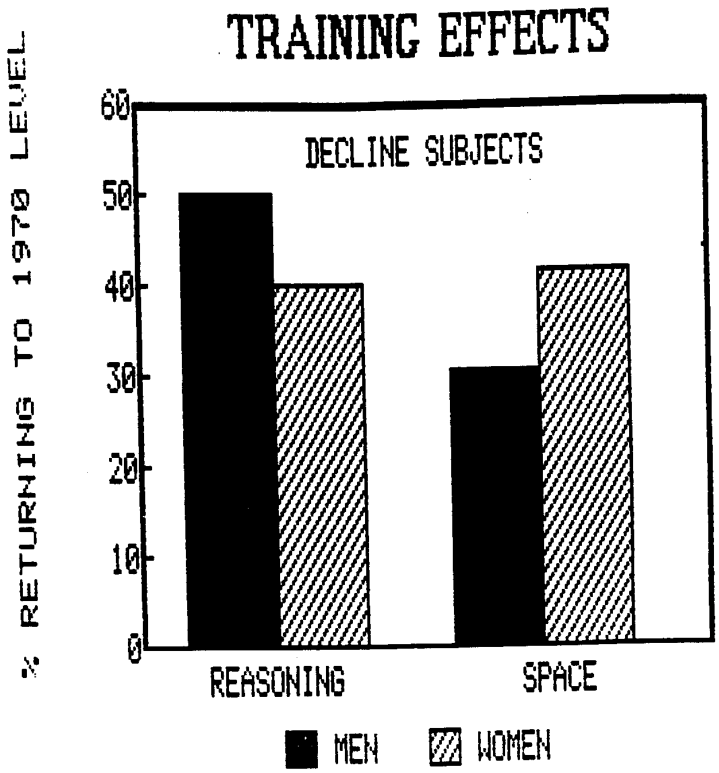


*Figure 3.* Proportion of decline subjects remediated to 1970 score level.

hour sessions), compared with the massive educational interventions common earlier in the lifespan, and the breadth of remediation possible with more extensive training efforts has yet to be examined.

Most importantly, our findings lend support to contentions regarding the plasticity of behavior into late adulthood. They suggest that for at least a substantial proportion of the community dwelling elderly, observed cognitive decline is not irreversible, is likely to be attributable to disuse, and can be subjected to environmental manipulations involving relatively simple and inexpensive educational training techniques.

## References


Baltes, P. B., Nesselroade, J. R., Schaie, K. W., & Labouvie, E. W. (1972). On the dilemma of regression effects in examining ability-level-related differentials in ontogenetic patterns of intelligence. *Developmental Psychology, 6,* 78–84.

Baltes, P. B., & Willis, S. L. (1977). Toward psychological theories of aging and development. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 128–154). New York: Van Nostrand Reinhold.

Baltes, P. B., & Willis, S. L. (1982). Enhancement (plasticity) 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.

Bearison, D. (1974). The construct of regression: A Piagetian approach. *Merrill-Palmer Quarterly, 20,* 21–30.

Birren, J. E., Cunningham, W. R., & Yamamoto, K. (1982). Psychology of adult development and aging. *Annual Review of Psychology, 34,* 543–575.

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.

Botwinick, J. (1977). Intellectual abilities. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 580–605). New York: Van Nostrand Reinhold.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cooper, L. B., & Shepard, R. N. (1973). Chronometric studies of rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–96). New York: Academic Press.

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.

Detterman, D. K., & Sternberg, R. J. (Eds.). (1982). *How much can intelligence be increased?* Norwood, NJ: Ablex.

Donaldson, G. (1981). Letter to the editor. *Journal of Gerontology, 36,* 634–636.

Dudek, F. J. (1979). The continuing misinterpretation of the standard error of measurement. *Psychological Bulletin, 86,* 335–337.

Egan, D. E. (1981). An analysis of spatial orientation test performance. *Intelligence, 5,* 85–100.

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

Giambra, L. M., & Arenberg, D. (1980). Problem solving, concept learning, and aging. In L. W. Poon (Ed.), *Aging in the 1980s* (pp. 253–259). Washington, DC: American Psychological Association.

Gribbin, K., Schaie, K. W., & Parham, I. A. (1980). Complexity of life style and maintenance of intellectual abilities. *Journal of Social Issues, 21,* 47–61.

Hertzog, C. K., Schaie, K. W., & Gribbin, K. (1978). Cardiovascular disease and changes in intellectual functioning from middle to old age. *Journal of Gerontology, 33,* 872–883.

Holzman, T. G., Pellegrino, J. W., & Glaser, R. (1982). Cognitive dimensions of numerical rule induction. *Journal of Educational Psychology, 74,* 360–373.

Horn, J. L., & Donaldson, G. (1976). On the myth of intellectual decline in adulthood. *American Psychologist, 31,* 701–719.

Kail, R., Pellegrino, J., & Carter, P. (1980). Developmental changes in mental rotation. *Journal of Experimental Child Psychology, 39,* 102–116.

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

Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

Nunnally, J. C. (1978). *Psychometric theory* (2nd Ed.). New York: McGraw-Hill.

Overton, W. F., & Newman, J. L. (1982). Cognitive development: A competence-activation/utilization approach. In T. Field, A. Huston, H. Quay, L. Troll, & G. Finlay (Eds.), *Review of human development* (pp. 217–241). New York: Wiley.

Poon, L. W., Walsh-Sweeney, L., & Fozard, J. L. (1980). Memory training for the elderly. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New directions in memory and aging* (pp. 461–484). Hillsdale, NJ: Erlbaum.

Salthouse, T. A. (1982). *Adult cognition.* New York: Springer-Verlag.

Schaie, K. W. (1983). The Seattle Longitudinal Study: A 21-year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64–135). New York: Guilford.

Schaie, K. W. (1984). Midlife influences upon intellectual functioning in old age. *International Journal of Behavioral Development, 7,* 463–478.

Schaie, K. W. (1985). *Manual for the Schaie-Thurstone Adult Mental Abilities Test (STAMAT).* Palo Alto, CA: Consulting Psychologists Press.

Schaie, K. W., & Hertzog, C. (1983). Fourteen-year cohort-sequential studies of adult intelligence. *Developmental Psychology, 19,* 531–543.

Sterns, H. L., & Sanders, R. E. (1980). Training and education in the elderly. In R. E. Turner & H. W. Reese (Eds.), *Life-span developmental psychology: Intervention* (pp. 307–330). New York: Academic Press.

Stone, V. (1980). *Structural modeling of the relations among environmental variables, health status and intelligence in adulthood.* Unpublished doctoral dissertation, University of Southern California, Los Angeles, CA.

Thurstone, L. L. (1948). *Primary mental abilities.* Chicago: University of Chicago Press.

Willis, S. L. (1985). Towards an educational psychology of the adult learner: Cognitive and intellectual 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., & Baltes, P. B. (1980). Intelligence in adulthood and aging: Contemporary issues. In L. W. Poon (Ed.), *Aging in the 1980s* (pp. 260–272). Washington, DC: American Psychological Association.

Received March 4, 1985
Revision received August 29, 1985 ■