Developmental Psychology
1980, Vol. 16, No. 6, 625–635

Integration Versus Differentiation of Fluid/Crystallized Intelligence in Old Age

Paul B. Baltes, Steven W. Cornelius, Avron Spiro, John R. Nesselroade, and Sherry L. Willis
College of Human Development, Pennsylvania State University

Using an intra-age simulation approach, the purpose of the study was to evaluate several models of psychometric intelligence differing primarily along the developmental continuum of differentiation–integration for older adults. A sample of 109 elderly adults (age range = 60–89) was administered a battery of 17 ability tests selected to mark Cattell's and Horn's fluid (Gf) and crystallized intelligence (Gc) and speed dimensions. Using the Gf/Gc model of intelligence and intermediate outcomes as guidelines, structural models with between seven and one factors were successively evaluated by confirmatory factor analysis. Three main findings emerged. First, models with fewer factors provided better fits to the data. Second, a model with a general factor and three group factors was especially acceptable based on empirical and theoretical criteria. Third, it was not possible to obtain an acceptable solution for older subjects that would directly parallel the Gf/Gc structural pattern of first-stratum factors reported for younger age groups. In concert, the findings support a neointegration, or dedifferentiation, view of psychometric intelligence in old age.

A central issue in developmental research concerns changes in the structure or organization of behavior in both evolutionary (e.g., Schneirla, 1966) and ontogenetic (e.g., Riegel & Rosenwald, 1975; Werner, 1948) manifestations. One possible research strategy for attacking the question of structural change in life-span behavioral development involves the application of multivariate data analytic methods and, in particular, comparative factor analysis. Within this general context, the primary purpose of this study is to examine the fit of alternative structural models of psychometric intelligence to data representing ability-related performance in old age. This question is related to the long-standing search for constancy versus change in the structural relationships among intellectual abilities over the life span.

Historically, the age-differentiation hypothesis of psychometric intelligence (e.g., Burt, 1954; Garrett, 1946) originally covered only the period from childhood to early adulthood. From a life-span orientation and following early work by Balinsky (1941), the Burt–Garrett hypothesis was extended to later adulthood and resulted in the formulation of an integration → differentiation → neointegration (dedifferentiation) hypothesis. In other words, it is hypothesized that intellectual ability structures become integrated in early childhood, undergo a process of differentiation during late childhood and adolescence, exhibit a fair degree of invariance during adulthood, and manifest a process of neointegration, or

This research was supported by a grant from the National Institute of Aging (#5-R01-AG004403) to Paul B. Baltes and Sherry L. Willis, co-investigators. In addition, the second and third authors were supported in part by graduate fellowships from the National Science Foundation and the Administration on Aging, respectively.

The first author expresses appreciation to the late Günter Reinert (1928–1979) who, as thesis mentor, introduced him to the differentiation hypothesis of intelligence (Reinert, Baltes, & Schmidt, 1965). The authors would like to acknowledge the assistance of Ruth Bittner, Margie Lachman, Vincent Morello, and Carolyn Nesselroade in various phases of this study.

Requests for reprints should be sent to Paul B. Baltes, Max-Planck Institute for Human Development and Education, Lentzeallee 94, 1 West Berlin 33, West Germany.

dedifferentiation, in old age. In general, reviews of research on life-span changes in the structure of psychometric intelligence (e.g., Anastasi, 1970; Baltes & Labouvie, 1973; Lienert & Crott, 1964; Reinert, 1970) tend to conclude that the findings at least partially support the model. In part, a definite conclusion is difficult because past research has suffered from a number of methodological flaws and shortcomings in factor-analytic procedure (e.g., Baltes, Nesselroade, & Cornelius, 1978; Cunningham, 1978). Moreover, there is relatively little evidence for late adulthood, although it is predicted that a major structural transformation of ability patterns may occur during old age.

During the last decade there has been a continuation of research on this topic spurred perhaps by two developments. First, there has been a new interest in examining the question of differentiation versus integration in old age (Cunningham, 1980; Cunningham & Birren, 1980; Hayslip & Sterns, 1979; Rudinger, 1976). Researchers have lamented the fact that much aging research uses instruments and structural models developed for the young rather than the older adult. Second, the continued development of new factor-analytic methods capable of providing a statistical basis for fitting and evaluating models (e.g., Bentler, 1980) offers the possibility of a higher level of precision in factor-analytic research. In the past, much factor-analytic work was judged inadequate primarily because extraction and rotation procedures were exceedingly arbitrary and clear criteria did not exist for assessing the adequacy of interpretation, rigor of comparisons, and appropriateness of model testing (Cunningham, 1978). Although a final evaluation of the usefulness of the more recently developed techniques (e.g., Sörbom & Jöreskog's, 1976, confirmatory factor analysis) is premature, the current strong concerns about the validity of model-testing procedures generally provides a clear gain over the past orientation and practices.

Within the context of developmental research on the structure of psychometric intelligence, the present study is characterized by several features and objectives. First, although Cattell's and Horn's (Cattell, 1971; Horn, 1970, 1978) theory of fluid/crystallized (Gf/Gc) intelligence has been formulated for the entire life span, this is the only study using the measurement framework provided by that theory for testing structural models with older adults. Pertinent research by Hayslip and Sterns (1979) on Gf/Gc relationships with older adults is based on a restricted battery of measures and, instead of factor-analytic methodology, involves direct comparison of test intercorrelations to examine Gf/Gc relationships. In the present study, specification of the structural models chosen reflects the putative first-stratum factor patterns on which Cattell's and Horn's second-stratum theory of Gf/Gc intelligence is based. If possible, the goal is to fit at an acceptable level a pattern that matches that theory.

Second, as in most developmental research, there are two primary strategies to contribute to knowledge about the structure of psychometric intelligence through the life span (Baltes, Reese, & Nesselroade, 1977). The most frequently applied strategy is to compare structural patterns of two or more age groups. The other research strategy, which was adopted in the present study, involves the successive evaluation and fitting of alternative models of intellectual structure based on life-span developmental considerations. Using the hypothesized integration-differentiation-neointegration sequence, the fitted models vary along a developmental continuum of differentiation (many factors) to integration (few factors). Thus, the strategy used in this study is to focus on a single age group (elderly adults), use information gained from studies with other age groups (e.g., integration-differentiation-neointegration sequence) as a framework for hypothesis testing, and examine the fit and usefulness of alternative "developmental" models for the particular age group under investigation.

Method

Subjects

The sample consisted of 122 community-dwelling older adults from rural areas of central Pennsylvania. Subjects were recruited through community organizations. Each subject was either paid $15 directly for participating or a $15 contribution was made to the organization on their behalf. Data for 13 subjects were incomplete and were excluded from analyses. The mean age of the remaining 109 subjects (28 males, 81 females) was 69.9 years ($SD = 6.9$). The age range was 60–89 years with about 85% of the sample between the ages of 62 and 77 years. Female subjects (M age = 72.8; $SD = 8.5$) were older, $t(107) = 2.62$, $p < .01$, than male subjects (M age = 68.9; $SD = 6.0$). The mean educational level of the sample was 11.7 years ($SD = 3.1$; range, 6–19) with no significant sex differences in educational level. The self-reported health status of participants was good, and subjects reported no substantial hearing or visual impairments. On the average, the sample is likely to be somewhat positively selected on educational and health indices compared to cohort-specific census data of Pennsylvania (Institute for The Study of Human Development, Note 1). However, the subjects resided in a variety of communities and geographical locations in central Pennsylvania. Thus, the composition is fairly heterogeneous despite a positive selection in average education and health.

It may be useful to offer a few observations on the sizable age range included in this study. Since the differentiation hypothesis of intelligence is based on age contrasts, there is a question of whether the given age group studied should exhibit any age variation itself. Because age–behavior correlations are possible within the age group studied, the key problem is the role (amount, direction) that age-related covariance plays in defining structural relationships among the behaviors (tests) studied. Merz and Kalveram (1965), for example, argued that such within-group age correlations produce statistical artifacts because they add to the "pure" (age-free) correlations among the measures observed. In research on the differentiation hypothesis, they have labeled this effect a *simultaneous coaddition* (*Überlagerung*) whenever the age-related change functions exhibit common directionality for the tests considered. Reinert, Baltes, and Schmidt (1966), on the other hand, though acknowledging the statistical argument of Merz and Kalveram, pointed out that there are often conceptual arguments and questions of sample representativeness (generalizability), both of which can suggest that within-group age differences are an appropriate focus for developmental analysis.

In our view, there is no simple solution to the question of age range. Aside from the issue of sample availability, we consider within-group age variance to be one of the salient ingredients of the data matrix to be analyzed because our interest is in identifying patterns appropriate for the larger age range (e.g., elderly adults) usually under investigation. As a consequence, although we were sensitive to the issue of age heterogeneity, it was decided to leave the sample intact and not to remove age-related variability within the sample. At the same time, however, information is provided (see Table 2) that permits a reanalysis of the data with age-related covariance partialed. If such a procedure is chosen, it is necessary to recognize that age differences are confounded with cohort differences in the data.[1]

Measurement Battery

The intelligence test battery was composed within the framework of the Gf/Gc intelligence theory (Cattell, 1971; Horn, 1970, 1975, 1978). In Table 1, a conceptual framework for the design of the measurement battery is presented at three levels of representation: broad, second-stratum dimensions; associated first-stratum primary mental abilities; and representative marker tests. Three general, second-stratum dimensions were identified: fluid intelligence, crystallized intelligence, and broad speediness, which is a nonintellectual performance factor (Horn, 1978). A set of relatively "pure" primary mental abilities representative of the three general dimensions were identified from previous research (e.g., Horn, 1975). Figural Relations, Induction, and Memory Span were chosen to represent fluid intelligence.[2] Experiential Evaluation and Verbal Comprehension were identified to represent crystallized intelligence. Semantic Relations was selected to represent abilities involving both crystallized and fluid intelligence. Finally, Perceptual Speed was identified to represent broad speediness. For reasons outside the purposes of this study, the dimension of fluid intelligence is somewhat overrepresented within the battery.

Multiple marker tests for each primary mental ability were included in the battery. The criteria for test selection involved use, when possible, of published tests from well-known test batteries; tests that could be administered to groups of older adults; and measures previously used in the Gf/Gc literature. Published tests included in the battery were modified slightly in test format (e.g., enlargement, simplification of response mode), and instructions were clarified to facilitate group administration to elderly adults. No modification was considered sufficiently extensive to alter the measurement validity. Two new tests (Figural Relations Diagnostic Test and Induction Diagnostic Test) were included in the battery. The reliabilities of these newly developed tests (see Table 2) are comparable to the remainder of the test battery.

Procedure

A white 69-year-old woman administered all tests with one of two white young adults (graduate research assistants) assisting as test proctors. Testing sites were community organizational facilities such as senior citizen centers. Subjects were tested during 1977 in 13 groups of 4–12 persons per group. Groups were

[1] Note in Table 2 that the within-group correlations are not large compared to the magnitude of the test correlations. Indeed, an analysis of a matrix based on age-free correlations did not alter the thrust of the results reported in this study.

[2] Horn (1975, 1978) also now recognizes memory (short-term acquisition and retrieval) as a general second-stratum dimension.

Table 1
Schematic Design of Measurement Battery: Hypothesized General Intellectual Dimensions, Hypothesized Primary Mental Abilities, and Marker Tests

General dimension	Primary ability	Test	Source
Gf	CFR	Culture Fair Test (Scale 2, Form A) and Power Matrices (scale 3, Form A, 1963 ed., and Form B, 1961 ed.)	Cattell & Cattell (1957, 1961, 1963)
	CFR	Adept Figural Relations Diagnostic Test (Form A)[a]	Plemons, Willis, & Baltes (1978)
	CFR	Raven's Advanced Progressive Matrices (Set II)	Raven (1962)
Gf	I	Adept Induction Diagnostic Test (Form A)[a,b]	Blieszner, Willis, & Baltes (Note 2)
	I	Induction Standard Test[b]	Ekstrom, French, Harman, & Derman (1976); Thurstone (1962)
Gf	MS	Visual Number Span	Ekstrom et al. (1976)
	MS	Auditory Number Span	After Ekstrom et al. (1976)
	MS	Auditory Number Span–Delayed Recall	After Ekstrom et al. (1976)
Gf/Gc	CMR	Verbal Analogies I	Guilford (1969a)
	CMR	Word Matrix	Guilford (1969b)
Gc	EMS	Social Translations (Form A)	O'Sullivan & Guilford (1965); O'Sullivan, Guilford, & de Mille (Note 3)
	EMS	Social Situations (EPO3A)	Horn (Note 4)
Gc	V	Verbal Meaning (9–12)	Thurstone (1962)
	V	Vocabulary (V-2, V-3, V-4)[b]	Ekstrom et al. (1976)
Gs	P	Finding A's	Ekstrom et al. (1976)
	P	Number Comparison	Ekstrom et al. (1976)
	P	Identical Pictures	Ekstrom et al. (1976)

Note. Hypothesized relationships are based primarily on Cattell (1971) and Horn (1970, 1975, 1978). Hypothesized general intellectual dimensions include fluid intelligence (Gf), crystallized intelligence (Gc), and broad speediness (Gs). Hypothesized primary mental abilities include Figural Relations (CFR), Induction (I), Memory Span (MS), Semantic Relations (CMR), Experiential Evaluation (EMS), Verbal Comprehension (V), and Perceptual Speed (P).
[a] These tests were developed by the investigators for assessment purposes in cognitive training research.
[b] Induction and Vocabulary are composites of several subtests (see text). The Induction Standard subtests include letter sets (Ekstrom et al., 1976), number series, and letter series (Thurstone, 1962).

formed on the basis of geographical proximity to testing sites. The test battery was divided into three test sessions (subbatteries A, B, and C), each approximately 2.5 hours long including multiple rest periods. Several tests of Gf and Gc abilities were assessed in each session. Tests were administered in an invariant order within a test session. The three test sessions were usually held on 3 consecutive days. To minimize possible order effects, three orders of test sessions (Subbattery Orders ABC, BCA, and CAB) were used. Each group of subjects was randomly assigned to one of the three orders. There was no statistical evidence for systematic differences among the three orders.

Scoring and Descriptive Statistics

In three instances, scores from several highly correlated tests assumed to measure the same primary mental ability were combined. For the fluid ability of Induction, composite Induction scores were computed for the Induction Diagnostic and Induction Standard Tests by summing the standardized T scores of the three respective "subtests": Letter Sets, Number Series, and Letter Series Tests. Similarly, standardized T scores on Vocabulary Tests 2, 3, and 4 (Ekstrom, French, Harman, & Derman, 1976) were summed to form a composite vocabulary score.

Descriptive statistics for the 17 ability measures

Table 2
Correlations Among Intellectual Tests and Descriptive Statistics

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Culture Fair	—																
2. Figural Relations Diagnostic	.81	—															
3. Raven's Progressive Matrices	.69	.67	—														
4. Induction Diagnostic	.80	.71	.68	—													
5. Induction Standard	.82	.80	.63	.81	—												
6. Visual Number Span	.45	.44	.27	.39	.43	—											
7. Auditory Number Span	.47	.49	.33	.41	.46	.69	—										
8. Number Span—Delayed Recall	.53	.52	.46	.49	.48	.62	.75	—									
9. Verbal Analogies	.59	.55	.51	.60	.57	.28	.31	.44	—								
10. Word Matrix	.57	.46	.47	.59	.52	.30	.42	.43	.56	—							
11. Social Translations	.72	.73	.62	.69	.71	.34	.38	.46	.45	.47	—						
12. Social Situations	.70	.63	.56	.64	.62	.33	.43	.51	.53	.50	.61	—					
13. Verbal Meaning	.72	.68	.56	.69	.69	.47	.52	.53	.55	.51	.66	.74	—				
14. Vocabulary	.67	.65	.56	.68	.66	.51	.52	.55	.56	.51	.64	.69	.94	—			
15. Finding A's	.46	.48	.42	.59	.53	.26	.27	.26	.38	.34	.46	.36	.52	.50	—		
16. Number Comparison	.61	.51	.44	.68	.61	.37	.31	.35	.39	.38	.46	.46	.58	.54	.55	—	
17. Identical Pictures	.59	.52	.46	.61	.51	.45	.45	.46	.42	.44	.51	.50	.59	.54	.50	.59	—
Reliability (alpha)	.92	.87	.74	.93	.93	.80	.80	.83	.69	.50	.83	.72	.97	.96	.92	.84	.95
M	27.76	18.66	6.44	50.00	50.00	11.22	3.97	4.29	10.61	4.20	10.54	13.63	34.42	49.97	22.86	17.12	24.29
SD	11.35	7.64	4.08	8.85	9.01	3.33	2.98	3.24	4.10	2.25	5.00	3.88	15.30	9.56	7.60	4.61	7.86
Correlation with age cohort	−.35	−.28	−.19	−.27	−.23	−.08	−.08	−.15	−.27	−.25	−.18	−.27	−.15	−.06	−.08	−.21	−.32

are reported in Table 2. The correlation matrix was used in the analyses of different structural models. The correlations are moderately high and positive, ranging between .2 and .9, with the majority in the range from .4 to .6 The alpha reliability coefficients, means, and standard deviations are also listed in Table 2. The reliabilities of the measures are relatively high and satisfactory, with the possible exception of the Word Matrix Test.

Method of Data Analysis

Structural models ranging between 7 and 1 factors were compared to investigate various degrees of structural differentiation in psychometric intelligence for older age cohorts. A hypothesized seven-factor model, representing seven primary mental abilities identified in previous research with younger adult age cohorts, was examined to test the extreme limit of a differentiated structure. Models with 1–5 factors were hypothesized to examine the developmental neointegration hypothesis.

Confirmatory factor analysis was used to evaluate separately the fit of these structural models to the data. The computer program confirmatory factor analysis with model modification (COFAMM; Sörbom & Jöreskog, 1976) was chosen to obtain estimates of salient factor loadings, factor intercorrelations, unique variances, and a chi-square goodness-of-fit statistic. Different classes of models were specified by the hypothesized number of factors. Within a particular class of models, testing focused on hypotheses around the pattern of salient factor loadings. Initially, all factor loadings hypothesized to be nonsalient were set equal to zero. Predicted salient loadings were left as free parameters except for one salient loading per factor that was fixed equal to unity to establish a metric for the estimation. Factor intercorrelations and unique variances were left as free parameters. This strategy was adopted because previous research contains more systematic information about loading patterns and number of factors than factor intercorrelations and unique variances. Moreover, factor-analytic researchers (e.g., Meredith, 1964; Thurstone, 1947) have suggested that constraints on factor covariance matrices must be relaxed to identify invariant loading patterns. Proceeding from the most differentiated to the least differentiated class of models, models were altered on the basis of both theoretical (i.e., predictions derived from the expected structure of Gf/Gc) and empirical considerations (i.e., predictions developed by sequential examination of the fit of alternative models to the present data).

Except for single tests, the application of COFAMM involves a number of aspects where decision making is subjective and based on standards developed from experience. This is particularly true for the decision to accept or reject a given model as plausible and decisions involved in the modification of a given model (e.g., Bentler, 1980). Our decisions were based on two primary criteria. First, models with structural characteristics that were as similar as possible to existing knowledge about psychometric intelligence, especially the Gf/Gc theory, were examined. Second, recommendations (e.g., Bentler, 1980) regarding the interpretation of the statistical test of a model's goodness of fit were followed. In the present study, it was decided to consider solutions plausible if the chi-square for a solution was less than twice the number of degrees of freedom. This standard, compared with others, is fairly conservative in statistical inference.

Results

Seven-Factor Model

The hypothesized seven-factor model, representing the primary mental abilities of Figural Relations, Induction, Memory Span, Semantic Relations, Experiential Evaluation, Verbal Comprehension, and Perceptual Speed, was a test of the extreme boundary of a highly differentiated structure (see Table 1). The standardized solution for the seven-factor model is listed in Table 3. Note that the factor loadings are quite large and in accord with the hypothesized pattern of salient factor loadings. The statistical fit is acceptable, $\chi^2(98) = 102.23$, $p = .36$. The unique variances for about two thirds of the variables are relatively small, indicating a high degree of communality. The factor correlations are high, with four correlations (1 and 2, 1 and 5, 2 and 5, and 5 and 6) exceeding .9. The magnitude of these correlations suggests a high degree of redundancy among some of the factors. Although the fit for the seven-factor model is acceptable, this solution does not have much merit from the perspective of simplicity and parsimony because of the extremely high factor intercorrelations. Such high factor correlations suggest that less differentiated models may be viable alternatives.

A less differentiated five-factor model was tested next. The model was constructed by combining the loading patterns of several factors. First, the Semantic Relations factor was combined separately with each of the fluid abilities of Figural Relations and Induction. Second, a general Crystallized Knowledge factor was formed by combining Semantic Relations with the crystallized ability dimensions of Verbal Comprehension and Experiential Evaluation. Memory Span and Perceptual Speed were retained as distinct factors as in the seven-factor model.

Table 3
Seven-Factor Model, With Factor Loadings and Intercorrelations

Test	Factors 1	2	3	4	5	6	7	Unique variance
		Loadings						
Culture Fair	.93							.14
Figural Relations Diagnostic	.88							.24
Raven's Progressive Matrices	.76							.43
Induction Diagnostic		.91						.18
Induction Standard		.89						.21
Visual Number Span			.76					.42
Auditory Number Span			.88					.23
Number Span—Delayed Recall			.85					.28
Verbal Analogies				.77				.41
Word Matrix				.73				.47
Social Translations					.78			.39
Social Situations					.78			.39
Verbal Meaning						.99		.03
Vocabulary						.95		.11
Finding A's							.69	.52
Number Comparison							.80	.36
Identical Pictures							.73	.47
		Intercorrelations						
Factor								
1. Figural Relations	—							
2. Induction	.96	—						
3. Memory Span	.63	.59	—					
4. Semantic Relations	.82	.85	.59	—				
5. Experiential Evaluation	.99	.95	.64	.83	—			
6. Verbal Comprehension	.78	.78	.63	.72	.91	—		
7. Perceptual Speed	.79	.89	.55	.69	.78	.76	—	

The factor loadings for the five-factor model were large, and, according to the criteria specified earlier, the fit was acceptable, $\chi^2(105) = 156.88$, $p = .00$. However, the factor intercorrelation between the modified Figural Relations and Induction factors was extremely high (.96). This outcome, like the result obtained for the seven-factor model, suggested a model with fewer factors.

Four-Factor Model

A four-factor model was constructed by collapsing the fluid dimensions of Figural Relations and Induction with Semantic Relations into a general Reasoning factor. The Memory Span, Crystallized Knowledge, and Perceptual Speed factors were retained, as in the five-factor model. Analysis of this basic model yielded large factor loadings, moderately high factor intercorrelations, and an acceptable fit, $\chi^2(111) = 165.90$, $p = .00$.

Because this four-factor model was theoretically and empirically acceptable, two types of modifications in the model were made that might improve, or at least not substantially worsen, its statistical fit. First, some parameters (loadings) in the factor pattern matrix, whose values had been fixed initially at zero under the assumption of an extreme simple structure solution (i.e., most variables had a salient loading on only one factor), were selected (based on the magnitude of the first-order derivatives, e.g., Sörbom, 1975) and left free to be estimated. Second, factor loadings that did not differ significantly from zero were fixed equal to zero (e.g., Long, 1976). The combined outcome of this series of modifications is presented in the standardized solution in Table 4. This model incorporates the following modifications: (a) loadings of the Social Translations and Social Situations Tests on the Reasoning factor, Iden-

Table 4
Modified Four-Factor Model, With Factor Loadings and Intercorrelations

Test	Factors 1	2	3	4	Unique variance
	Loadings				
Culture Fair	.92				.16
Figural Relations Diagnostic	.87				.24
Raven's Progressive Matrices	.75				.44
Induction Diagnostic	.58			.37	.19
Induction Standard	.89				.21
Visual Number Span		.76			.42
Auditory Number Span		.88			.22
Number Span—Delayed Recall		.85			.29
Verbal Analogies	.66				.57
Word Matrix	.62				.62
Social Translations	.80				.36
Social Situations	.40		.43		.39
Verbal Meaning			.99		.02
Vocabulary			.94		.11
Finding A's				.70	.51
Number Comparison				.81	.34
Identical Pictures		.26		.59	.44
	Intercorrelations				
Factor					
1. Reasoning	—				
2. Memory Span	.62	—			
3. Crystallized Knowledge	.79	.62	—		
4. Perceptual Speed	.78	.46	.71	—	

tical Pictures Test on the Memory Span factor, and Induction Diagnostic Test on the Perceptual Speed factor and (b) loadings set to zero for the Verbal Analogies, Word Matrix, and Social Translations Tests on the Crystallized Knowledge factor. The modified four-factor model has moderate to large factor loadings, small to moderate unique variances, and moderately large factor correlations with a slight overfit, $\chi^2(110) = 103.95$, $p = .66$. The model includes two fluid factors—Memory Span and a broad Reasoning factor encompassing Figural Relations, Induction, Semantic Relations, and, unexpectedly, Experiential Evaluation. The third factor is a narrow factor of Crystallized Knowledge, and the fourth factor represents Perceptual Speed.

The relatively high factor correlations in the modified four-factor model and our research strategy suggested the examination of less differentiated solutions. A three-factor model was constructed by combining the general Reasoning and Perceptual Speed factors; Memory Span and Crystallized Knowledge factors were retained as in the basic four-factor model. This model was hypothesized based on the relatively large correlation of the Reasoning and Perceptual Speed factors (see Table 4) and an assumption that Perceptual Speed is more highly related to fluid than crystallized abilities (Cattell, 1971). The solution had moderate to large factor loadings, moderately large factor correlations, and a marginally acceptable fit, $\chi^2(114) = 186.16$, $p = .00$. A fairly integrated solution then provides a viable representation.

A two-factor model was examined next. This model involved combining the general Reasoning, Crystallized Knowledge, and Perceptual Speed factors and retaining the Memory Span factor as in the basic four-factor model. The Memory Span factor was retained because it exhibited the lowest correlations with other factors in the previously tested models. The fit of this two-factor model is unacceptable, $\chi^2(118) = 277.00$, $p = .00$. Thus, the two-factor model is unlikely to be a plausible representation.

General-Factor Models

Finally, models involving a general factor were examined. A one-factor model with all variables loading it represents an extreme test of an integrated pattern. The fit of this model was unacceptable, $\chi^2(119) = 370.88$, $p = .00$, even though the factor loadings were relatively large (range, .52–.89) and the unique variances for many of the variables were moderate to small. Thus, a simple general-factor model was rejected as a plausible structural representation of the abilities assessed in the test battery.

In a further examination of a general-factor model, a model was investigated with a general factor and three group factors—Memory Span, Verbal Comprehension, and Perceptual Speed. This model combines characteristics of integration (general factor) and differentiation (group factors). The loadings of variables on the general factor were fixed to equal the values obtained from the one-factor model described earlier, and the variance for the general factor was fixed equal to unity. The standardized solution for this model is presented in Table 5. The fit of the model is relatively good, $\chi^2(122) = 134.02$, $p = .22$. As noted earlier, the loadings of variables on the general factor were relatively large. In addition, the loadings of variables on the group factors were significant. Only the factor intercorrelation between Memory Span and Verbal Comprehension differed significantly from zero. The unique factor variances are approximately the same as those reported earlier for the modified four-factor model.[3]

Table 5
Model With General and Three Group Factors

Test	Factors 1	2	3	4	Unique variance
Loadings					
Culture Fair	.90				.16
Figural Relations Diagnostic	.85				.26
Raven's Progressive Matrices	.73				.44
Induction Diagnostic	.87				.23
Induction Standard	.87				.22
Visual Number Span	.52	.60			.42
Auditory Number Span	.57	.76			.17
Number Span—Delayed Recall	.62	.59			.31
Verbal Analogies	.66				.57
Word Matrix	.63				.61
Social Translations	.79				.36
Social Situations	.77				.43
Verbal Meaning	.85		.55		.06
Vocabulary	.83		.59		.07
Finding A's	.59			.39	.53
Number Comparison	.67			.45	.37
Identical Pictures	.67			.39	.44
Intercorrelations					
Factor					
1. General	—				
2. Memory Span	−.08	—			
3. Verbal Comprehension	−.11	.29	—		
4. Perceptual Speed	−.07	.08	.25	—	

Discussion

Before discussing the findings, a qualification on the general nature of the evaluative procedure used may be appropriate. Despite the strength of confirmatory factor analysis and the use of within-age comparisons as procedures for model testing and building, there are limitations in statistical inference and restrictions on interpretation. First, the procedure of statistical inference associated with COFAMM is not a simple, straightforward process. Our general strategy involved proceeding from a strict and fairly independent test of a differentiated model to other classes of less differentiated models that were formulated in part based on information from previous analyses. Although this procedure is justifiable, it emphasizes the need for additional validation (e.g., cross-validation with new samples). Second, the acceptance of a hypothesized structural model as plausible does not signify that the model is *the* true structural representation but only one of the many possible. To identify a particular model as the only valid one, it would be necessary to reject all possible alternatives. In our analyses, multiple models are plausible. Therefore, additional criteria are necessary for selecting one model over another or in specifying the conditions under which a particular model may be more useful. Finally, not only the strengths but also limitations of a within-age comparative approach must be recognized. Our inferences about life-span changes in ability structures are indirect and rely on comparisons with other research involving different cohorts and procedures. For example, because past research on younger adults has not employed the same tests and factor-analytic procedures used here, we can speculate only that the general finding of more differentiated ability structures for younger adults would hold up if the present procedures were employed.

Recognizing these limitations, the following conclusions are offered. Regarding the integration versus differentiation question (Reinert, 1970), the general conclusion

[3] A general-factor model with four group factors paralleling the four factors shown in Table 4 was also tested. The solution and fit were less satisfactory.

is that the pattern of first-stratum abilities associated with the Gf/Gc framework is fairly integrated in old age. This finding supports the neointegration hypothesis. Though it is possible to obtain loading patterns that parallel the expectations derived from Gf/Gc theory in most models examined, the factor intercorrelations in highly differentiated structures were unacceptably high. Similarly, Hayslip and Sterns (1979) found that intercorrelations among tests of fluid and crystallized intelligence were higher for older than for younger adults. In addition, Cunningham (1980), comparing the ability structures for several adult age groups, found that similar factor-loading patterns could be obtained for adults of increasing age but that there was an age-related increase in the magnitude of factor covariances.

In the analyses reported here, models with fewer factors (including those with a general factor) are judged more acceptable because the factor intercorrelations are of a reasonable magnitude. Although it appears justifiable to accept either the four-factor model (see Table 4) or the general group-factor model (see Table 5), the solution for the latter model is preferable for several reasons. The negligible or relatively small correlations among group factors indicates some differentiation, whereas neointegration is directly manifest in the general factor. This general group-factor model also suggests that the first stratum of the Gf/Gc model of intelligence developed with younger adults needs to be modified when it is applied to older adults. Specifically, there is a general collapse of the fluid/crystallized distinction with certain components of memory and speed tests joining in the formation of a general factor. Also note that of the three fluid factors (Figural Relations, Induction, Memory Span) anticipated, only Memory Span exists as a separate fluid factor. In addition, the crystallized factor is narrower than expected because tests indexing Experiential Evaluation load only on the general factor. Finally, tests of speed identify a separate factor but also show a strong relationship to the general factor. Thus, a highly differentiated structure hypothesized on the basis of research with younger adults (for reviews see Cattell, 1971; Horn, 1970; Reinert, 1970) does not provide a very useful structural pattern of psychometric intelligence in old age.

It may be helpful to relate the present findings to the general context of Cattell's and Horn's theory of fluid/crystallized intelligence. In early writings, their position was oriented more toward structural invariance rather than ontogenetic change in patients. However, Horn (1978) more recently has recognized some form of structural change and acknowledges the possibility of neointegration. In addition, he has increasingly emphasized memory as a second-stratum dimension that is distinct from Gf. In this sense, the present findings support the general direction of Horn's more recent work on the structure and ontogenetic transformation of psychometric intelligence.

Reference Notes

1. Institute for the Study of Human Development. *The continuing social survey: Pennsylvania Turnpike District*. Unpublished manuscript, Pennsylvania State University, College of Human Development, 1975.
2. Blieszner, R., Willis, S. L., & Baltes, P. B. *Training of fluid intelligence (Induction) in the elderly*. Unpublished manuscript, Pennsylvania State University, College of Human Development, 1979.
3. O'Sullivan, M., Guilford, J. P., & de Mille, R. *Measurement of social intelligence* (Tech. Rep. No. 34). Los Angeles: University of Southern California, Psychology Laboratory, 1965.
4. Horn, J. L. *Social situations–EP03A*. Unpublished test, University of Denver, 1967.

References

Anastasi, A. On the formation of psychological traits. *American Psychologist*, 1970, *25*, 899–910.

Balinsky, B. An analysis of mental factors of various age groups from nine to sixty. *Genetic Psychology Monographs*, 1941, *23*, 191–234.

Baltes, P. B., & Labouvie, G. V. Adult development of intellectual performance: Description, explanation, and modification. In C. Eisdorfer & M. P. Lawton (Eds.), *The psychology of adult development and aging*. Washington, D.C.: American Psychological Association, 1973.

Baltes, P. B., Nesselroade, J. R., & Cornelius, S. W. Multivariate antecedents of structural change in development. A simulation of cumulative environmental patterns. *Multivariate Behavioral Research*, 1978, *13*, 127–152.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. *Life-span developmental psychology: Introduction* *to research methods*. Monterey, Calif.: Brooks/Cole, 1977.

Bentler, P. M. Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 1980, *31*, 419–456.

Burt, C. The differentiation of intellectual abilities. *British Journal of Educational Psychology*, 1954, *24*, 76–90.

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

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

Cattell, R. B., & Cattell, A. K. S. *Measuring intelligence with the Culture-Fair Tests: Manual for Scales 2 and 3*. Champaign, Ill.: Institute for Personality and Ability Testing, 1961.

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

Cunningham, W. R. Principles for identifying structural differences: Some methodological issues related to comparative factor analysis. *Journal of Gerontology*, 1978, *33*, 82–86.

Cunningham, W. R. Age comparative factor analysis of ability variables in adulthood and old age. *Intelligence*, 1980, *4*, 133–149.

Cunningham, W. R., & Birren, J. E. Age changes in the factor structure of intellectual abilities in adulthood and old age. *Educational and Psychological Measurement*, 1980, *40*, 271–290.

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

Garrett, H. E. A developmental theory of intelligence. *American Psychologist*, 1946, *1*, 372–378.

Guilford, J. P. *Verbal analogies test, I*. Beverly Hills, Calif.: Sheridan Psychological Services, 1969. (a)

Guilford, J. P. *Word matrix test*. Beverly Hills, Calif.: Sheridan Psychological Services, 1969. (b)

Hayslip, B., Jr., & Sterns, H. L. Age differences in relationships between crystallized and fluid intelligences and problem solving. *Journal of Gerontology*, 1979, *34*, 404–414.

Horn, J. L. Organization of data on life-span development of human abilities. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory*. New York: Academic Press, 1970.

Horn, J. L. Psychometric studies of aging and intelligence. In S. Gershon & A. Raskin (Eds.), *Aging* (Vol. 2). New York: Raven Press, 1975.

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

Lienert, G. A., & Crott, H. W. Studies on the factor structure of intelligence in children, adolescents, and adults. *Vita Humana*, 1964, *7*, 147–163.

Long, S. Estimation and hypothesis-testing in linear models containing measurement error: A review of Jöreskog's model for the analysis of covariance structures. *Sociological Methods and Research*, 1976, *5*, 157–206.

Meredith, W. Notes on factorial invariance. *Psychometrika*, 1964, *29*, 187–206.

Merz, F., & Kalveram, K. T. Kritik der Differenzierungshypothese der Intelligenz. *Archiv für die Gesamte Psychologie*, 1965, *117*, 287–295.

O'Sullivan, M., & Guilford, J. P. *Social Translations, Form A*. Beverly Hills, Calif.: Sheridan Psychological Services, 1965.

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

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

Reinert, G. Comparative factor analytic studies of intelligence throughout the human life span. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory*. New York: Academic Press, 1970.

Reinert, G., Baltes, P. B., & Schmidt, L. R. Faktorenanalytische Untersuchungen zur Differenzierungshypothese der Intelligenz: Die Leistungsdifferenzierungshypothese. *Psychologische Forschung*, 1965, *28*, 246–300.

Reinert, G., Baltes, P. B., & Schmidt, L. R. Kritik einer Kritik der Differenzierungshypothese der Intelligenz. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 1966, *13*, 602–610.

Riegel, K. F., & Rosenwald, G. C. (Eds.). *Structure and function: Developmental and historical aspects*. New York: Wiley, 1975.

Rudinger, G. Correlates of changes in cognitive functioning. In H. Thomae (Ed.), *Patterns of aging: Findings from the Bonn Longitudinal Study*. New York: Karger, 1976.

Schneirla, T. C. Behavioral development and comparative psychology. *Quarterly Review of Biology*, 1966, *41*, 283–302.

Sörbom, D. Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology*, 1975, *28*, 138–151.

Sörbom, D., & Jöreskog, K. G. *COFAMM: Confirmatory factor analysis with model modification*. Chicago: National Educational Resources, 1976.

Thurstone, L. L. *Multiple factor analysis*. Chicago: University of Chicago Press, 1947.

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

Werner, H. *Comparative psychology of mental development*. New York: International Universities Press, 1948.

Received March 11, 1980 ■