Application of a Simplex Process Model to Six Years of Cognitive Development in Four Demographic Groups

Lloyd G. Humphreys, Randolph D. Park, and Charles K. Parsons
University of Illinois, Urbana–Champaign

A simplex process model of the cross-lagged correlation paradigm was applied to 16 tests administered to samples of black and white males and black and white females in Grades 5, 7, 9, and 11. Listening, a measure of aural comprehension, consistently anticipated individual differences on an intellectual composite in all four groups. The other achievement test of the STEP series anticipated individual differences on the so-called aptitude tests of SCAT, which in turn anticipated individual differences on the narrow information scores obtained from the Test of General Information (TGI). This model may be more powerful in revealing lags than the traditional methods of analyzing cross-lagged differences in longitudinal data. The model does not require stationarity and can produce a meaningful outcome in its absence.

The cross-lagged panel correlation methodology described by Campbell and Stanley (1963) requires as a minimum two measures—X and Y— and two occasions of measurement for both. When certain conditions are met, if $r_{x_1 y_2} > r_{x_2 y_1}$, an inference that individual differences on X anticipate individual differences on Y is valid. It is only a small extension of this inference to conclude that X assesses the causes for the changes in Y prior to the changes in Y, or that the predominant direction of causality is from X to Y. A conclusion that X causes changes in Y, however, clearly requires additional evidence.

There are six correlations among two measures, each measured on two occasions. The two correlations between X and Y holding occasion constant are typically called synchronous correlations; the two correlations between occasions for X and Y separately are stability coefficients; and the final two between X at Time 1 with Y at Time 2 and Y at Time 1 with X at Time 2 are the cross-lagged correlations. Unfortunately, these latter correlations can differ from each other if the reliability of X decreases over time while the reliability of Y increases, or the reverse. If it can be assumed, however, that the causal processes operate on both X and Y in a stationary manner (see Kenny, 1975) from Occasion 1 to Occasion 2, differences in reliability can be corrected and the corrected values can be compared.

Since many psychometricians are not accustomed to thinking of stationary causal processes, it is useful to translate this term into correlational statistics. If causal processes are stationary, the synchronous correlations at Time 1 and Time 2 between X and Y should equal each other after they have been corrected for attenuation. This is the basis for Kenny's correction for uniqueness. In order for the estimated correlations between true scores of X and Y on separate occasions to be equal to each other, it is neces-

sary that the estimated stabilities of true scores of X and Y be equal. If individual differences on X are changing at a more rapid rate than individual differences on Y (i. e., if the causal processes are not stationary), it is not possible for the two synchronous true score correlations to be equal.

Investigators using the methodology have been critically dependent on the validity of the assumption of stationarity. Lack of stationarity can simulate a cross-lagged difference in true score correlations. On the other hand, a correction (which assumes stationarity) applied to observed correlations can abolish evidence for a lag. More recently, Humphreys and Parsons (in press) have described an alternative model which does not require stationarity. While the model could be applied in a limited way for only two occasions if independent estimates of reliabilities were available, more occasions are highly desirable. A minimum of four occasions are required if reliabilities and specificities are to be estimated from the intercorrelations alone.

The model assumes that the intercorrelations of true scores over occasions follow the simplex pattern (Guttman, 1955; Humphreys, 1960). When this assumption is valid, reliabilities can be estimated from the intercorrelations without exhausting degrees of freedom down to a minimum of four occasions. With estimates of reliabilities in the diagonal, whether obtained independently or based on the simplex assumption, multiple matrices of diagonal factors for X and Y are computed separately.

Choice of the diagonal method of factoring is the logical consequence of the assumption concerning the nature of the pattern of intercorrelations. Guttman (1955) had found that the intercorrelations of tests measuring the same content but differing in degree of complexity formed simplex matrices. Thus, diagonal factors extracted in the order of the measures in the simplex describe the changes in complexity from one measure to another. Humphreys (1960) applied the simplex to learning and developmental measures obtained over several occasions lon-

gitudinally. Here again, the diagonal factors extracted in the order of the occasions described the learning and growth increments that occurred between occasions.

The multiple matrices of diagonal factors are needed to test the possibility of lag. They are obtained by starting each set of factors with a first factor defined by a different occasion and proceeding systematically forward and/or backward in time from the initial occasion. The diagonal factors of X can then be equated hypothetically with diagonal factors in Y without any lag or with a lag of up to $n - 1$ occasions in either direction. That is, one can assume that X anticipates Y by one or more occasions, that Y anticipates X by one or more occasions, or that neither anticipates the other.

The final step is to determine which pairing of diagonal factor matrices most accurately describes the cross-correlations between X and Y. These correlations are estimated by postmultiplying a standard diagonal factor matrix for one variable by the series of factor matrices, in turn, of the other and by allowing for the absence of complete identity arising from the specificities of the two variables being compared. Note that causes operating on both variables must have their effects on common factor content. Now, if expected values based on the assumption of a lag fit the observed correlations more accurately than the ones based on no lag, one has evidence that individual differences in one variable anticipate individual differences in the other.

In the development of their model, Humphreys and Parsons analyzed a portion of the data used by Atkin, Bray, Davison, Herzberger, Humphreys, & Selzer (1977). These latter authors published a cross-lagged panel correlation analysis of 16 cognitive measures for 4 demographic groups defined by race (black and white) and sex in which stationarity was assumed. Complete data were available for all members of the four samples in Grades 5, 7, 9, and 11. They found only one measure which was clearly different from the rest with respect to its cross-lagged difference. Listening, from the

STEP series published by Educational Testing Service (ETS), predicted a latter intellectual composite more highly at all lags than the same intellectual composite predicted later scores on the Listening test. One or two of the subtests of the Test of General Information (TGI), also published by ETS, showed less consistent cross-lagged differences in the opposite direction. That is, individual differences in a measure of aural comprehension anticipated individual differences in an intellectual composite, while the latter seemingly anticipated individual differences in one or two narrow information tests.

With respect to methods of analysis other than those associated with attempting to control for stationarity, the analyses of Atkin et al. (1977) were not traditional. They did not compare the various tests with each other in a pairwise fashion. Instead they formed rotating intellectual composites and compared each separate measure with a composite formed from the remaining 15. This methodology reduced the number of possible comparisons, thus reducing the opportunity to capitalize on chance. The composites were each highly reliable and had high communality with the omitted measure. Each composite also was a highly valid measure of the general factor in human intelligence.

Humphreys and Parsons (in press) in their test of the simplex process model for such data, combined the white male and female sample correlations for Listening and its composite. They selected these particular data because the factor structure of the variables was known to be almost identical in white males and females; the Listening test, which had produced the largest and most consistent cross-lagged differences, was the one of greatest interest; and in the combined samples ($N > 1400$), the correlations had a great deal of sampling stability. Under these conditions, Humphreys and Parsons reported an excellent fit of the model to the observed correlations for a three-year lag between Listening and the intellectual composite. A secondary finding concerned the assumption of stationarity. Although not assumed in the development

of the model, it turned out empirically that the estimated true score correlations among occasions were very nearly identical for Listening and for the composite. That is, individual differences were changing in the two measures being compared at almost the same rate.

With the model producing a good fit to the observed correlations under favorable conditions of high sampling stability and highly consistent cross-lagged differences revealed by a more traditional methodology, it now seems desirable to apply it further. The model can be applied, for example, to test the hypothesis of zero lag between two variables; i.e., it can be used as a test of spuriousness in the Kenny sense. Also, since lack of stationarity as well as changes in uniqueness can produce a spurious cross-lagged difference, and since the simplex process model does not require stationarity, it would be desirable to apply it in situations in which stationarity from measure to measure would not hold. The data used by Atkin et al. (1977) are appropriate for these purposes: 16 widely different cognitive measures (including Listening) in 4 sex-race groups were available for further applications of the model.

Design of the Analyses

Following the lead of Atkin et al., relationships were examined between each of the 16 cognitive tests in each of the 4 race-sex groups, with a composite formed from the remaining 15. While a standard, uniform criterion measure of general intellectual level would have been desirable, one was not available. With a rotating criterion, lags become relative; i.e., if most of the 16 tests exhibited individual differences which lagged behind individual differences on a standard test of intelligence, the present methodology would not reveal that fact. Instead, the typical lag would be close to zero. The methodology will allow one, however, to rank order the measures from those which show the largest amount of anticipation to those which show the greatest lags behind the rest.

The descriptive statistics used also followed the lead of the earlier research. Correlations between individual tests and composites are multiple correlations, and correlations between composites are canonicals. Both, of course, capitalize on chance in a sample; but the error is not very great when the number of variables is much smaller than the number of observations. *N*s available were as follows: white males—668, white females—762, black males—172, black females—215. Thus, there was more capitalization on chance in the black samples than in the white. There was also more capitalization on chance in the canonicals than in the multiple correlations. Conversion of sample statistics to population estimates would have been desirable, but there is no population estimate for a canonical correlation.

In fitting the model of the 64 cross-lagged comparisons resulting from 16 tests and 4 different samples, the first step involved the use of the Jöreskog (1970) restricted maximum likelihood program to estimate the reliabilities of the measures. Specificities were estimated by comparing these reliabilities with squared multiples. Diagonal factors were computed directly or estimated by the procedure described by Humphreys and Parsons (in press) for nine different lags between test and composite by four years down to zero lag, and then back up to the composite anticipating the test by four years.

Table 1 contains the names of the 16 cognitive tests used in these analyses. The School and College Ability Test (SCAT), from which the seventh and eighth measures were obtained, is well known, as are the Scholastic Tests of Educational Progress (STEP) represented by the first six. Tests 9 through 16 were obtained from the ETS's Test of General Information (TGI). In those cases in which there is similarity of names with tests in STEP, the TGI tests are narrower in scope. The latter were intended to be tests of information only.

Data for all correlations were obtained from the Growth Study of Educational Testing Service. The total sample was designed to represent the nation's students, but conditions were imposed which may have subjected the means and standard deviations to some degree of bias. The conditions required not only complete data for all tests administered at any one grade but complete data for all four grades.

Table 1

Identification of the Measures

1. Mathematics (STEP-M)[a]	10. Home Arts (TGI-HA)
2. Physical Science (STEP-PS)	11. Physical Science, Mathematics (TGI-PSM)
3. Social Science (STEP-SS)	
4. Reading (STEP-R)	12. Biological Science (TGI-BS)
5. Listening (STEP-L)	13. Music, Art (TGI-MA)
6. Writing (STEP-W)	14. History, Literature (TGI-HL)
7. Verbal (SCAT-V)[b]	15. Recreation, Entertainment (TGI-RE)
8. Quantitative (SCAT-Q)	
9. Industrial Arts (TGI-IA)[c]	16. Government, Public Affairs (TGI-GPA)

[a]Sequential Tests of Educational Progress
[b]School and College Ability Tests
[c]Tests of General Information

Results

Table 2 presents data obtained from the use of the Jöreskog (1970) program. On the left are the estimates of true score correlations between scores on adjacent time periods. On the right are the chi-square values testing goodness of fit of the simplex model to the obtained intercorrelations. With four time periods, there was one degree of freedom for testing goodness of fit.

There is useful psychological and methodological information in the estimates of true score correlations between adjacent time periods. In a few cases, this estimate was unity, which is unrealistic psychologically. The methodological significance is that it was not possible to check the simplex model of change in such cases. This outcome was restricted to the black samples, which are both smaller than the white. At certain grades, also, black performance on several tests was less reliable than white performance simply because difficulty levels of the items were inappropriate. Thus, the combination of larger sampling errors and larger corrections for attenuation probably produced these unrealistic results.

One similar example occurred in the white female sample. The estimated true score correlation for the Mathematics score from STEP was .995. Sampling error was about half that in the black samples. Whatever the reason may be for this seeming high stability over time, the methodological result is that the fitting of a model

Table 2

Estimated True Score Correlations Between Adjacent Time Periods
and X^2 for Goodness of Fit of the Simplex Individual Tests
and Canonical Composites

| | | Correlations | | | | X^2 Values | | | |
| | | White | | Black | | White | | Black | |
	Test	M	F	M	F	M	F	M	F
1.	STEP-M	93	99	89	97	01	00	1.11	1.24
2.	STEP-PS	94	93	92	96	1.38	3.88	28	3.46
3.	STEP-SS	96	97	91	1.00	1.76	1.57	44	2.84
4.	STEP-R	96	96	94	96	7.51	12.85	06	82
5.	STEP-L	95	96	92	94	3.56	60	71	27
6.	STEP-W	97	94	94	97	32	1.75	26	00
7.	SCAT-V	95	96	96	95	3.45	34	44	23
8.	SCAT-Q	92	89	78	84	03	1.70	1.67	78
9.	TGI-IA	85	77	69	91	1.04	15	09	2.54
10.	TGI-HA	88	85	72	84	2.87	04	1.62	00
11.	TGI-PSM	81	82	1.00	66	16	00	28	77
12.	TGI-BS	86	86	82	81	00	86	00	07
13.	TGI-MA	87	87	1.00	90	36	10	81	2.92
14.	TGI-HL	89	86	1.00	87	26	12	10	27
15.	TGI-RE	85	90	80	86	1.01	71	1.64	00
16.	TGI-GPA	94	80	86	75	1.08	02	08	1.97
	Composites	97	97	96	97	43	72	1.22	69

Note: Decimal points omitted for values less than 1.0.

which assumes change and the possibility of different lags between changes in individual differences from test to test cannot lead to any very definitive result for this test for this sample.

If one of the most important differences among cognitive tests is breadth of coverage (Humphreys, 1971), a logical inference is that broad tests should generally show more stability over time than narrow tests. The eight scores from the TGI were designed to be narrow in coverage and in function measured. The other eight tests did, with one notable exception, typically show relatively higher degrees of stability. The exception was Quantitative Aptitude from SCAT. The latter test did not show one of the characteristics, high stability, that might be expected from the connotations of the aptitude construct. The information tests, with the exception of three aberrant cases for black males, typically showed the lower level of stability which is in line with expectations.

There were only two values of chi-square out of 68 comparisions that were large enough for rejection of the null hypothesis at the .01 level; but both occurred for the same test, Reading Comprehension, in the two white samples. Furthermore, the signs of the residuals were identical in both samples. These white students showed slightly less stability than predicted by the model over a four-year interval, but slightly more stability than predicted over the six-year interval. Even though the model could be rejected, however, the fit was descriptively quite good. The single largest residual was .023 in the female sample for the 5th and 11th grade correlation.

Because chi-square is a function of the number of observations, the fits which were defined descriptively, rather than probabilistically, were not quite as close in the black as in the white samples. Thus, there were numerous cases in which the residuals were larger than .02 in the data for blacks. This characteristic of the four samples became important later when the simplex process model was being fitted to the observed correlations. The goodness of fit of the model to the observations is described by a statistic which is independent of N. Therefore, more error must be expected in the black samples.

Tables 3, 4, 5, and 6 present indices of goodness of fit of the model to the observed cross-correlations for white males and females and for black males and females. A plus lag tests the hypothesis that individual differences on the test anticipate individual differences on the composite, while a negative lag tests the reverse hypothesis. Both the algebraic sum of deviations and the sum of squared deviations are presented for the 12 correlations representing 2, 4, and 6 years of elapsed time between the measurement occasions. Discrepancies between estimates from the model and the observed values of the four synchronous correlations were omitted because these four correlations were each assumed to be inflated by correlated measurement error. More attention is paid to the sums of the squared errors than to the algebraic sums in interpreting the results. The latter indicate the amount of constant error in the application of the model. The most probable source of constant error is the estimation of specificities of the measures. However, the minimum sum of squared deviations among the nine lags provides an index of the lag between individual differences on the measures, even when sums are inflated by constant error.

Most of the indices of goodness of fit reported in these tables are based on the observed correlations among occasions, while some are based on the estimates. Diagonal factor matrices of full rank (4 x 4) were obtained whenever possible in both, but for a few tests a factor matrix of full rank could not be obtained in the observed. In these cases, the model was applied to the cross-correlations using the factors in the estimated correlations among occasions. There is no bootstrapping involved here, since the factoring of the latter is completely independent of the cross-correlations. Also, in those cases in which diagonal factoring of full rank could be accomplished on both bases, it was possible to

Table 3
Fit of the Model for 668 White Males at Nine Different Lag Times

Test		+4	+3	+2	+1	0	-1	-2	-3	-4
1. STEP-M	Σd	226	124	022	-033	-028	-077	-126	-017	092
	Σd^2	040	025	016	007	003	006	024	041	072
2. STEP-PS	Σd	077	-008	-093	-111	-130	-156	-182	-081	020
	Σd^2	028	019	016	007	003	009	025	038	062
3. STEP-SS	Σd	054	-035	-124	-125	-127	-134	-141	-062	016
	Σd^2	026	017	012	005	003	005	013	016	024
4. STEP-R	Σd	-021*	-107*	-193*	-201*	-209*	-226*	-242*	-153*	-064*
	Σd^2	031*	022*	019*	010*	006*	009*	017*	022*	034*
5. STEP-L	Σd	024	-048	-120	-136	-151	-164	-177	-087	003
	Σd^2	004	003	005	008	016	031	054	074	101
6. STEP-W	Σd	-246	-305	-365	-354	-344	-335	-326	-257	-188
	Σd^2	036	028	024	017	013	012	014	013	015
7. SCAT-V	Σd	313	229	145	135	125	109	094	191	287
	Σd^2	065	049	038	022	011	004	004	010	025
8. SCAT-Q	Σd	298	190	082	045	009	-065	-138	-029	080
	Σd^2	089	072	061	037	020	009	016	028	059
9. TGI-IA	Σd	416	302	188	150	112	013	-086	058	201
	Σd^2	130	107	092	065	044	015	028	039	090
10. TGI-HA	Σd	649	554	458	399	339	228	116	216	316
	Σd^2	213	187	166	125	093	038	026	031	076
11. TGI-PSM	Σd	1.415	1.290	1.165	1.108	1.051	906	760	891	1.022
	Σd^2	427	376	335	274	223	121	090	108	191
12. TGI-BS	Σd	722	606	492	448	404	296	187	335	483
	Σd^2	178	147	124	092	067	023	022	044	108
13. TGI-MA	Σd	894	772	651	603	554	437	320	439	558
	Σd^2	245	209	182	140	105	046	024	027	065
14. TGI-HL	Σd	710	594	478	437	396	305	214	333	451
	Σd^2	151	121	098	069	048	019	017	029	068
15. TGI-RE	Σd	854	737	619	567	515	390	265	398	531
	Σd^2	319	282	253	202	158	078	043	029	060
16. TGI-GPA	Σd	127	035	-057	-071	-085	-111	-137	-050	038
	Σd^2	045	033	025	013	006	003	009	012	023

*Measures were analyzed using the estimated simplex matrix for the diagonal
factoring. Observed correlations were used unless a full diagonal solution
could not be obtained. Substitutions of estimated correlations for the
observed has no systematic effect on the data.

compare indices of goodness of fit using both sets of factors. No consistent differences in accuracy as a function of the correlations factored were observed. As a matter of fact, it was disappointing to discover that the factors based on estimated correlations among occasions did not provide greater accuracy in estimating the cross-correlations between tests and the composite.

In a very few cases, indices of goodness of fit are not reported in these tables. The absence of data indicates that it was not possible to obtain a diagonal matrix of full rank on either basis.

The first thing to observe in these four tables is that there was a relatively smooth change in the sum of the squared deviations from a possible four-year lag from test to composite to a

Table 4
Fit of the Model for 762 White Females at Nine Different Lag Times

Test		+4	+3	+2	+1	0	-1	-2	-3	-4
1. STEP-M	Σd	-273	-304	-336	-294	-251	-234	-217	-187	-157
	Σd²	011	012	015	017	021	023	025	025	026
2. STEP-PS	Σd	180	097	014	-012	-034	-073	-108	-006	095
	Σd²	035	024	018	007	004	007	023	037	065
3. STEP-SS	Σd	-049	-109	-169	-165	-161	-156	-151	-074	004
	Σd²	016	010	008	004	003	006	013	017	026
4. STEP-R	Σd	-006*	-089*	-171*	-175*	-179*	-190*	-202*	-118*	-035*
	Σd²	032*	022*	016*	008*	005*	006*	013*	016*	025*
5. STEP-L	Σd	053	-031	-114	-117	-121	-131	-141	-065	010
	Σd²	011	007	008	008	013	024	039	050	066
6. STEP-W	Σd	071	-032	-134	-154	-174	-214	-253	-162	-071
	Σd²	080	064	054	034	020	009	008	009	020
7. SCAT-V	Σd	362	274	186	179	172	157	142	225	308
	Σd²	063	045	032	019	010	005	006	013	026
8. SCAT-Q	Σd	493	382	272	227	182	073	-035	088	212
	Σd²	139	116	100	069	046	021	025	036	075
10. TGI-HA	Σd	661	546	432	381	330	205	080	188	297
	Σd²	236	206	184	142	107	042	022	022	067
11. TGI-PSM	Σd	954	847	740	690	639	525	411	518	626
	Σd²	337	300	272	221	178	083	047	050	105
12. TGI-BS	Σd	586	470	354	315	276	174	071	208	346
	Σd²	164	138	119	087	062	019	015	027	078
13. TGI-MA	Σd	736	616	496	446	395	267	139	257	375
	Σd²	249	215	189	146	110	047	021	020	056
14. TGI-HL	Σd	765	644	524	477	429	305	181	308	434
	Σd²	218	183	156	120	091	033	016	028	080
15. TGI-RE	Σd	451	350	249	208	167	079	-009	097	203
	Σd²	157	134	118	036	062	028	016	010	024
16. TGI-GPA	Σd	798	680	562	515	467	316	164	324	485
	Σd²	302	269	245	198	159	071	047	047	113

*Measures were analyzed using the estimated simplex matrix for the diagonal factoring. Observed correlations were used unless a full diagonal solution could not be obtained. Substitutions of estimated correlations for the observed has no systematic effect on the data.

possible four-year lag from composite to test. Each test with the exception of the Mathematics test from STEP for white females also showed a minimum for this summation within the time span studied. The exception is probably not an important one because the estimate of the true score correlation between adjacent occasions was very high (.995). This estimate makes the model insensitive to lag; it is also so high that it is almost certainly not replicable.

As one reads down each of the four tables searching for the minimum sum, it becomes apparent that there was a preponderance of negative lags; i.e., individual differences on the composite more frequently anticipated individual differences on the test than the reverse. Since the composite was formed from the tests with which it is compared, the a priori expectation of a balance between positive and negative lags was not confirmed. At least part of the

Table 5
Fit of the Model for 172 Black Males at Nine Different Lag Times

Test		+4	+3	+2	+1	0	-1	-2	-3	-4
1. STEP-M	Σd	486	388	290	270	251	222	194	293	393
	Σd^2	110	085	068	045	032	015	020	036	079
2. STEP-PS	Σd	209	112	015	-002	-020	-052	-083	023	129
	Σd^2	080	062	052	030	017	013	025	038	067
3. STEP-SS	Σd	395	263	130	100	071	006	-059	064	187
	Σd^2	115	083	062	036	018	008	022	032	062
4. STEP-R	Σd	164	059	-046	-061	-076	-107	-139	-032	075
	Σd^2	055	040	032	018	012	017	034	042	062
5. STEP-L	Σd	034	-060	-154	-179	-204	-244	-284	-199	-133
	Σd^2	018	016	019	024	037	066	109	144	192
6. STEP-W	Σd	006	-080	-167	-179	-191	-211	-230	-137	043
	Σd^2	044	031	024	012	008	010	022	032	053
7. SCAT-V	Σd	200	120	038	044	050	050	050	139	228
	Σd^2	065	048	036	022	014	013	018	022	033
8. SCAT-Q	Σd	1.313	1.168	1.022	960	897	683	470	608	746
	Σd^2	344	296	281	200	152	082	089	123	229
9. TGI-IA	Σd	1.346	1.242	1.137	1.102	1.068	967	866	977	1.089
	Σd^2	376	337	308	260	220	152	126	154	301
10. TGI-HA	Σd	1.846	1.746	1.646	1.579	1.511	1.338	1.165	1.224	1.282
	Σd^2	712	665	627	536	460	309	267	251	312
12. TGI-BS	Σd	1.019	884	749	700	651	516	381	508	634
	Σd^2	289	242	208	158	119	043	028	063	155
15. TGI-RE	Σd	1.368	1.240	1.113	1.050	987	817	646	758	870
	Σd^2	481	429	388	311	249	138	099	094	159
16. TGI-GPA	Σd	610	504	398	345	292	193	094	152	211
	Σd^2	181	149	125	095	074	038	036	082	117

explanation for the asymmetry lies in the reliabilities, stabilities, and intercorrelations of the tests. If the 16 tests are divided into 2 sets of 7 and 9 based on the typical level of stability from one occasion to another, the more stable 7 had substantially higher canonical weights in the composite. They also had higher intercorrelations and larger loadings on the general factor. Thus, they are more central to the composite, while the remaining 9 are more peripheral. The more peripheral tests tended to show the negative lags.

It is also possible that the use of unshrunken composites contributed to the displacement of the typical lag from zero to negative. There were appreciable differences among the correlations used in these analyses in the extent to which they capitalize on chance. This means that greater stress in interpretations should be placed on relative size of lags from test to test than on the number of years indicated for a particular lag.

Listening, as in the earlier analyses, consistently anticipated the composite in the four samples. It was consistently first in rank order in this regard for both white and black males. It was tied for first rank for black females and was ranked second for white females. However, the first ranked test for this last group was Mathematics achievement, which did not show enough instability from occasion to occasion for the minimum sum of squared deviations to be sensitive to differences in lag. There may also be a

Table 6
Fit of the Model for 215 Black Females at Nine Different Lag Times

Test		+4	+3	+2	+1	0	-1	-2	-3	-4
1. STEP-M	Σd	236*	181*	126*	130*	133*	134*	134*	192*	244*
	Σd²	019*	014*	012*	010*	012*	015*	021*	028*	036*
2. STEP-PS	Σd	222*	150*	079*	072*	065*	048*	031*	104*	192*
	Σd²	026*	019*	016*	011*	009*	012*	020*	026*	046*
4. STEP-R	Σd	133	068	003	-011	-025	-041	-055	024	103
	Σd²	033	024	018	009	004	003	006	009	018
5. STEP-L	Σd	312	220	129	104	081	035	-011	070	151
	Σd²	018	010	006	003	005	016	037	054	080
6. STEP-W	Σd	-073	-138	-202	-202	-201	-204	-208	-146	-084
	Σd²	045	035	029	019	012	008	007	006	009
7. SCAT-V	Σd	610	516	421	398	374	327	281	364	447
	Σd²	074	055	041	028	020	013	016	029	051
8. SCAT-Q	Σd	982	866	749	697	644	475	307	411	516
	Σd²	316	278	248	202	162	097	079	062	098
9. TGI-IA	Σd	235*	160*	084*	061*	037*	-023*	-084*	003*	091*
	Σd²	114*	098*	086*	070*	056*	034*	023*	016*	021*
10. TGI-HA	Σd	1.266	1.148	1.029	975	922	758	593	700	807
	Σd²	343	301	267	222	185	094	058	068	131
11. TGI-PSM	Σd	1.980*	1.900*	1.819*	1.782*	1.745*	1.669*	1.420*	1.465*	510*
	Σd²	693*	651*	614*	564*	519*	405*	332*	246*	278*
12. TGI-BS	Σd	1.051	929	807	756	704	532	359	490	621
	Σd²	389	344	309	260	218	102	059	048	108
13. TGI-MA	Σd	573*	472*	372*	330*	288*	189*	116*	184*	252*
	Σd²	133*	114*	099*	078*	063*	030*	022*	023*	048*
14. TGI-HL	Σd	992	885	778	726	674	530	386	496	605
	Σd²	200	171	149	115	008	041	028	044	097
15. TGI-RE	Σd	799	689	579	528	476	327	178	263	348
	Σd²	170	148	132	099	072	042	046	052	093
16. TGI-GPA	Σd	1.407	1.280	1.153	1.108	1.062	908	754	845	936
	Σd²	676	621	580	515	456	273	181	093	107

*Measures were analyzed using the estimated simplex matrix for the diagonal factoring. Observed correlations were used unless a full diagonal solution could not be obtained. Substitutions of estimated correlations for the observed has no systematic effect on the data.

sex difference in the amount of lag between Listening and the intellectual composite. The symmetries about the minimum sum of squared deviations indicate that the lag was at least three years for males, while it was at most three and one, respectively, for white females and black females.

As a set, the STEP achievement tests tended to anticipate the SCAT aptitude tests, while the latter in turn anticipated the narrow information tests of the TGI. This finding requires an extension of the number of dimensions along

which different kinds of cognitive[1] tests can be ordered. Anticipating or being anticipated by individual differences on other measures is undoubtedly not unrelated to breadth, relationship to a particular curriculum, or age of the learning (Humphreys, 1971); but it does represent a fourth descriptive dimension. On this dimension, tests which are generally labeled achievement (STEP), but which are not restricted to recall of information, more nearly fit the connotations of aptitude (SCAT). Broad achievement tests more directly reflect intellectual development than Verbal and Quantitative aptitude.

One has to look hard for any racial differences; the prime candidate is Industrial Arts, but there are no data for white females. There are, however, two measures other than Listening on which there were possible sex differences. The Mathematics achievement test appeared to anticipate later individual differences on the composite for females but not for males, although the qualification for white females discussed earlier must be kept in mind. Economics and Political Science information was anticipated by the composite for females with a three-year lag, while the lag was seemingly smaller for males.

Discussion

The analysis by Atkin et al. (1977) revealed only one measure, Listening, that differed unequivocally from the rest. Both in the observed and in the correlations corrected for uniqueness, Listening showed consistent cross-lagged differences from one occasion to another and from group to group. Individual differences in Listening anticipated the composite. There were only hints in the analysis that there might be other differences. For example, data were presented for Recreation, which indicated that individual differences on that measure lagged behind the composite. In contrast, the present analysis, which made use of a different model, showed a larger number of cross-lagged differences which were quite substantial in size. The explanation lies in the stationarity assumption.

As discussed earlier, Kenny's method of correcting cross-lagged differences for differences in uniqueness depends on the equality of all synchronous correlations for a given pair of measures after they have been corrected for attenuation. When individual differences on one measure are changing at a more rapid rate than on another, such correlations cannot be equal.

Some readers may be disturbed by the lack of tests of significance to support the fit of the model to the observed cross-correlations. While the simplex matrix provides a good fit both descriptively and statistically to intercorrelations of measures over occasions, only a descriptive test is available for the cross-correlations. Furthermore, the errors for some of these were fairly large. Both constant and random errors are involved.

The most probable source of large constant errors is the procedure for estimation of specificities. The difference between an estimate of reliability and a squared multiple correlation represents one method of estimation, but it is well known that squared multiples are lower bound estimates of communalities. The error in the specificity estimate also tends to be larger the lower the reliability of the variable. The model also requires estimates of true score correlations in order to estimate reliabilities which in turn are used in the diagonal factoring to obtain correlations between fallible measures and factors which represent the true scores of those

---

[1] There is no single term generally agreed to which covers the gamut of so-called aptitude, ability, achievement, information, and intelligence tests. Cognitive may be a reasonable choice for a generic term. More concretely, if the reader is puzzled by our meaning, we are referring essentially to tests which have correct answers; i.e., in which there is some modicum of problem solving. Thus, the number of attempts on a highly speeded clerical test containing very simple items might not qualify as a cognitive measure, but a formula score involving a correction to numbers of correct answers based on the number of errors would qualify.

Table 7

Reliability Estimates of the 16 Tests and the Composite
at Four Grade Levels for White Males and Females

Test	Males				Females			
	5	7	9	11	5	7	9	11
1. STEP-M	755	817	823	691	577	694	684	571
2. STEP-PS	734	759	759	709	781	712	706	638
3. STEP-SS	752	821	806	762	752	846	745	733
4. STEP-R	745	767	805	766	748	823	734	771
5. STEP-L	771	819	748	653	711	817	757	730
6. STEP-W	728	745	771	672	767	830	738	778
7. SCAT-V	848	922	884	876	836	947	893	906
8. SCAT-Q	852	808	803	889	787	814	726	958
9. TGI-IA	712	598	582	692	766	388	484	100
10. TGI-HA	767	389	585	803	726	460	687	761
11. TGI-PSM	947	448	676	968	847	297	443	818
12. TGI-BS	823	591	551	785	731	624	530	735
13. TGI-MA	773	596	659	921	757	608	728	934
14. TGI-HL	737	726	726	920	694	663	599	857
15. TGI-RE	855	637	712	899	566	551	699	896
16. TGI-GPA	624	640	536	686	723	716	546	644
Composite	945	968	968	956	931	992	947	973

measures. Whenever reliabilities are low to moderate, corrections for measurement error are made with a great deal more error than when reliabilities are high. Thus, the estimates of reliabilities used in this model fitting are relevant to an evaluation of the goodness of fit. These reliabilities appear in Tables 7 and 8.

For some of the measures, there was a good deal of variability in reliability from one occasion to another. Since more than one test form had to be used to span the grades represented in these data, the best guess is that certain forms were either too easy or too difficult at certain grade levels. The appropriateness of difficulty levels would vary from race to race and, for some of the tests, from sex to sex as well. There were also clear differences from test to test irrespec-

tive of grade. The TGI tests, being shorter, consistently had lower reliabilities than the others. The test of so-called Verbal Aptitude, as would be expected, had consistently high reliabilities.

When the reliability estimates reported in these tables are compared with the errors of fit in Tables 3 through 6, there is an obvious relationship. Errors are smaller for the more reliable tests of STEP and SCAT. Although errors tended to be large for the TGI scores, the consistency of results from test to test and from group to group still allows one to have confidence in the trends revealed.

Up to this point the data have been interpreted in the most conservative manner; i. e., in terms of anticipation or being anticipated. This choice does not mean that any causal relation-

Table 8

Reliability Estimates of the 16 Tests and the Composite
at Four Grade Levels for Black Males and Females

Test	Males				Females			
	5	7	9	11	5	7	9	11
1. STEP-M	610	564	808	461	632	433	760	297
2. STEP-PS	716	558	716	508	653	664	706	491
3. STEP-SS	822	694	733	787	611	692	709	672
4. STEP-R	738	728	714	739	807	806	733	818
5. STEP-L	785	621	767	660	781	869	797	720
6. STEP-W	685	604	697	668	734	771	764	710
7. STEP-V	810	837	867	762	777	920	920	920
8. SCAT-Q	977	594	696	1000	803	781	861	930
9. TGI-IA	1000	222	513	910	379	402	323	317
10. TGI-HA	972	228	465	892	773	508	794	913
11. TGI-PSM	185	080	558	506	1000	075	404	1000
12. TGI-BS	787	436	540	727	836	524	578	780
13. TGI-MA	267	267	613	514	546	490	721	710
14. TGI-HL	230	216	529	444	519	676	593	791
15. TGI-RE	601	410	687	1000	462	531	667	690
16. TGI-GPA	702	204	607	722	579	458	454	1000
Composite	949	906	980	981	933	976	976	957

ship has been precluded between tests that anticipate the composite and tests that are anticipated by the composite. A causal relationship may be present, but it requires a design that goes beyond the limits of longitudinal correlational data. An hypothesis that specifies a causal link between aural comprehension and intellectual development in terms of attentive behaviors in the home and in the classroom which are tapped more directly by the Listening tests than any other is plausible. A contrasting hypothesis that aural comprehension is more closely tied to a general factor, genetically determined, is also plausible. Both hypotheses converge, however, on a recommendation that more research should be concentrated on measures of aural comprehension during cognitive development.

References

Atkin, R., Bray, R., Davison, M., Herzberger, S., Humphreys, L., & Selzer, U. Cross-lagged panel analysis of sixteen cognitive measures at four grade levels. *Child Development*, 1977, *48*, 944–952.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally & Company, 1963.

Guttman, L. A generalized simplex for factor analysis. *Psychometrika*, 1955, *20*, 173–192.

Humphreys, L. G. Investigations of the simplex. *Psychometrika*, 1960, *25*, 313–323.

Humphreys, L. G. Theory of intelligence. In R. Cancro (Ed.), *Intelligence—genetic and environmental influences*. New York: Grune & Stratton, 1971.

Humphreys, L. G., & Parsons, C. K. A simplex process model to describe differences between

cross-lagged correlations. *Psychological Bulletin,* in press.

Jöreskog, K. G. Estimation and testing of simplex models. *The British Journal of Mathematical and Statistical Psychology,* 1970, *23,* 121–145.

Kenny, D. A. Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin,* 1975, *82,* 887–903.

Acknowledgements

Author's Address

Requests for reprints should be sent to Lloyd G. Humphreys, 425 Psychology Building, University of Illinois, Champaign, IL 61820.