THE AGING OF TESTS:
A METHODOLOGICAL PROBLEM OF LONGITUDINAL STUDIES [1]

Kathy Gribbin & K. Warner Schaie
Andrus Gerontology Center

Paper presented at the 30th Annual Meeting of the Gerontological Society, San Francisco, November, 1977

[1] This project was supported in part by research grant AG00480 from the National Institute on Aging. The cooperation of Group Health Cooperative of Puget Sound is gratefully acknowledged.

INTRODUCTION

Longitudinal designs are perhaps the most powerful method for detailing changes that occur with increasing age. There are, however, several disadvantages inherent in the utilization of such designs (cf. Botwinick, 1967, 1973). One of these is that outmoded measurement instruments must continue to be employed, even though newer and better ones may later become available, in order to permit orderly comparison of the measurement variables.

If one is primarily interested in comparing directly observable behavior, the above concern is well taken. However, since most psychological measurement instruments are no more than arbitary samples of behavior designed to assess individual differences with respect to certain psychological constructs, the comparison of some derived scores which retain the same meaning over all comparisons (cf. Baltes, Nesselroade, Schaie & Labouvie, 1972; Nesselroade, Schaie & Baltes, 1972; Schaie, 1973) is actually the question of interest. Thus it is possible to convert from one set of measures to another if the appropriate linkage studies--studies which give some indication of the common factor structure for both old and new measures--are undertaken.

Considerable thought must be given to the design of such linkage studies. New instruments must be chosen which, on either a theoreteral or empirical basis, should be measuring the same constructs as the previously used instruments. Thus it is advisable to include a variety of tasks purported to measure the same constructs. Determination of which new measures best describe the same information as gained from the older measures can then be made on an empirical basis with scores on the resultant new battery closely matching those from the older battery.

Another factor which is important to the design of linkage studies concerns the sample for which comparison of the new and old measures is made. In order to gain information on changes in the range of performances, reliability, and construct validity, the "linkage" sample should be drawn from the same parent population and should be comprised of individuals of the same sex(es) and age range as those in the longitudinal study.

Since the linkage study demands that the same individuals are given

both the old and the new measurement variables, suitable regression techniques permit judgment regarding whether to convert to new measures, and if so, which measures to include in the new battery. Alternatively, results may show that switching to the new tests would result in significant information loss which would argue for retention of one or more sub-tests from the old form.

Since 1956, Schaie and his associates have been using the 1948 edition of the Primary Mental Abilities Test (PMA) (Thurstone & Thurstone,1948), a test of intellectual functioning, in a longitudinal study with repeated testings in 1963 and 1970. During the 1970 data collection it was noted that a few younger subjects seemed to be finishing certain subtests shortly before the allotted time for test administration had elapsed. Since sequential analyses of the data had shown that those subjects aged 22 to 28 have successively higher scores at each time of measurement (Schaie, Labouvie & Buech, 1973), the question was raised as to whether or not a "ceiling effect" would be reached in the 1977 follow-up. In other words, although this test may be a valid measure for older people, the question was raised as to whether or not over the time period monitored, the test had "aged" so as not to allow for a full range of variability of measurement for younger people. Such an occurrence would question whether or not the measures retain appropriate construct validity. Although the old test may no longer be appropriate for successive younger age levels, it must be kept in mind that new tests may not be appropriate for older cohorts. For example, Monge and his associates[2] found that while 20- and 30 years olds performed significantly better on items entering the language after 1960, 40- to 60-year olds performed significantly better on items entering the language in the late 1920's. Consequently, changing to new tests may significantly alter the construct validity for older subjects.

The purpose of the present study was to assess the continuing reliability of the 1948 PMA by jointly administering it with a revised 1962 version, and selected tests from the Kit of Reference Tests for Cognitive Factors published by the Educational Testing Service (ETS) (1963). The 1962 PMA was chosen because it was felt that it would be most similar to the 1948 version, while the ETS tests were included with the expectation that

although they were slightly different from the PMA tests, they might account for some additional variance which would reduce the information loss should a decision be made to switch PMA tests for future testings.

METHOD

Subjects

The population base consisted of the approximately 128,000 individuals aged 22 to 82 who, in 1975, were members of a prepaid medical plan in a Pacific Northwest Metropolitan area (See Gribbin, Schaie & Stone, 1976[2], for a discussion of population differences between the 1956 and 1974 member population). Data were collected from 242 men and women (aged 22 to 88) randomly sampled by sex and within year of birth.

Measurement Variables

The 1948 PMA consists of five subtests--Verbal Meaning (V_{48}), Space (S_{48}), Reasoning (R_{48}), Number (N_{48}), and Word Fluency (W_{48}). The 1962 version differs from the earlier format by omitting W_{48}; by having N_{62} include subtraction, multiplication, and division instead of just addition as does N_{48}; and by having R_{62} include number series and word groupings as well as the letter series which is included in R_{48}. More vocabulary items are included V_{62} than are in V_{48}. Selected tests from the ETS battery included: Hidden Patterns, a measure of flexibility of closure; Letter Sets, a measure of inductive reasoning; Length Estimation, the ability to judge and compare visually perceived distances; Finding A's and Identical Pictures, measures of Perceptual Speed; Nonsense Syllogisms, a measure of syllogistic reasoning; Maze Tracing, which requires spatial scanning; and Paper Folding, which requires transforming the image of spatial patterns into other visual arrangements. The ETS tests all have two parts with each form similar to the other.

Tests were administered in a modified counterbalanced order; i.e., PMA (1948), ETS tests, PMA (1962), or PMA (1962), ETS tests (in reverse order from initial order), and PMA (1948). A 20 minute break, where refreshments were served, was given after one-half of the ETS tests had been administered.

Design and Data Analysis

Regression techniques were employed to determine the relationship between the tests. For each subtest to be predicted (i.e., subtests from both versions of the PMA), scores on all subtests from the alternative version plus each part of the eight ETS tests were used as predictor variables across all subjects. Since the relationship between the variables for different age groups is also of significant importance, similar analyses were again conducted separately by age groups (those aged 22 to 51, and those aged 52 to 82) in order to determine whether predictability of the tests differed by age groupings.

RESULTS

Table 1 presents the R^2 (amount of variance accounted for) for each subtest of the 1948 PMA, as well as the beta weights (β) (standardized regression coefficients) for each predictor variable. Similarly, Table 2 presents the same information for the 1962 PMA. The reliabilities of each test are presented at the bottom of each of the tables. As can be seen by comparing the R^2 of the relevant subtests from each version, in general, the 1962 version is better predicted than is the 1948 edition. This finding is particularly apparent in the case of N_{48} and W_{48}. For these subtests, only 43 and 40 per cent of the variance, respectively, can be accounted for. Thus, although accountable variance for the other subtest is similarly high, that of N_{48} and W_{48} is too low to consider in terms of changing subtests.

Results of the age group regressions are presented in Tables 3-6. Table 3 describes the 1948 tests and Table 4 the 1962 tests for younger subjects (ages 22-51); Table 5 describes the 1948 tests and Table 6 the 1962 tests for older subjects (ages 52-82).

Comparison of Tables 3 and 4 show that N_{48} and W_{48} are even more poorly predicted for young subjects than for the entire sample. Results from the other tests are again relatively comparable, with good prediction of all subtests.

Comparison of Tables 5 and 6 show that for older subjects all subtests except N_{48} and W_{48} are predicted well. It is interesting to note that N_{48} is better predicted for older (R^2=.52) than for younger (R^2=.37) subjects, while the reverse is true for W_{48}(R^2=.34 and .45 respectively).

DISCUSSION

Results suggest that, along with certain ETS tests, one could replace V_{48} with V_{62}, S_{48} with S_{62}, and R_{48} with R_{62}, and sustain relatively little loss of information. For example, for V_{48}, R^2=.69. Since the reliability of the test is .92, only 84 percent of the variance of the test is reliable variance. Thus not much information is lost if V_{48} were to be replaced by other subtests. This finding is even more pronounced for R_{48} where R^2=.84 and the reliable variance is .86, and for S_{48} where R^2=.70 and the reliable variance is .74. For these tests practically no information is lost.

For N_{48} and W_{48}, however, analysis of regression results dictate continued usage of these tests. Although the amount of variance accounted for differs significantly from zero (.43 and .40 respectively), it was too small to justify replacement of the tests.

Given the above results, one could easily determine which set of tests to use in order to replace V_{48}, S_{48}, and R_{48}. In fact, since some of the ETS tests contribute very little additional information, another set of regression analyses could be performed, without including these tests, so as to get suitable regression equations on a more limited set of tests for replacement purposes. Then one could proceed with the longitudinal study now utilizing the new set of tests along with N_{48} and W_{48}.

Comparison of regression equations for young and old groups suggests that although, in general, the amount of variance accounted for is similar, the contribution of each variable to the equation differs for each group. Since the R^2 for the entire group is at least as high (if not higher) than the separate analyses for young and old, it is advised that if one were to use the

new tests, the regression equation for the entire sample should be utilized.

In the present case, however, it was decided to continue with the 1948 edition of the PMA. Since this edition is composed of one test booklet, separation of the tests in order to maintain usage of W_{48} and W_{48} might bias the results since administration procedures would have to be changed. It must be kept in mind, however, that the mean score for younger cohorts on some subtests may be artificially constricted since a "ceiling effect" may be reached on certain subtests by a few individuals.

FOOTNOTES

[2]Monge, R. Personal Communication, 1971

[3]Gribbin, K., Schaie, K. W., & Stone, V. Ability differences between established and redefined populations in sequential studies. Paper presented at the American Psychological Association, Washington, D.C., 1976.

REFERENCES

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

Botwinick, J. Cognitive processes in maturity and old age. New York: Springer, 1967.

Botwinick, J. Aging and Behavior. New York: Springer, 1973.

French, J.W., Ekstrom, R.B., & Price, L.A. Kit of reference tests for cognitive factors. Princeton, New Jersey: Educational Testing Service, 1963.

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, 1972, 27, 222-228.

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: Methodological issues. New York: Academic Press, 1973, 253-280.

Schaie, K.W., Labouvie, G.V., & Buech, B.U. Generational and cohort-specific differences in adult cognitive functioning: A fourteen-year study of independent samples. Developmental Psychology, 1973, 9, 151-166.

Thurstone, L.L., & Thurstone, T.G. SRA Primary mental abilities: Inter-mediate--ages 11-17. Chicago: Science Research Associates, 1948.

Thurstone, T.G. Primary mental abilities for grades 9-12. Chicago: Science Research Associates, 1962.

TABLE 1

Regression Analyses Predicting the 1948 PMA: ALL SUBJECTS

BETA WEIGHTS

	$VERBAL_{48}$	$SPACE_{48}$	$REASONING_{48}$	$NUMBER_{48}$	$WORD_{48}$
$VERBAL_{62}$	.50	-.07		.26	.24
$SPACE_{62}$		.58			-.11
$REASONING_{62}$:					
LETTER SERIES		-.18	.43		
NUMBER SERIES	.24		.18		.18
WORD GROUPINGS	-.24	.14		.12	
$NUMBER_{62}$	.13	.17		.33	
HIDDEN PATTERNS 1					
HIDDEN PATTERNS 2	.11		.10		
FINDING A'S 1			.04	.11	
FINDING A'S 2	.05				.18
MAZE TRACING 1				-.33	.16
MAZE TRACING 2	-.15		-.09	.16	-.23
PAPER FOLDING 1	.06	.08			
PAPER FOLDING 2				-.08	.09
IDENTICAL PICTURES 1			.12		
IDENTICAL PICTURES 2		.07			
LETTER SETS 1	.14	.09	.15		
LETTER SETS 2	.11	.07	.14	.15	.25
LINE ESTIMATION 1			.04		
LINE ESTIMATION 2			-.07		-.11
NONSENSE SYLLOGISMS 1				-.06	.06
NONSENSE SYLLOGISMS 2		-.04			
R^2:	.69	.70	.84	.43	.40
Reliable Variance[b]:	.81	.92	.86	.80	N.A.

[b] r^2 as provided by authors in test manual

TABLE 2

Regression Analyses Predicting the 1962 PMA: ALL SUBJECTS

BETA WEIGHTS

	$VERBAL_{62}$	$SPACE_{62}$	$REASONING_{62}$	$NUMBER_{62}$
$VERBAL_{48}$	.41			.12
$SPACE_{48}$		.54	.10	.18
$REASONING_{48}$			.39	-.07
$NUMBER_{48}$	.13		.10	.20
WORD FLUENCY	.10		.06	
HIDDEN PAT's 1		.12		
HIDDEN PAT's 2				
FINDING A'S 1	.13	.13		
FINDING A'S 2		-.10		
MAZE TRACING 1		.10		.12
MAZE TRACING 2	.14		.14	.16
PAPER FOLDING 1				.09
PAPER FOLDING 2				
IDEN. PICTURES 1		.13		
IDEN. PICTURES 2	.11			
LETTER SETS 1			.22	.23
LETTER SETS 2	-.12			
LINE ESTIMATION 1				
LINE ESTIMATION 2	.12	.06	.09	
NONSENSE SYL's 1			.05	.16
NONSENSE SYL's 2	.15	.07	.08	
R^2:	.72	.73	.87	.72
Reliable Variance*:	.81	.73	.58	.64

*r^2 as provided by authors in test manual

TABLE 3

Regression Analyses Predicting the 1948 PMA: YOUNG SUBJECTS

BETA WEIGHTS

	$VERBAL_{48}$	$SPACE_{48}$	$REASONING_{48}$	$NUMBER_{48}$	$WORD_{48}$
$VERBAL_{62}$	.59	-.17	.09	.21	.27
$SPACE_{62}$	-.11	.64	-.13		
$REASONING_{62}$:					
LETTER SERIES		-.26	.42	.18	.12
NUMBER SERIES	.39	.22	.24		.18
WORD GROUPINGS	-.23				
$NUMBER_{62}$	.18	.27	-.21	.18	
HIDDEN PATTERNS 1	-.14				
HIDDEN PATTERNS 2	.22		.20	.10	
FINDING A'S 1				.14	-.11
FINDING A'S 2					.27
MAZE TRACING 1				-.37	
MAZE TRACING 2	-.22	-.08		.25	
PAPER FOLDING 1				-.13	
PAPER FOLDING 2					
IDENTICAL PICTURES 1			.26		
IDENTICAL PICTURES 2		.15	-.16		
LETTER SETS 1			.28	.14	
LETTER SETS 2	.10	.11			.17
LINE ESTIMATION 1	-.09				
LINE ESTIMATION 2					-.11
NONSENSE SYLLOGISMS 1			.11		
NONSENSE SYLLOGISMS 2					
R^2:	.65	.61	.72	.37	.45

TABLE 4

Regression Analyses Predicting the 1962 PMA: YOUNG SUBJECTS

BETA WEIGHTS

	$VERBAL_{62}$	$SPACE_{62}$	$REASONING_{62}$	$NUMBER_{62}$
$VERBAL_{48}$	.50	-.10		.24
$SPACE_{48}$		.59	.17	.16
$REASONING_{48}$			.38	-.16
$NUMBER_{48}$	.08	.13	.10	.11
WORD FLUENCY	.11		.12	
HIDDEN PAT's 1				
HIDDEN PAT's 2				
FINDING A'S 1	.16			-.10
FINDING A'S 2				
MAZE TRACING 1				
MAZE TRACING 2	.13		.19	.36
PAPER FOLDING 1		.16		
PAPER FOLDING 2				
IDEN. PICTURES 1	.12	.20	-.10	
IDEN. PICTURES 2				
LETTER SETS 1			.22	.37
LETTER SETS 2	-.13			
LINE ESTIMATION 1	.14		.09	
LINE ESTIMATION 2				.10
NONSENSE SYL's 1		-.08		.26
NONSENSE SYL's 2	.19	.13	.18	
R^2:	.67	.62	.76	.63

TABLE 5

Regression Analyses Predicting the 1948 PMA: OLDER SUBJECTS

BETA WEIGHTS

	$VERBAL_{48}$	$SPACE_{48}$	$REASONING_{48}$	$NUMBER_{48}$	$WORD_{48}$
$VERBAL_{62}$	.28		-.12	.26	
$SPACE_{62}$	.14	.46			
$REASONING_{62}$:					
LETTER SERIES	.19		.47		
NUMBER SERIES		-.17	.14		.18
WORD GROUPINGS	-.18	.17		.30	-.13
$NUMBER_{62}$				.31	
HIDDEN PATTERNS 1	.27		.17		
HIDDEN PATTERNS 2	-.14	.10			-.17
FINDING A'S 1	-.15	.15	.09		
FINDING A'S 2	.19				.14
MAZE TRACING 1				-.18	.18
MAZE TRACING 2			-.10		-.24
PAPER FOLDING 1		.14			
PAPER FOLDING 2		.12			
IDENTICAL PICTURES 1				.11	.18
IDENTICAL PICTURES 2	.16				
LETTER SETS 1	.23	.15	.10	-.26	
LETTER SETS 2			.19	.25	.37
LINE ESTIMATION 1			.11		
LINE ESTIMATION 2					
NONSENSE SYLLOGISMS 1		-.11		-.10	.15
NONSENSE SYLLOGISMS 2		-.08	.06		
R^2:	.68	.72	.85	.52	.34

TABLE 6

Regression Analyses Predicting the 1962 PMA: OLDER SUBJECTS

BETA WEIGHTS

	$VERBAL_{62}$	$SPACE_{62}$	$REASONING_{62}$	$NUMBER_{62}$
$VERBAL_{48}$	.25	.13		
$SPACE_{48}$		.51		.20
$REASONING_{48}$			.35	.04
$NUMBER_{48}$	.22		.14	.24
WORD FLUENCY				.08
HIDDEN PAT's 1	.13			.18
HIDDEN PAT's 2		.29		
FINDING A'S 1	.18	.21	.11	
FINDING A'S 2		-.14		
MAZE TRACING 1		.08		
MAZE TRACING 2				
PAPER FOLDING 1		-.20	-.07	.18
PAPER FOLDING 2			.11	
IDEN. PICTURES 1	.11			
IDEN. PICTURES 2			.09	
LETTER SETS 1			.22	.15
LETTER SETS 2	-.13			
LINE ESTIMATION 1				
LINE ESTIMATION 2	.21		.14	.10
NONSENSE SYL's 1	.08			.13
NONSENSE SYL's 2	.09	.08		
R^2:	.67	.68	.87	.72