Effects of Previous Assessment in the Study of Adult Intellectual Ontogeny[1]

K. Warner Schaie, Gisela V. Labouvie, and T. J. Barrett

West Virginia University

Paper presented at the

24th Annual Meeting of the

Gerontological Society,

Houston, October, 1971

[1]Preparation of this paper was facilitated by a grant from the National Institute of Child Health and Human Development (HD 044766-02).

Effects of Previous Assessment in the Study of Adult Intellectual Ontogeny

K. Warner Schaie, Gisela V. Labouvie, and T. J. Barrett
West Virginia University

Abstract

It is argued that the study of testing effects in intellectual ontogeny by means of designs involving posttest-only control groups will be confounded by the fact that retest participants are a biased sample from the original parent population. Two analyses were performed on subjects aged 32 to 74 years, both comparing retested samples with samples tested for the first time (controls). The first analysis involves no attempt to equate retestees and controls in terms of selective attrition, while the second design does incorporate such an attempt. Significant effects or prior testing were obtained in the first analysis only. These results suggest that "testing" effects may be artifacts caused by the phenomenon of selective attrition.

Effects of Previous Assessment in the Study of Adult Intellectual Ontogeny

K. Warner Schaie, Gisela V. Labouvie, and T. J. Barrett

West Virginia University

During the past decade, considerable interest has been directed to the problem of developing adequate methodologies for the assessment of ontogenetic change (cf., Baltes, 1968; Schaie, 1965; Wohlwill, 1970). Originally stimulated by discrepant findings from cross-sectional and longitudinal findings (cf., Kuhlen, 1963; Schaie, 1965), these discussions have demonstrated that both methods are fraught with such a severe lack of control over error factors such as generation differences, and testing and selection effects, that it has become increasingly hazardous to interpret resulting age-performance functions as indications of ontogenetic change.

Emphasizing the distinction between individual (ontogenetic) and generational change, both Baltes (1968) and Schaie (1965) have therefore proposed to combine series of cross-sectional and short-term longitudinal studies in complex sequential designs aimed at differentiating the effects of chronological age and generation (cohort) differences. Applications of such strategies in the area of intellectual ontogeny--based on comparisons between longitudinal gradients of single cohorts, and cross-sectional gradients across many cohorts--so far have clearly substantiated the over-riding effect of generational differences over that of differences related to chronological age (e.g., Nesselroade, Schaie, & Baltes, 1971; Riegel,

Riegel, & Meyer, 1967; Schaie, 1970, 1971b; Schaie & Strother, 1968).

The validity of inferences based on the comparison between cross-sectional and longitudinal gradients is, however, jeopardized to the extent that changes in longitudinal patterns from one time of measurement to the next may not reflect genuine developmental change, but may be artifactual to the extent that the measurement operation itself has served to modify the trait whose development we wish to observe. The control of such testing effects (e.g., Baltes, 1968; Campbell & Stanley, 1963; Schaie, 1965, 1971a, 1972b) requries the incorporation of control groups sampled from the same base population but tested on a single occasion only.

It is the purpose of the present study to report the application of such designs involving posttest-only control groups in order to control for the effect of repeated assessment in adult intellectual ontogeny. In addition, the present study incorporates another important feature so far neglected in discussions of the control of testing effects (see, however, Schaie, 1971a). That is, we know from the longitudinal literature that random samples do not maintain their original sampling characteristics with respect to psychological variables (cf. Baltes, Schaie, & Nardi, 1971; Schaie, 1972a), but that attrition operates selectively to leave positively biased samples. The present paper, therefore, will also include an attempt to equalize the control group with the pretested groups in terms of such attrition characteristics.

Method

Deisgn. In order to unconfound the sources of variance which artifactually may suggest adult intellectual change it is necessary to assess

two or more cohorts (generations) at two or more measurement points to obtain samples that at one time of measurement have been tested either for the first or the second time (cross-sequential method). Alternatively, a design might use a breakdown by age (time-sequential method) if the assumption is warranted that ontogenetic trends account for a stronger variance component than differences between generations and transient secular trends. Since the literature reviewed strongly substantiates the impact of cohort differences, a cross-sequential design is adapted for the present study. A sampling plan and the corresponding analysis of variance model are presented in Table 1.

To further differentiate the effects of prior experience and selective attrition, it is necessary to attempt to equate unpretested control subjects and pretested subjects with regard to attrition characteristics. With the three measurement points (T_1, T_2, and T_3 in Table 1) now available in this study, it is possible to perform an analysis on the T_2 data by taking the retest scores of subjects first tested at T_1 and followed up at T_2 (sample $S_aO_2T_2$ in Table 1) and only that subsample of subjects first tested at T_2 (subsample of $S_bO_1T_2$) for which retest data are available at T_3.

Subjects. All Ss were members of a pre-paid medical plan in a metropolitan area of the Pacific Northwest, with a population base of approximately 18,000 members at the time of initial data collection. Detailed accounts of the sampling plan and procedures have been reported elsewhere (Schaie, 1958, 1959). In summary, quota sampling was conducted in 1956 for each 5 year interval from 21 to 70 years of age and in 1963 from 21 to 75 years of age with approximately equal of men and women in each age interval. In addition approximately 60% of the 1956 sample was

retested in 1963. Because of the seven year testing interval the samples have now been reorganized into seven year cohorts with mean years of birth ranging from 1889 to 1945.

In 1970 a second follow-up was conducted in which the residual samples from the 1956 and 1963 studies were retested. Also new random samples were drawn from the parent population for all cohorts tested previously plus the next younger one. As a consequence these are now available repeated measurement data for all three data points for a sample of 162 Ss now ranging in mean age from 21 to 84 years, for two data points (1956-63) a sample of 300 Ss ranging in age from 28 to 77 years and another sample (1963-70) of 409 Ss ranging in age from 28 to 84 years. Single point independent random sampling data are available for the 1956 series on 490 Ss, for the 1963 series on 960 Ss, and for the 1970 series on 701 Ss. Although attrition in the repeated measurement samples has appeared to be random with respect to most socio-economic variables, subject loss has been found to be biased with respect to the psychological variables (cf. Baltes, Schaie, & Nardi, 1971).

Measurement Variables. The SRA Primary Mental Abilities Test (PMA), Schaie's Test or Behavioral Rigidity, a socioeconomic status questionnaire, and a survey of satisfaction with the pre-paid medical plan were administered in group sessions handling from 10 to 50 subjects and lasting approximately two hours each. The present report deals with the PMA data only. Specifically, this test includes five factors: Verbal Meaning (V), Space (S), Reasoning (R), Number (N), and Word Fluency (W). In addition, a composite measure of IQ was included as a sixth dependent variable.

Data Analysis. To permit cross-scale comparisons, raw scores were converted to T-scores (mean: 50, standard deviation: 10), using as reference

the first test administration for all three times of measurement. From the available data subsamples were then combined for a cross-sequential analysis of variance involving the independent measurement factors of Cohort (7 levels), Time (1963, 1970), Prior Testing (1st vs. 2nd test), and Sex.

In addition, a cross-sequential analysis for the single time of measurement 1963 was performed for subjects previously tested in 1956, and for those subjects first tested in 1956 who were known to have returned for a second test in 1970. This analysis involved the factors Cohort (7 levels), Prior Testing (1st vs. 2nd test), and Sex. This analysis attempts to equate the two sets of samples for effects of experimental mortality.

Results

Summary results for the main cross-sequential analysis are presented in Table 2. Main effects for Cohort are obtained for all variables ($p < .01$), and main effects of Sex for all variables except IQ ($p < .01$). Only two interactions with Sex reach significance at the .05 level (Cohort by Time by Sex for Verbal Meaning, and Cohort by Sex for Word Fluency). A significant Time of Measurement Effects is obtained for Verbal Meaning only ($p < .05$).

Of primary interest in the present context, however, are effects involving Prior Testing. Mean scores for subjects tested previously are raised significantly over those of subjects tested first on all variables (see Table 4). Three variables (N, W, and IQ) show an Cohort by Prior Testing interaction indicating that the testing effect is somewhat stronger for the older than the younger cohorts. For four variables (V, R, W, IQ), furthermore, there is an interaction of Cohort, Prior Testing and Time of Measurement. Consequently, differential Cohort by Practice interactions are obtained for the 1963 and 1970 Time of Measurement. Inspection of the data reveals that this trend is

consistent for all four variables. That is, for the 1963 data the retest effect is largest for the middle cohorts (approximately 45 to 60 years of age) while for the 1970 data, these cohorts show the smallest retest effect.

Before accepting and interpreting these results it is necessary to consider the outcome of a second analysis, which attempts to equalize the pretested and unpretested groups for effects of selective attrition. Summary results for this analysis are presented in Table 3. Cohort main effects again are obtained for all variables ($p < .01$), and Sex main effects for S and W at the .01 level and for N at the .05 level. A Cohort by Sex interaction is obtained for W.

Most significantly, however, all main effects of Prior Testing now disappear. Only two interactions of Cohort and Prior Testing (R, W) reach significance ($p < .05$). Inspection of the means in Table 4 suggests that for these variables the means for two cohorts (mean ages, 39 and 53) are raised above that of their unpretested counterparts. Note also that this trend is quite consistent for each of the variables, even though it does not reach significance for all of them.

Discussion

The discrepancy between the two analyses reported in this study suggests that what traditionally has been interpreted as a re-test effect indeed may be an artifact created by the operation of selective attrition (experimental mortality) of retested samples. Thus initial analysis, involving no attempt to equate pretested and unpretested groups, strongly suggests that previous assessment may significantly raise test scores, thereby suggesting ontogenetic changes which in fact have not occurred. The second analysis, however, did

attempt to select control subjects who are selectively biased similarly to previously tested subjects. Here, it is found that testing effects are of minimal significance in modifying ontogenetic patterns. The present data clearly substantiate therefore the notion that the study of testing effects by means of deisgns involving posttest control groups requires more careful consideration of selective attrition effects than has been given traditionally.

Note, however, that this inference is restricted to the extent that volunteering behavior may vary over time, either as a function of age or cultural change. That is, although the data for the samples included in our second analysis were gathered at one measurement point (1963), the basis for differentiating retest-resisters and retest-participants is, of course, seven years apart (namely, 1963 for the pretested sample, and 1970 for the unpretested sample). Although at present there is no evidence to suggest an interaction of age and/or cohort differences and experimental mortaility, the possibility of age and/or cohort-related changes in recruiting behavior requires closer examination (e.g., Baltes, Schaie, & Nardi, 1971). In fact, data from the present study are currently being examined for this possibility.

In the absence of such interaction effects, however, the present data suggest that retest effects are not a major source of error, at least in the area of intellectual ontogeny (for other variables see, however, Schaie, 1971a). From a measurement perspective, this is extremely encouraging, since the presence of prior testing effects could potentially jeopardize the distinction between ontogenetic and generational change, insofar as this distinction is derived from comparisons between longitudinal and cross-sectional gradients.

References

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

Baltes, P. B. Longitudinal and cross-sectional sequences in the study of age and generation effects. Human Development, 1968, 11, 145-171.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963. Pp. 171-246.

Kuhlen, R. G. Age and intelligence: The significance of cultural change in longitudinal vs. cross-sectional findings. Vita Humana, 1963, 6, 113-124.

Nesselroade, J. R., Schaie, K. W., & Baltes, P. B. Ontogenetic and generational components of structural and quantitative change in adult cognitive behavior. Journal of Gerontology, 1971, in press.

Riegel, F. F., Riegel, R. M., & Meyer, G. A study of the drop-out rates in longitudinal research on aging and the prediction of death. Journal of Personality and Social Psychology, 1967, 4, 342-348.

Schaie, K. W. Rigidity-flexibility and intelligence: A cross-sectional study of the adult life span from 20 to 70 years. Psychological Monographs, 1958, 72, (9, Whole No. 462).

Schaie, K. W. Cross-sectional methods in the study of psychological aspects of aging. Journal of Gerontology, 1959, 14, 208-215.

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

Schaie, K. W. Cultural change and repeated assessment in the study of the adult personality. Paper presented as part of a symposium on "Personality assessment of the aged" held at the 79th Annual Convention of the Amiercan Psychological Association, Washington, 1971. (a)

Schaie, K. W. Generational vs. ontogenetic components of change: A second follow-up. Paper presented as part of a symposium on "Intellectual Development in Adulthood and Old Age" held at the 79th Annual Convention of the Ameircan Psychological Association, Washington, 1971. (b)

Schaie, K. W. Methodological problems in descriptive developmental research on adulthood and aging. In J. R. Nesselroade & H. W. Reese (Eds.), Life-span developmental psychology: Methodology. New York: Academic Press, 1972, in press. (a)

Schaie, K. W. Can the longitudinal method be applied to studies of human development? Proceedings of first symposium of the International Society for the Study of Behavioral Development, Basel: Karger, 1972, in press. (b)

Schaie, K. W., & Strother, C. R. A cross-sequential study of age changes in cognitive behavior. Psychological Bulletin, 1968, 70, 671-680.

Wohlwill, J. F. Methodology and research strategy in the study of developmental change. In L. R. Goulet & P. B. Baltes (Eds.), Life-span developmental psychology: Research and theory. New York: Academic Press, 1970. Pp. 149-191.

Table 1

Sampling Plan and Analysis of Variance Model

for the Cross-Seuqential Model Controlled for Practice

a. Sampling Plan

Time of Measurement	Tested Samples			Untested Samples	
T (1956)	$S_a O_1 T_1$			$S_b O_0 T_1$	$S_c O_0 T_1$
T (1963)	$S_a O_2 T_2$	$S_b O_1 T_2$			$S_c O_0 T_2$
T (1970)	$S_a O_3 T_3$	$S_b O_2 T_3$	$S_c O_1 T_3$		

Note.--S_a, S_b, and S_c are random samples from base population. O refers to observation (O_0 = no observation, O_1 = first observation, etc.) and T to time of measurement. Framed samples are included in analysis.

b. Analysis of Variance Model

Source of Variation	Degrees of Freedom
Between Cohorts (C)	C - 1
Between Times (T)	T - 1
Between Practice Levels (P)	P - 1
Cohort x Time interaction	(C - 1) (T - 1)
Cohort x Practice interaction	(C - 1) (P - 1)
Time x Practice interaction	(T - 1) (P - 1)
Cohort x Time x Practice interaction	(C - 1) (T - 1) (P - 1)
Error	N - (C) (T) (P)
Total variation	N - 1

Table 2

Cross-Sequential Analysis of Variance for Effects of Cohort, Time, Sex, and Prior Testing (Significant F Ratios; df for error terms = 2122)

		P M A Variables					
Source of Variation	df	V	S	R	N	W	IQ
Cohort (C)	6	137.2**	121.3**	188.3**	41.9**	41.1**	158.8**
Time of Measurement (T)	1	5.7*					
Sex (S)	1	13.2**	149.**	10.2**	13.7**	36.6**	
Prior Testing (P)	1	64.3**	27.4**	48.4**	5.0*	24.8**	44.3**
C x T	6				2.4*		2.9*
C x S	6					2.6*	
C x P	6				2.2*	3.5*	2.3*
T x S	1						
T x P	1						
S x P	1						
C x T x S	6	2.1*					
C x T x P	6	3.2**		2.1**		3.9**	2.5**
C x S x P	6						
T x S x P	1						
C x T x S x P	6						

Table 3

Cross-Sequential Analysis of Variance for Effects of Cohort, Sex, and Previous Testing Controlled for Experimental Mortality (Significant F Ratios, df for error terms = 646)

		P M A Variables					
Source of Variation	df	V	S	R	N	W	IQ
Cohort (C)	6	31.8**	25.2**	49.1**	6.8**	6.8**	31.1**
Sex (S)	1		56.5**		5.2*	11.7**	
Prior Testing (P)	1						
C x S	6					2.6*	
C x P	6			2.2*		3.2*	
S x P	1						
C x S x P	6						

Table 4

Mean T-Scores of Pretested and

Unpretested Subjects without and with Control for Experimental Mortality

	Intact Samples		Attrited Samples	
Cohort	Pretested	Not Pretested	Pretested	Not Pretested
		Verbal Meaning		
1	42.7	39.4	42.2	44.6
2	45.9	41.7	46.0	47.3
3	50.9	46.4	51.9	49.5
4	54.1	50.2	56.9	52.8
5	55.3	53.3	54.5	55.1
6	57.3	54.7	57.7	55.3
7	56.1	54.2	55.9	54.9
		Space		
1	42.9	40.9	43.2	46.3
2	44.6	42.7	45.2	46.7
3	47.3	45.9	47.9	49.4
4	50.8	49.1	51.9	50.0
5	53.2	51.5	52.3	53.3
6	54.6	53.4	55.2	53.2
7	56.5	54.1	57.4	54.5

Table 4 (continued)

Cohort	Intact Samples		Attrited Samples	
	Pretested	Not Pretested	Pretested	Not Pretested
		Reasoning		
1	41.6	40.0	41.5	43.8
2	43.4	40.9	43.8	44.2
3	48.1	44.2	49.2	47.4
4	51.6	48.4	53.8	49.6
5	52.6	51.2	51.8	53.6
6	56.1	54.2	57.4	55.3
7	58.2	55.8	58.0	57.5
		Number		
1	45.4	42.7	45.8	46.2
2	46.3	44.7	47.3	48.4
3	51.6	47.7	53.8	50.2
4	53.4	51.3	53.8	52.2
5	52.8	53.0	52.7	52.4
6	52.7	53.7	52.8	52.7
7	51.7	52.5	50.6	52.2

Table 4 (continued)

Cohort	Intact Samples		Attrited Samples	
	Pretested	Not Pretested	Pretested	Not Pretested
		Word Fluency		
1	44.3	43.5	44.9	48.0
2	47.3	43.2	47.7	47.8
3	50.9	46.7	51.1	52.4
4	51.9	48.7	54.8	49.6
5	50.2	51.5	51.2	50.9
6	53.3	51.9	54.1	50.7
7	53.3	51.3	51.1	54.8
		IQ		
1	41.7	39.0	41.9	44.6
2	44.2	40.8	45.0	46.2
3	50.0	45.3	51.5	49.6
4	53.2	49.7	55.4	51.3
5	53.4	52.8	53.2	53.8
6	55.8	54.6	56.6	54.3
7	56.1	54.5	55.3	55.9