Practice Effects in the Fluid Intelligence Performance of the Elderly: Longitudinal Followup and Replication

Michael Marsiske and Sherry L. Willis

Department of Human Development and Family Studies
Pennsylvania State University

Running Head: FLUID INTELLIGENCE PRACTICE EFFECTS

Paper presented at the 43rd Annual Scientific Meeting of the Gerontological Society of America, Boston MA, November 1990.

The research reported in this article was supported by grants R01-AG05304 and R01-AG08082 from the National Institute on Aging to S. L. Willis

Abstract

This study constitutes a longitudinal examination of the long-term effects of practice on the fluid ability performance of older adults. The returning sample included 7 males and 21 females, with a mean age of 74.57 years (S.D. = 4.75 years, range = 66-84 years). At Time One, following an extensive pretest of fluid and crystallized intelligence measures and measures of perceptual speed, subjects were assigned to either a practice condition or a control condition. Following the pretest, practice subjects received 8 or 10 hours of practice on two measures representing the fluid abilities of Figural Relations and Induction. Control subjects received no contact until the posttest, when both practice and control subjects were readministered the same measurement battery they had received at pretest.

The returning sample at Time Two included 18 practice subjects and 10 control subjects; a similar pretest-posttest control group design was used, with Time One practice subjects re-assigned to a practice condition at Time Two. There was some evidence for a significant residual effect of practice at Time One on the Induction performance of returning subjects at Time Two. Furthermore, when the number of practice sessions was held constant across the two retest episodes, there was no evidence of a significant difference in practice gains between the two occasions, despite the advanced age of the returning participants. Retest gain was small and steady at both Time One and Time Two, with no evidence of a performance asymptote after five practice sessions at Time Two.

The study findings support a broad literature which demonstrates cognitive plasticity in adult intellectual performance. Furthermore, in the present sample, there is little evidence for a restriction in plasticity across the seven-year interval studied.

Practice Effects in the Fluid Intelligence Performance of the Elderly: Longitudinal Followup and Replication

The study of intellectual change in late life has been a source of some controversy (Horn & Donaldson, 1976; Baltes & Schaie, 1976), for both methodological (Schaie, 1965) and conceptual (Barton et al., 1975) reasons. Even when significant decline in intellectual measures can be identified (e.g., Schaie, 1989), it may not be correct to attribute this decline solely to biological factors (e.g., Horn, 1982). There may also be environmental sources of intellectual decrement (Barton et al., 1975). If intellectual abilities showing decline are susceptible to environmental manipulations in late life, it follows that environmental factors may be partially implicated in the observed lower performances of some older adults (Baltes, 1987; Baltes & Lindenberger, 1988).

There is a growing body of literature which supports the notion of plasticity of intellectual functioning in late life, especially in the domain of fluid intelligence (Cattell, 1971). A number of short-term training studies which either attempted to instruct older adults in the rules required to successfully solve fluid intelligence measures or encouraged them to discover these rules for themselves have demonstrated significant improvement in the elderly's cognitive performance (e.g., Anderson, et al., 1986; Baltes et al., 1986; Baltes et al., 1988; Baltes & Willis, 1982; Blackburn & Papalia-Finlay, 1988; Blackburn et al., 1988; Blieszner et al., 1981; Denney & Heidrich, 1990; Hayslip, 1989; Plemons et al., 1978; Schaie & Willis, 1986; Willis et al., 1981; Willis & Schaie, 1986).

Hofland, Willis & Baltes (1981) and Hofland (1979, 1981) demonstrated that the fluid intelligence performance of older adults could be improved simply through the provision of no-feedback practice opportunities. Following eight or ten practice sessions with measures of Figural Relations and Inductive Reasoning (two fluid intelligence measures), the researchers demonstrated significant and disproportionate mean gain for subjects who received practice relative to controls who did not, on the measures they had practiced. Furthermore, there was evidence of significant transfer of practice to another measure of Inductive Reasoning on which subjects had not received practice (Hofland, 1981).

Hofland et al. (1981) also provided important information about the pattern of change produced under practice conditions. They reported that gains for older adults were small and steady across practice sessions, with the only significant difference between contiguous practice sessions occurring between the the pretest and the first practice session on the practiced measure of Figural Relations. They also reported that only a linear trend was significant in describing the pattern of practice change across sessions, suggesting that improvement was non-asymptotic. At Time One, there was also little evidence that practice had changed the correlational validity of the practiced measures: as a function of practice, the fluid intelligence measures did not show increasing correlation with crystallized measures. This argued against the notion that practice, as a process of acculturation and education, rendered the practiced measures better markers of crystallized intelligence than fluid intelligence (Donaldson, 1981).

Other studies have also demonstrated the effectiveness of practice as a cognitive intervention. Baltes et al. (1988) provided subjects with ten hours of practice on measures of fluid and crystallized intelligence, as well as perceptual speed. Gain for practice subjects was significant and comparable to that found for training subjects. In a subsequent study, Baltes et al. (1989) found few differences between subjects who received five hours of training in a fluid intelligence measure and subjects who received five hours of no-feedback practice with the training stimuli. Similar findings were reported by Blackburn & Papalia-Finlay (1988).

Taken as a whole, there is now fairly consistent evidence that environmental manipulations can improve the fluid intelligence performance of older adults, and that there is a substantial range of plasticity in the intellectual functioning of the elderly. As Willis & Nesselroade (in press) have noted, however, the developmental issue of the effects of advancing old age on intellectual plasticity has not received as much attention. None of the studies of practice have examined their long-term impact (beyond a six-month posttest), and none have examined the plasticity of intellectual functioning following practice when subjects are in their seventies and eighties.

The present study examined the long-term effects of the practice interventions reported by Hofland et al. (1981) and Hofland (1981). Experimental subjects had participated in eight or ten practice sessions at Time One (1979-1981), on fluid intelligence measures of Figural Relations and Inductive Reasoning. Subjects were re-examined at a longitudinal followup at Time Two, in 1987. Following a pretest at Time Two, experimental subjects again received five no-feedback practice sessions with the same measures.

This study will address three questions. The first question focuses on the long-term effects of practice. Can any residual benefit of practice at Time One be detected for returning practice subjects? The second question compares the magnitude of practice gain at Time One with the practice gain experienced by the older, returning subjects at Time Two. If aging is associated with increasing restrictions in the range of intellectual plasticity, will returning practice subjects demonstrate smaller practice gains at Time Two? The third question examines the pattern of change at Time Two. Will the pattern of change continue to be linear, incremental and will the correlational validity of the measures remain invariant across practice sessions?

Method

Subjects

The initial sample at Time One (1978-81) comprised 107 individuals. Of this initial sample, 28 subjects (7 males, 21 females) returned for participation at Time Two (1986). At Time Two, the mean age of the returning participants was 74.57 years (*S.D.* = 4.75 years, *range* = 66-84) and the mean educational level was 12.42 years (*S.D.* = 2.56 years, *range* = 7-16). The mean subject self-ratings for health, vision and hearing were 2.14 (*S.D.* = 0.76), 2.71 (*S.D.* = 0.98), and 2.64 (*S.D.* = 1.16) respectively.

Participants were Caucasian community-dwelling Pennsylvania residents. They were recruited from church groups and community centers; leaders of the groups and organizations were initially recruited, and interested leaders informed group members about the study. If subjects indicated interest, another presentation was made to discuss the study more fully. Subjects were paid at the rate of $2.00/hour at Time One; subjects who opted to return at Time Two were paid at the rate of $5.00/hour. Subjects could keep their payment, or have it donated to their community organizations.

Attrition effects

A 2 Group (Practice, Control) X 2 Attrition (Dropout, Return) between-subjects ANOVA was conducted on selected demographic and intellectual ability variables to examine Time One differences between returning and non-returning subjects. The demographic variables examined were gender, education, age, and self-ratings of health, vision and hearing. There was a significant attrition main effect for age: dropout subjects were significantly older than returning subjects (F = [1,98] = 10.76, p < .001). There was a significant group main effect for health, with experimental subjects rating their health as significantly better than controls (F [1,66] = 4.77, p < .05). The Attrition X Group interaction did not reach significance for any demographic variables.

For the intellectual ability variables a signficant attrition main effect was found for ADEPT Figural Relations (F [1,103] = 11.66, p < .001), Number Comparisons (F [1,103] = 8.20, p < .01), Induction Composite (F [1,103] = 4.65, p < .05), ADEPT Induction Composite (F [1,103] = 12.80, p < .001), and Culture Fair (F [1,103] = 9.38, p < .01), with returning subjects performing at a higher level than non-returning subjects. A significant group main effect was found for the ability variables of Vocabulary (F [1,103] = 4.09, p < .05), Ravens Progressive Matrices [F [1,103] = 4.04, p < .05) and Induction Composite (F [1,103] = 5.12, p < .05]. The practice group performed at a higher level than controls.

Although mean level differences at pretest were greater for returning subjects than for the total sample, the Attrition X Group interaction did not reach significance for any of the intellectual variables. Mean performance of returning and non-returning subjects is displayed in Table 1.

Insert Table 1 about here

Design and Procedure

The design used in this study was a longitudinal pretest-posttest control group design. Subjects initially participated in two 2-hour pretest sessions at Time One (1979-80). The pre- and posttest battery is described below. Subjects were then randomly assigned to the practice or control group. In the practice condition, subjects received eight or ten practice sessions with the Culture Fair test (a measure of Figural Relations) and the Induction Composite test (a measure of Inductive Reasoning ability). Control subjects received no contact. Following the practice intervention for the practice group, all subjects received a posttest battery. At Time Two (1987), subjects who returned again received similar pretest and posttest batteries. Practice subjects received five booster practice sessions between the pre- and posttests at Time Two.

Practice sessions. At both Time One and Time Two, all 8 or 10 practice sessions were conducted in small groups (12 or fewer subjects). Each session lasted approximately one hour, and was conducted according to standard test instructions and time limits. All retest sessions were conducted by a middle aged woman, both at Time One and Time Two.

Pretest and posttest sessions. All pretesting and posttesting was conducted in groups of 15 or fewer people. At Time One, all the testing sessions were conducted by graduate students, with the assistance of a proctor. At Time Two, all testing was conducted by women ranging from young-to-middle adulthood, usually in pairs, who functioned only as testers or proctors. To minimize fatigue, multiple rest breaks were provided within the testing sessions.

Measures.

The psychometric ability battery used in the present study was developed within the fluid (Gf) and crystallized (Gc) model of intelligence (e.g., Cattell, 1971). Multiple marker tests of Gf were included in the pre- and posttest batteries, as well as individual marker tests of Gc and Perceptual Speed. In addition, a personal data questionnaire was included in the measurement battery. The measures used in various phases of the study are outlined below.

Retest measures. Two primary mental abilities, Figural Relations (CFR) and Induction (I) were selected to represent the broad domain of Fluid Intelligence (Gf). These primary abilities had been identified in the literature (e.g., Horn & Cattell, 1966, Cattell, 1971) as relatively "pure" markers of Gf. The Culture Fair test (Cattell and Cattell, 1961) was selected to represent Figural Relations. Induction was represented by the Induction Composite test, which was a combination of Letter Sets (Ekstrom, French, Harman, & Derman, 1976), Letter Series, and Number Series (Thurstone, 1962).

Pretest and posttest batteries. A broad battery of Fluid (Gf), Crystallized (Gc), and Perceptual Speed (Ps) primary abilities was administered to all subjects.

The practice and control subjects received the battery outlined in Table 2 at the pretest and posttest at both Time One and Time Two. Practice effects on the primary abilities of Figural Relations and Inductive Reasoning were not measured only with the specific retest measures (Culture Fair and Induction

Composite, respectively), but with other measures of the same primary ability. Thus, the Figural Relations factor is further represented in the pre- and posttest battery by the ADEPT Figural Relations Diagnostic Test, and Raven's Progressive Matrices. Similarly, the Induction factor is further represented by the ADEPT Induction Composite tests: Letter Sets, Number Series and Letter Series. These tests are ordered along a theoretical continuum; retest effects should be strongest on the practice measures (near-near transfer). Secondarily, retest effects should be observed on other measures of the same latent ability, even though these measures were not practiced after the pretest. This represents near-fluid transfer. Finally, there are measures in the battery which do not represent either the same primary ability, or the same broad second-order ability. The Vocabulary measure assesses the crystallized ability of Verbal Comprehension. The Number Comparison measure assesses the speeded ability of Perceptual Speed. These would not be expected to show any ability-specific improvement as a function of practice on Induction and Figural Relations. Rather, any improvements observed on these measures were hypothesized to be due to improvement in generalized test-taking skills, and should be common to both experimental and control subjects.

Insert Table 2 about here

Subjects also completed a personal data questionnaire. In this questionnaire, subjects provided information about age, marital status, education, physical health, vision, and hearing ability.

Results

Data standardization. All Time One pretest measures were standardized to a mean of 50 and a standard deviation of 10. This permitted cross-test comparisons. The standardization procedure used total group means and standard deviations at Time One pretest as the base. By using total group (i.e., returning and non-returning subjects at Time One pretest; both practice subjects and controls) data as the standardization base, subgroup differences (i.e., between returning and non-returning subjects, between practice and control subjects) were preserved. Posttest data at Time One, and all Time Two data, were similarly standardized; each measure was scaled using Time One pretest scores of the total sample (returning and non-returning subjects, practice and controls) as the base. This facilitated comparison of the magnitude of practice-related and age-related change across measures.

Long-term maintenance of practice effects

Longitudinal change. To obtain an initial estimate of long-term practice effects, an overall analysis of covariance was conducted, using a mixed between-within design. The between subjects factor was Group (Practice, Control) and was crossed in a 2 X 7 design with the within-subjects factor of Measure (Culture Fair, Induction Composite, ADEPT Figural Relations, ADEPT Induction, Ravens Progressive Matrices, Vocabulary, Number Comparison). The ANCOVA was used to reduce the large group differences between the practice and control groups. The covariates were the pretest score at Time One on the measures of Culture Fair and Induction Composite. In this overall analysis, there was a trend toward a significant Group main effect (F [1,24] = 3.57,

$p < .10$), with control subjects showing more overall decline than practice subjects. Neither the main effect of Measure, nor the Measure X Group interaction reached significance.

Following this overall analysis, practice and transfer effects were examined more specifically at the individual test level. Additional followup univariate analyses of covariance (ANCOVAs) were conducted on the longitudinal change scores for each measure. Time One pretest Induction and Culture Fair performances were again used as the covariates, and Group (Practice, Control) was the between-subjects factor. These followup tests were conducted, despite the absence of significant main effects or interactions in the overall analysis, since long-term practice effects were only preducted for practice measures and possibly near-transfer measures.

There was a significant main effect of Group on longitudinal change in Number Comparisons (F [1,24] = 4.75, $p < .05$], suggesting an anomalous finding that practice subjects improved over the longitudinal interval, while control subjects declined. There was also a trend toward a significant main effect of Group on longitudinal change in ADEPT Induction (F [1,24] = 3.83, $p < .10$). Control subjects declined more than Practice subjects over the longitudinal interval.

Salience of practice for the longitudinal prediction of ability performance. Although the preceding analyses of variance and t-tests did not generally suggest that there were significant group differences in the magnitude of longitudinal change, followup stepwise regression analyses were conducted to examine the salience of group membership and Time One pretest-to-posttest change for the prediction of Time Two pretest performance level, controlling for Time One performance level and a number of personal variables.

Prediction of Time Two pretest performance level for each ability measure involved the use of a stepwise regression algorithm. The iterative procedure selected the best one-variable model from the predictors provided, then the best two-variable model, and so on, until the next largest model would contain non-significant regression parameter estimates. For each ability variable, the following predictors were provided to the program for possible inclusion in the final model: Group (practice or control), Time One pretest level, Time One pre-post gain, Gender, Education, and Age. The results of these regressions are presented in Table 3 below.

Insert Table 3 about here

For the crystallized ability of Vocabulary, and the two marker tests of Figural Relations (Culture Fair and ADEPT Figural Relations), the single best predictor of Time Two pretest level was Time One pretest performance. For Number Comparison, three predictors were significant: Time One pretest, educational level, and Group status. This result again appears to reflect the anomalously high Time Two pretest performance of experimental group subjects. For both of the Induction marker tests (ADEPT Induction Composite, and Induction Composite), two predictors were salient for the prediction of Time Two pretest performance level. The single best predictor was Time One pretest level. Additional significant variance was accounted for by Time One pretest-to-posttest retest gain. Although the group (practice, control) effect was non-significant, the significance of this retest variable (which had higher values for practice subjects than controls) suggests at least some salience for prior practice in the prediction of later performance level.

<u>Practice effects at Time Two</u>

To obtain an initial assessment of practice effects following booster retest sessions at Time Two, a 2 (Group: Practice vs. Control) X 2 (Occasion: Time 2 pretest, Time 2 posttest) X 7 (Measure) Analysis of Covariance (ANCOVA) was conducted. Pretest Induction Composite and Culture Fair performance at Time One were used as the covariates, to reduce large individual differences at Time One pretest. There were significant main effects of Group (F [1,24] = 5.92, $p < .05$), and Measure (F [6,19] = 2.61, $p < .05$). Neither the main effect of occasion nor the 2- and 3-way interactions reached significance.

Followup univariate repeated measures analyses of covariance were conducted on the separate Time Two pretest and posttest ability measures. Initial level of performance at Time One pretest on Induction Composite and Culture Fair were used as the covariates in each analysis, to reduce the large practice and control group differences. The design used was a two-factor ANCOVA approach, with Group (Practice, Control) as the between-subjects factor, and Occasion (Time Two pretest, Time Two posttest) as the within-subjects factor, with scores on each test as the dependent measure. There were significant group main effects for Culture Fair (F [1,24] = 5.82, $p < .05$) and Induction Composite (F [1,24] = 7.61, $p < .01$), with practice subjects outperforming controls on both measures. There was also a significant Occasion X Group interaction for the practiced measure of Induction Composite (F [1,24] = 14.67, $p < .001$) with a trend toward a significant Occasion X Group interaction for the other practice measure, Culture Fair (F [1,24] = 2.914, $p < .10$). For both variables, Practice subjects improved more than controls.

Comparison of Time One and Time Two retest gain. The magnitude of the retest gain at Time Two was compared to that obtained by the same subjects at Time One. To adjust for the fact that the pretest-to-posttest interval was not the same for Practice subjects at Time One and Time Two (8-10 sessions at Time One, compared with 5 at Time Two), the correlations between retest gains at Time Two and retest gains at Time One were examined following the first five practice sessions at Time One were obtained. There was some evidence for stability in retest gain across the longitudinal interval. For three of the four measures examined, correlations were positive and significant: Letter Sets ($r = .65$, $p < .05$), Letter Series ($r = .61$, $p < .05$), and Culture Fair ($r = .51$, $p < .05$).

A correlated samples t-test was conducted, with individual differences in pre-post change following five practice sessions at both Time One and Time Two as the dependent variable. None of the four ability variables examined (Culture Fair, Letter Sets, Number Series, or Letter Series) showed significant differences between Time One and Time Two in terms of retest gain. Figure 1 illustrates retest gain following 5 practice sessions at Time One and Time Two.

Insert Figure 1 about here

Pattern of practice change at Time Two.

Following up the work of Hofland et al. (1981), the pattern of change at Time Two was examined. Three related issues were considered: 1) Was the slope of change linear, or is there evidence for aymptotic performances being reached within the five practice sessions? 2) Was change across practice sessions small and incremental, as reported by Hofland et al., or did returning subjects display more discontinuous or erratic change patterns across practice sessions? 3) Was there any evidence for a change in the correlational validity of practiced fluid intelligence measures across practice sessions, such that correlations between the practiced measures and measures of crystallized intelligence (particularly) increased?

Slope of practice change. Linearity of change slopes connotes the absence of limits to plasticity over the interval studied; curvilinearity connotes asymptotic performance functions, and limits to plasticity. To examine the slopes of change across practice sessions, individuals sessions scores were regressed on orthogonal polynomial vectors representing linear, quadratic, and cubic time series. Only the linear trend was significant for Culture Fair (b = 0.64, t [1] = 3.04, $p < .01$, r^2 = .075), Letter Sets (b = 0.25, t [1] = 2.04, $p < .05$, r^2 = .035), and Number Series (b = 0.24, t [1] = 2.23, $p < .05$, r^2 = .042). No directional trend reached significance for the Letter Series test. Figure 2 illustrates changes across practice sessions at Time Two.

Insert Figure 2 about here

Magnitude of session-to-session change. To examine continuities and discontinuities in practice change, the means of contiguous practice sessions were compared. Changes from session to session were small, and generally positive. For Culture Fair, there were significant gains between Pretest and Practice Session 1 (t [11] = 2.60, $p < .05$), Practice Session 1 and 2 (t [17] = 2.15, $p < .05$), Practice Session 3 and 4 (t [17] = 2.88, $p < .01$).

For Number Series, significant gains occurred between the Pretest and Session 1 (t [11] = 3.32, $p < .01$).

On the Letter Series test, significant gains occurred between Practice Sessions 3 and 4 (t [17] = 2.62, $p < .05$) and Practice Session 5 and the posttest (t [14] = 2.35, $p < .05$)

Correlates of practice session performance

Organismic and intellectual factors thought to be associated with performance in the Time Two practice sessions were correlated with ability scores across the five practice sessions. Prior intellectual performance (as measured at the Time One pretest) was significantly and positively associated with performance in each of the practice sessions (ranging from r = ·80 between Session 2 and the Time One pretest for Culture Fair, to r = .34 between practice Session 1 and the Time One pretest for Letter Sets). Correlations between session scores and Time One pretest Vocabulary and Number Comparisons were also examined. None of these correlations, across sessions, were significantly greater than zero. In this small sample, neither Age nor Education were significantly associated with within-session performance. Comparison of session performance correlations with baseline ability performance across practice sessions, using a formula for the comparison of dependent correlations with a common predictor (Ferguson, 1953), revealed no significant differences in correlation between contiguous practice sessions, or between the first and the last practice session. Thus, there was sizable interindividual stability in rank ordering of ability performance across the 5 practice trials.

Discussion

This study provides another positive perspective on the cognitive plasticity of the old old. It suggests that a minimal intervention (no feedback practice) can exert significant effects on the intellectual performances of old adults, even in the mid-seventies. The following section summarizes and discusses the key findings of this study.

Salience of practice for the longitudinal prediction of ability performance

Practice at Time One can be considered in one of two ways. It can be used to describe a dichotomous categorization variable: one group received multiple retest sessions, while control subjects did not. Practice can also be used to describe retest gain at Time One, as reflected in pre-post score increases at Time One. In this latter sense, both "practice" and "control" subjects received practice. While only the "practice" subjects received multiple practice opportunities, the control subjects received practice from the pretest to the posttest; by retaking the tests at posttest, even the control subjects could be expected to show some increase over their pretest level.

The guiding question for this set of analyses was as follows: controlling for other likely predictors of Time Two performance level, does practice at Time One exert a significant effect on Time Two performance level, controlling for other predictors? Overall, few predictors reached the conventional ($p < .05$) criterion of significance for any ability variable, undoubtedly another consequence of low statistical power. For all ability measures studied, the Time One pretest was the single best predictor of the Time Two pretest (with R^2 values ranging from 0.72 for both Induction measures, to 0.42 for the Culture Fair). Interestingly, for both Induction measures (Induction Composite and ADEPT Induction), Time One pre-post change was also a significant predictor of Time Two performance level, even after initial Time One pretest was controlled for. No Time One demographic variables reached significance as predictors of Time Two performance level. The additional variance (above that explained by the Time One pretest) explained by Time One retest gain was .10 for the Induction Composite, and .04 for ADEPT Induction.

These regression results are interesting in several ways. First, they confirm the findings from the previous analyses of variance, that for Induction, there may be some long-term residual effect of practice. It is important to note, however, that this effect of practice is not directly tied to the number of sessions subjects participated in, but to the magnitude of gain they experienced. Although the mean retest gain at Time One was greater for practice subjects than controls, there was clearly individual variability, with some practice subjects gaining less than some controls. For both measures of Inductive Reasoning the greater the magnitude of this retest gain, the higher the Time Two pretest score was, controlling for initial level of performance.

It is interesting that the majority of these subjects were drawn from Hofland's (1981) dissertation sample. In that study, Hofland reported significant practice effects for the near-near transfer measures of Induction Composite and Culture Fair, as well as the near transfer measure of ADEPT

Induction. There was also a trend ($p < .06$) toward a significant practice effect for ADEPT Figural Relations. The present findings regarding residual effects of the Time One practice interventions reported by Hofland suggest a similar pattern; these results suggest, at least tentatively, that practice exerts stronger effects on Induction measures than Figural Relations measures, and that these effects have some detectable durability 6-to-8 years after the intervention. Confidence in this assertion is strengthened by the fact that both Induction measures still show significant relationships with Time One retest gain.

Note that the presence of selective attrition effects may be a threat to the external validity of the longitudinal results above, if the age-performance functions obtained no longer represent those of the original parent population (Schaie, Labouvie & Barrett, 1973), and also to the internal validity, since models which fail to account for sample selectivity are inherently misspecified (Berk, 1983). Fortunately, since the Attrition by Experimental Group interaction did not reach significance in this study, there is reason to believe that group differences can still be examined; that is, although both practice and control subjects who returned were more advantaged than non-returning subjects, there is no statistical evidence that either experimental group has become disproportionately advantaged as a function of attrition. This assumption must, of course, be questioned by the low statistical power of the attrition X group analyses, in the face of low sample sizes

Practice effects at Time Two

An overall analysis of covariance comparing practice and control subjects Time Two retest change found significant main effects of group (reflecting the overall superiority of the practice subjects) and measure (reflecting some hierarchical transfer pattern in training gain), but found no overall effects of occasion, nor were any two- or three-way interactions significant.

Examining Occasion (Time Two pretest, Time Two posttest) by Group (Practice, Control) effects separately for each measure (since Occasion by Group effects were only hypothesized for practiced measures), there were significantly larger gains for practice subjects than controls on the practiced Induction Composite measure. There was also a trend ($p < .10$) toward a significant practice group superiority in pre-post gain on the practiced Culture Fair measure. There was no significant group difference in pre-post gain on any other measure. Again, power is an important consideration. Even for a strong effect, Cohen (1988) reports a maximum power of $B = .54$ to detect an effect significant at $p < .05$ in this design.

This pattern of practice transfer is narrower than even that found by Hofland (1981) at Time One. At Time Two, the evidence is consistent with the earlier findings (stronger practice effect for Induction than Figural Relations), but the band of transfer is even narrower than that obtained at Time One.

One possible reason, beside low statistical power, for the even narrower band of transfer at Time Two than at Time One is the smaller number of retest sessions given at Time Two (five sessions at Time Two compared with eight-to-ten at Time One).

<u>Comparison of Time One and Time Two retest gain</u>

Given the narrower band of transfer at Time Two, a comparison of retest gains at Time One and Time Two was undertaken. It is not immediately clear, however, whether this narrower transfer at Time Two is attributable to age-related declines in cognitive plasticity, or whether it is attributable to the smaller number of practice sessions at Time Two. Statistical power is also a consideration. In an attempt to help resolve the relative contribution of advanced age and a decreased number of sessions, retest gains following the first five practice sessions at Time One were compared with retest gains at Time Two.

The results suggest that gain on practiced measures at Time One was positively correlated with Time Two gain when only the first five sessions were compared, suggesting at least moderate interindividual stability of retest gain across the longitudinal interval. Correlations reported on such a small sample must be viewed carefully, however. More directly, a correlated samples t-test examining individual differences in pre-post change following five practice sessions at both Time One and Time Two failed to find a significant difference for any of the four nearest transfer measures: Culture Fair, Letter Sets, Number Series, and Letter Series. These results provide limited support for the notion that smaller overall retest gains for practice subjects at Time Two is primarily due to the difference in the number of practice sessions at Time One and Time Two, and has less to do with age-related restrictions in the range of cognitive plasticity. Again, however, sensitivity to differences must be questioned. The maximum $\underline{B}$ for this design under the assumption of a strong effect is 0.69 (Cohen, 1988).

Slope of practice change

Hofland (1981) found steady increments for practice subjects across 10 practice trials, with slight evidence for a performance asymptote at Practice Session 9 for Induction, and at Practice Session 10 for Culture Fair. Since the current study was a longitudinal followup, in part, of the Hofland studies, the linearity of the improvement slope across the five practice sessions at Time Two was again examined. In the present study only the orthogonal polynomial representing the linear trend yielded a significant parameter estimate for three of the four retest measures : Culture Fair, Letter Sets, and Number Series. No directional trend reached significance for the Letter Series test.

These results provide little evidence for a restriction in the range of cognitive plasticity over the longitudinal interval. Since there was some evidence for performance asymptotes toward the end of the practice trials at Time One (Hofland, 1981), one could argue that an asymptote might have been reached at an earlier trial for the older returning subjects, if their range of improvement has become more restricted. There was no evidence for a performance asymptote at Time Two, arguing against a developmental restriction of plasticity.

Magnitude of session-to-session change

Hofland (1981) found that the only two consecutive data points to differ significantly at Time One were the Time One pretest and Practice Session 1; that is, the largest increment in performance occurred between the pretest and the first practice session, with subsequent session-to-session practice gains being small enough not to reach the criterion of significance.

The means of contiguous practice sessions were also compared at Time Two in this study via t-tests; a less clear pattern regarding the pattern of practice gains was obtained. While both Culture Fair and the Number Series test showed significant gains between the pretest and Practice session 1, the Culture Fair test also showed significant gains between Practice Sessions 1 and 2, and Practice Sessions 3 and 4. Conversely, the Letter Series tests showed significantly different means between contiguous practice sessions only toward the end of the practice trials: increases were significant between Sessions 3 and 4, and between Session 5 and the posttest.

It is not immediately clear whether this more dispersed pattern of practice gain at Time Two reflects a developmentally meaningful change. It may suggest that the older returning subjects are slower to profit from the cognitive activation/competence utilization (Overton & Newman, 1982) opportunities that practice affords. This would certainly argue against interpreting one-shot assessments of older adults, especially those in their seventies and eighties, as true reflections of their underlying cognitive competence.

Correlates of practice session performance

Donaldson (1981) has argued that cognitive training interventions may render fluid intelligence measures, like Inductive Reasoning and Figural Relations, more crystallized. Repeated exposure to fluid intelligence stimuli may transform the tasks from novel situations requiring the application of problem-solving skills (prototypical fluid intelligence task) to automated responses on familiar, accultured stimuli (prototypical crystallized intelligence task). Addressing this question, Hofland, Willis and Baltes (1981) found little evidence for a change in correlational stability of the retest measures across eight practice trials. In the present study, there was also no evidence for a change in the correlational validity of the practiced measures across practice trials. The very low sample size for these correlations may be implicated in these findings at Time Two.

Relationship of study findings to previous research

The data in the present study attempt to extend numerous findings from the Hofland ADEPT Studies (Hofland, Willis, & Baltes, 1981; Hofland, 1981).

The present study confirmed the results of many other longitudinal studies of cognitive aging, finding that returning subjects were significantly more advantaged, relative to dropouts, on the measured intellectual ability variables (Baltes, Schaie, & Nardi, 1971; Cooney, Schaie, & Willis, 1988; Jarvik & Falek, 1963; Marsiske & Willis, 1989; Schaie, Labouvie, & Barrett, 1973; Siegler & Botwinick, 1979). These findings also support findings from previous cognitive intervention research, in that substantial plasticity of cognitive performance was demonstrated (e.g., Baltes, Dittmann-Kohli & Kliegl, 1986, Blackburn, Papalia-Finlay, Foye, & Serlin, 1988; Blieszner, Willis & Baltes, 1981; Willis, Blieszner & Baltes, 1981; Willis & Schaie, 1986). More recently, Willis and Nesselroade (in press) have demonstrated that subjects gained as much from Figural Relations training when they were in their late seventies as they did when they were in their late sixties. The present findings support the same general idea: when the number of practice sessions is controlled for, practice subjects did not gain significantly less in their late seventies as they did in their late sixties.

The present study found a narrower band of transfer than has been found in previous training studies (e.g., Baltes, Dittmann-Kohli & Kliegl, 1986; Blieszner, Willis, & Baltes, 1981; Willis, Blieszner, & Baltes, 1981), or even in previous practice studies (Hofland, 1981). At least three complementary hypotheses must be entertained as an explanation. It may be possible that the advanced age of the participants at Time Two has exerted an effect on the extent to which they can generalize the skills they activate/acquire in the course of practice. It may also be possible, however, that the smaller number of practice sessions at Time Two (five compared to eight or ten) is responsible for the reduced transfer. If subjects had had an equal number of practice opportunities at Time Two, their transfer may have been equally broad. Finally, it is possible that the smaller number of subjects at Time Two, relative to Time One, may have reduced the statistical power sufficiently to make the design insensitive to smaller transfer effects to non-practiced measures of the practiced abilities.

The current results also agree with Hofland (1981), in that there was no evidence of an asymptote of performance after five practice sessions. At Time One, Hofland did not find a levelling of performance increases before the second-last or last of the ten practice trials. While age-related losses in cognitive plasticity (Baltes, 1987) have been hypothesized from testing-the-limits research on the methods-of-loci memory strategy with young and old adults, the current analyses find no evidence of a reduction of in cognitive plasticity even as subjects reach their mid-seventies and eighties, using a longitudinal method.

Conclusions

This study has provided further support for the notion that simple practice is a powerful cognitive intervention. As a function of five opportunities to practice on measures of Figural Relations and Induction, practice subjects showed a total gain of 8/10 of a standard deviation on the practiced measure of Culture Fair, and a total practice gain of about 7/10 of a standard deviation on the practice measure of Induction Composite.

This study also provides some preliminary support for the idea that long-term, residual effects of practice interventions may actually affect the developmental trajectory of some practiced abilities. In this study, subjects showed small but significant positive effects of practice gain on Induction measures on their performance on those measures six-to-eight years later. No similar effect was found for Figural Relations, even though this was another practiced ability.

Most importantly, this study provides further positive support for the cognitive plasticity of older adults, and suggests that very minimal interventions (no-feedback practice, in this study) can produce improvements of about 3/4 of a standard deviation on practiced measures.

Table 1

Returning and dropout subjects' standardized means, by experimental group, at Time One and Time Two

	Time One					
	Practice Group			Control Group		
	Pretest		Post	Pretest		Post
	Ret[a] n = 18	Drop[b] n = 54	Ret n = 18	Ret n = 10	Drop n = 25	Ret n=10
Culture Fair[c]	57.33 (7.02)	48.98 (10.18)	69.39[c] (8.15)	52.40 (7.32)	47.12 (10.63	55.10 (6.10)
Induction[d]	57.07 (13.21)	49.89 (7.80)	70.57[d] (13.60)	49.67 (8.02)	47.76 (8.60)	54.73 (7.12)
ADEPT Figural Relations	57.28 (8.61)	49.02 (9.51)	64.50 (8.19)	53.10 (6.49)	46.52 (10.51)	58.50 (5.80)
ADEPT Induction	57.00 (9.96)	49.11 (7.22)	66.91 (13.77)	52.40 (6.51)	46.65 (9.48)	53.00 (7.47)
Ravens Matrices	55.11 (11.51)	49.91 (10.21)	59.17 (7.87)	47.30 (5.52)	48.28 (10.31)	56.10 (8.70)
Vocabulary	55.11 (8.32)	49.67 (9.38)	59.94 (6.64)	47.60 (11.67)	48.12 (10.36)	50.90 (8.77)
Number Comparison	55.00 (9.10)	48.76 (8.84)	59.83 (7.91)	53.90 (13.78)	47.40 (9.71)	56.60 (12.12)

(cont. on next page)

Table 2 (cont.)

Time Two				
	Practice Group n = 18		Control Group n = 10	
	Pretest	Posttest	Pretest	Posttest
Culture Fair	56.38 (5.54)	64.44 (7.96)	51.50 (5.66)	54.20 (5.96)
Induction	56.20 (8.43)	63.28 (9.39)	49.30 (5.58)	50.00 (6.36)
ADEPT Figural Relations	57.22 (8.02)	59.83 (7.06)	50.10 (6.66)	56.30 (6.72)
ADEPT Induction	55.33 (10.69)	59.44 (11.24)	48.07 (4.85)	50.20 (7.46)
Ravens Matrices	52.78 (5.59)	56.06 (9.66)	52.50 (8.66)	53.00 (7.12)
Vocabulary	56.28 (7.54)	57.11 (5.86)	47.10 (9.43)	49.90 (10.31)
Number Comparison	61.72 (17.87)	54.00 (20.52)	46.80 (13.34)	48.70 (11.34)

Standard deviations are in parentheses.

a Ret = Returning subject
b Drop = Non-returning subjects
c After first five practice sessions, mean = 62.80, SD = 7.11
d After first five practice sessions, mean = 66.00, SD = 12.32

Table 2

Pretest and posttest assessment batteries

General Dimension	Primary Ability	Marker Measures	Source
Fluid	Induction	Induction Composite Test Letter Sets Number Series Letter Series	Ekstrom et al., 1976; Thurstone, 1962
		ADEPT Induction Test, Form A Form B Letter Sets Number Series Letter Series	Blieszner, Willis & Baltes, 1981
	Figural Relations	Culture Fair Test (Scale 2, Form A) and Power Matrices [Scale 3, Form A (1963 ed.) and Form B (1961 ed.)]	Cattell & Cattell, 1957; Cattell & Cattell, 1961, 1963
		ADEPT Figural Relations Diagnostic Test Form A[a] Form B[b]	Plemons, Willis, & Baltes, 1978
		Raven's Advanced Progressive Matrices (Set II)	Raven, 1962
Speed	Perceptual Speed	Number Comparison	Ekstrom et al., 1976
		Identical Pictures	Ekstrom et al., 1976
Crystal-lized	Verbal Comprehension	Vocabulary, V-2	Ekstrom et al., 1962

Table 3

Summary of stepwise regression results for the prediction of Time Two pretest ability performance level

Dependent Variable:	Selected Independent Variable	*b*-value	*F*-ratio	*p*	partial R^2
Vocabulary	Time One pretest[a]	0.756	56.47	.0001	0.68
Culture Fair	Time One pretest	0.521	18.52	.0002	0.42
ADEPT Figural Relations	Time One pretest	0.784	37.80	.0001	0.59
Number Comparisons	1. Group	-11.586	4.27	.05	0.17
	2. Time One pretest	0.737	8.38	.008	0.14
	3. Educ.	-2.812	6.21	.02	0.14
Induction Composite	1. Time One pretest	0.658	116.44	.0001	0.72
	2. Time One pre-post change[b]	0.225	14.57	.0008	0.10
ADEPT Induction Composite	1. Time One pretest	0.827	56.92	.0001	0.72
	2. Time One pre-post change	0.278	4.31	.05	0.04

[a] Pretest predictors were measured at the Time One pretest
[b] Pre-post changes reflect Time One pretest-to-posttest change

References

Anderson, J. W., Hartley, A. A., Bye, R., Harber, K. D., & White, O. L. (1986). Cognitive training using self-discovery methods. Educational Gerontology, 12, 159-171.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. Developmental Psychology, 23, 611-626.

Baltes, P. B., Dittmann-Kohli, F., & Kliegl, R. (1986). Reserve capacity of the elderly in aging-sensitive tests of fluid intelligence: Replication and extension. Psychology and Aging, 1, 172-177

Baltes, P. B., Kliegl, R., & Dittmann-Kohli, F. (1988). On the locus of training gains in research on the plasticity of fluid intelligence in old age. Journal of Educational Psychology, 80, 392-400.

Baltes, P. B., & Lindenberger, U. (1988). On the range of cognitive plasticity in old age as a function of experience: 15 years of intervention research. Behavior Therapy, 19, 283-300.

Baltes, P. B., & Schaie, K. W. (1976). On the myth of intellectual decline in adulthood. American Psychologist, 31, 701-724.

Baltes, P. B., Schaie, K. W., & Nardi, A. H. (1971). Age and mortality in a seven-year longitudinal study of cognitive behavior. Developmental Psychology, 5 18-26.

Baltes, P. B., Sowarka, D., & Kliegl, R. (1989). Cognitive training research on fluid intelligence in old age: What can older adults achieve by themselves? Psychology and Aging, 4, 217-221.

Baltes, P. B., & Willis, S. L. (1982). Plasticity and enhancement of intellectual functioning in old age. In F. I. M. Craik & S. Trehub (Eds.). Aging and Cognitive Processes (pp. 353-389). New York: Plenum Press.

Barton, E. M., Plemons, J. K., Willis, S. L., & Baltes, P. B. (1975). Recent findings on adult and gerontological intelligence. American Behavioral Scientist, 19, 224-236.

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. American Sociological Review, 48, 386-398.

Blackburn, J. A., & Papalia-Finlay, D. (1988). Modifiability of figural relations performance: A comparison of young-old and old-old elderly adults. Paper presented at the 41st Annual Scientific Meeting of the Gerontological Society of America, San Francisco, CA. November, 1988.

Blackburn, J. A., Papalia-Finlay, D., Foye, B. F., & Serlin, R.C. (1988). Modifiability of figural relations performance among elderly adults. Journal of Gerontology: Psychological Sciences, 43, 87-89.

Blieszner, R., Willis, S. L., & Baltes, P. B. (1981). Training research in aging on the fluid ability of inductive reasoning. Journal of Applied Developmental Psychology, 2, 247-265.

Cattell, R. B. (1971). Abilities: Their structure, growth, and action. Boston: Houghton-Mifflin.

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

Cattell, R. B., & Cattell, A. K. S. (1961). Measuring intelligence with the Culture Fair tests: Manual for Scales 2 & 3. Champaign, IL: Institute for Personality and Ability Testing.

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

Cooney, T. M., Schaie, K. W., & Willis, S. L. (1988). The relationship between prior functioning on cognitive and personality dimensions and subject attrition in longitudinal research. Journal of Gerontology: Psychological Sciences, 43, P12-P17.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences, Second edition. Hillsdale, NJ: Erlbaum.

Denney, N. W., & Heidrich, S. M. (1990). Training effects on Ravens Progressive matrices in young, middle-aged, and elderly adults. Psychology and Aging, 5, 144-145.

Donaldson, G. (1981). Letter to the Editor. Journal of Gerontology, 36, 634-638.

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

Ferguson, G. A. (1953). Statistical Analysis in Psychology and Education. New York: McGraw-Hill.

Hayslip, B. (1989). Fluid ability training with aged people: A past with a future? Educational Gerontology, 15, 573-595.

Hofland, B. F. (1979). Intraindividual variability in fluid intelligence: Retest effects of younger and older adults. Unpublished Masters Thesis, Pennsylvania State University.

Hofland, B. F. (1981). Practice effects in the intellectual performance of the elderly: Retesting as an intervention strategy. Unpublished Doctoral Dissertation, Pennsylvania State University.

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

Horn, J. L. (1982). The aging of human abilities. In B. B. Wolman (Ed.). Handbook of Developmental Psychology. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Horn, J. L., & Donaldson, G. (1976). On the myth of intellectual decline in adulthood. American Psychologist, 31, 701-724.

Jarvik, L. F., & Falek, A. (1963). Intellectual stability and survival in the aged. Journal of Gerontology, 18, 173-176.

Marsiske, M., & Willis, S. L. (1989). Selective attrition effects in a longitudinal study of adult intelligence: Methodological considerations. Paper presented at the 42nd Annual Scientific Meeting of the Gerontological Society of America, Minneapolis, MN. November, 1989.

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

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

Raven, J. C. (1962). Advanced progressive matrices, Set II (1962 Rev.). London: H. K. Lewis.

Schaie, K. W. (1965). A general model of the study of developmental problems. Psychological Bulletin, 64, 92-107.

Schaie, K. W. (1989). The hazards of cognitive aging. Gerontologist, 29, 484-493.

Schaie, K. W., Labouvie, G. V., & Barrett, T. J. (1973). Selective attrition effects in a fourteen-year study of adult intelligence. Journal of Gerontology, 28, 328-334.

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

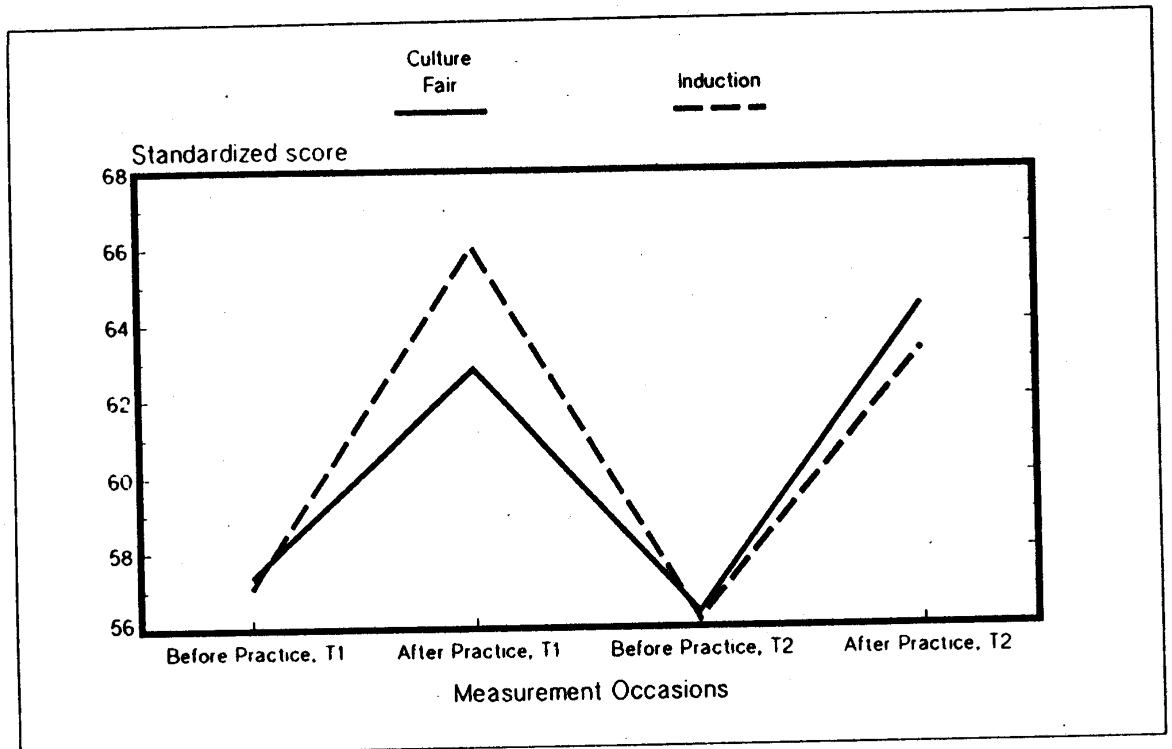
Siegler, I. C., & Botwinick, J. (1979). A long-term longitudinal study of intellectual ability in older adults: The matter of selective subject attrition. Journal of Gerontology, 34, 242-245.

Thurstone, T. G. (1962). Primary mental abilities for Grades 9-12 (Rev. ed.). Chicago, IL: Science Research Associates.

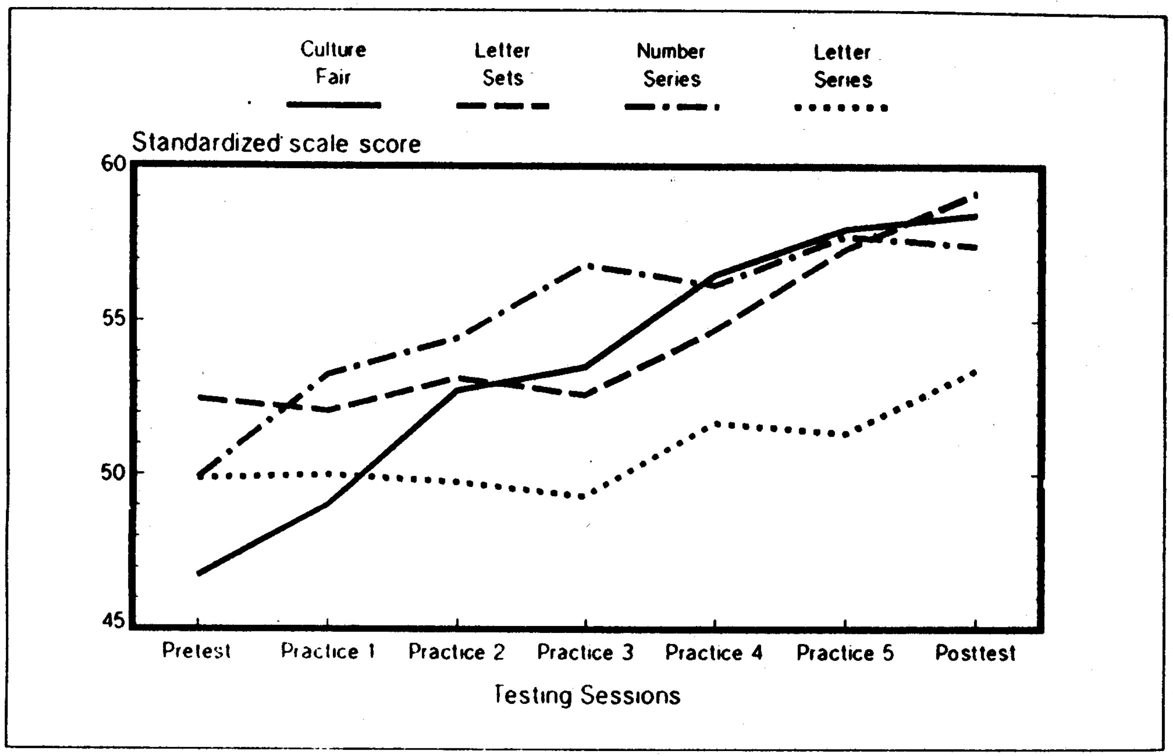
Willis, S. L., Blieszner, R., & Baltes, P. B. (1981). Intellectual training research in aging: Modification of performance on the fluid ability of figural relations. Journal of Educational Psychology, 73, 41-50.

Willis, S. L., & Nesselroade, C. S. (in press). Longterm effects of fluid ability training in old-old age. Developmental Psychology.

Willis, S. L., & Schaie, K. W. (1986). Training the elderly on the ability factors of spatial orientation and inductive reasoning. Psychology and Aging, 1, 7-12.



<u>Figure 1.</u> Change in performance after 5 practice sessions: Time One and Time Two



<u>Figure 2.</u> **Performance across practice sessions: Time Two, standardized scores**