DOCUMENT RESUME

ED 058 323 24 TM 001 052

AUTHOR Hopkins, Kenneth D.
TITLE The Stability and Change of Language and Non-Language IQ Scores. Final Report.
INSTITUTION Colorado Univ., Boulder.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
BUREAU NO BR-0-H-024
PUB DATE Dec 71
GRANT OEG-8-70-0007(509)
NOTE 41p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Correlation; Factor Analysis; *Group Intelligence Tests; *Intelligence Quotient; Multiple Regression Analysis; *Nonverbal Tests; Predictive Ability (Testing); Scores; Sex Differences; Statistical Analysis; Student Testing; Testing Problems; Test Interpretation; *Test Reliability; *Verbal Tests

ABSTRACT

The stability and change of verbal, non-verbal, and total IQ scores from group tests were investigated for students tested at Grades 1, 2, 4, 7, 9, and 11. Conclusions are discussed in detail, and the following recommendations are made: (1) Group intelligence tests should not be routinely administered in Grades 1 and 2 unless the users and interpreters of these scores have received extensive training in their meaning and limitations; (2) Non-language and non-verbal IQ scores must be interpreted even more cautiously than verbal IQ scores; (3) Teachers and other users of IQ scores from group tests need to be made aware of the fact that even when there is considerable overall stability, scores of a few individuals will vary greatly; and (4) No meaning should be attached to differences in verbal and non-verbal IQ scores prior to Grade 7. (AG)

FINAL REPORT
Project No. 0-H-024
Grant No. OEG-8-70-0007(509)

THE STABILITY AND CHANGE OF LANGUAGE AND NON-LANGUAGE IQ SCORES

Kenneth D. Hopkins
University of Colorado
Boulder, Colorado 80302

December 1971

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report

Project No. 0-H-024
Grant No. OEG-8-70-0007(509)

THE STABILITY AND CHANGE OF LANGUAGE AND NON-LANGUAGE IQ SCORES

Kenneth D. Hopkins
University of Colorado
Boulder, Colorado

December 1971

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Table of Contents

Tables and Figures

Acknowledgments

The author is indebted to several persons who made valuable contributions to this project. Glenn H. Bracht and Larry Nelson were responsible for most of the computer analyses. Beverly Anderson assisted in the review of related literature and in the data collection. Nancy Gallagher edited and typed the final report. George Kretke and Socrates Rapagna assisted with various statistical computations. Gene V Glass made helpful suggestions regarding the factor analytic portions of the study. Eugene Gullette and the Boulder Valley Public Schools cooperated by providing access to testing information. All of these individuals deserve credit for their contribution to the project.

SUMMARY

The stability and change of Verbal, Non-verbal, and Total IQ scores from group tests was investigated for students tested at grades 1, 2, 4, 7, 9, and 11. The study provided empirical support for the following conclusions:

1. Contrary to most textbook treatments of the topic and prevailing professional impression, there is little or no information on the stability of IQ scores from group tests especially (1) at the primary grades, or (2) for time intervals greater than three years, or (3) on currently used tests, or (4) on Non-verbal tests.
2. Verbal IQs are much more stable than Non-verbal IQ scores, especially prior to junior high school.
3. At grades one and two Verbal and Non-verbal IQs reflect relatively little lasting variance, even over short intervals.
4. By grade four, Verbal IQs show a very high relationship (about .8) with corresponding IQs thereafter.
5. Non-verbal IQs at grade seven reflect about the same degree of stability as Verbal IQs at grade four.
6. Total IQs have a pattern of stability highly similar to that of Verbal IQ scores.
7. IQ scores from group tests to be considerably less stable than Stanford-Binet IQs prior to grade seven, atter which there is little differences in stability.
8. In the main, there is considerable change in true Verbal IQ scores prior to grade four, but little change thereafter.
9. In the main, there is considerable change in true Non-verbal IQ scores in elementary school , but only minor changes thereafter.
10. Differences in Verbal and Non-verbal IQs show meager stability over time prior to grade seven, after which there is considerable stability.
11. The use of IQ scores in addition to the most recent score offers little promise for improving the determination of eventual intellectual status. Multiple regression can contribute slight improvement in prediction, whereas the averaging of prior performances lessened the accuracy in prediction.
12. The factor analytic study confirmed informal observations made from a study of the general correlation matrix -- Verbal IQ variancebecomes stable at about grade four, Non-verbal IQ variance stabilizes at about grade seven.

THE STABILITY AND CHANGE OF LANGUAGE AND NON-LANGUAGE IQ SCORES

Background for the Study.

Ever since the concept of an intelligence quotient, or IQ, was first applied to the measurement of intelligence by Terman in 1916 (Horrocks, 1964), the "constancy" of the IQ has been an important concern of educators and psychologists.

The question all too frequently is discussed as a general topic, as if the findings on a given test generalized to all tests and all groups. For example, Bloom (1963, p. 383) has stated,

> ...General intelligence as measured by either an individual intelligence test or a group test becomes relatively constant or stable by about the second or third year of school...after age eight the correlations between measurements over periods of six to ten years rarely are less than .80 when the same or parallel forms of the same test are used in repeated measurements.

Anastasi (1960, p. 232) also comments, "Empirically, IQ's have been found to remain sufficiently constant during the elementary schools years to make predictions over several years feasible." Most studies on the topic, however, have been made using individual tests, usually the Stanford-Binet; relatively few have used group tests, except for college and adult subjects. No previous study employing group tests, the only type ever taken by ninty percent or more of the population, has spanned the ten year interval mentioned in the Bloom quote above; in fact, only one previous study had encompassed a time interval of even six years and that was done years ago on a small, atypical sample using an obsolete test.

The literature search failed to reveal any study in which the IQ constancy on group language and/or non-language tests was investigated for either normal or special reference groups that continued from elementary through high school. Since inferences with long-range educational and vocational implications are important functions of intelligence tests, it is apparent that educational and psychological personnel must be aware of the degree of accuracy of such inferences made from these tests.

Non-language intelligence tests have been devised, in part, to provide an intellectual assessment of pupils with atypical cultural and/or lingual backgrounds. It is usually assumed that non-language assessment is more accurate for these types of pupils. The degree of accuracy of future predictions from these instruments must be known before their value can be ascertained.

The extent to which IQ scores reflect innate potential or abilities developed through various environmental stimuli is not addressed in, nor relevant to, the present study. Almost all pupils take several intelligence tests during the K-12 school years, and inferences are made from these scores. The critical question addressed by the present study is how much do obtained IQ scores change over time? The accompanying long-term inferences can be no more accurate than the degree of stability reflected in the IQ scores.

Related Literature.

The literature on IQ constancy was reviewed by Thorndike (1933, 1940) and more recently by Bloom (1964) who presented a summary of many previous studies. These reviews have been devoted almost exclusively to research involving the Stanford-Binet or adult subjects.

The Stanford-Binet, a verbal intelligence test that requires individual administration by a highly trained psychologist or psychometrist, and hence is very expensive, is the only test on which the stability of the IQ has been extensively investigated. It is paradoxical that long-term stability is usually treated as a general topic, apart from any particular test, as if Stanford-Binet data generalized to all tests and groups.

An exhaustive literature search identified twenty-one research studies that contained at least fragmentary data on the stability of mental ability performance for some portion of time during the developmental (pre-adult) period. Each of these studies is briefly described in chronological order.

Studies of IQ Stability from Group Tests with School-Age Children.

Garrison and Robinson (1925) found a very high correlation (.91 and .93) with 131 children in grades 3-6 in the Peabody Demonstration School between scores on the National Intelligence Tests and retests two (or three) years later. Descriptive data regarding variability was absent, although an average IQ of 133 was reported. It is unclear whether raw scores or IQ score were the basis for the correlation -- a critical factor since the age-range of the sample was considerable.

Hirsh (1930) (also Miller, 1933, using the same data) correlated grade-one Otis IQ scores obtained by 160 students with corresponding scores 1-5 years later. The correlation coefficients between grade one scores and grade 2-6 performance reflected an atypical pattern:

.79, .77, .76, .82, .88, respectively. It is unclear why the standard deviation of grade six scores was so large -- about 20 points.

Lincoln and Wadleigh (1932) studied 154 "regular pupils" tested in grades 3-6 with Otis Primary and retested 1-3 years later using either the National or Terman intelligence tests. No descriptive data on the sample such as means or variability were given, nor was a correlation coefficient reported -- only that the average difference between scores was about 7 IQ points.

Lauderbach and Hanse (1932) reported a correlation of .79 for a heterogeneous group of 150 pupils in grades 4-6 who were tested one year later on the now-obsolete McCall Multi-mental Scale.

Traxler (1934) obtained correlations of .76, .67, and .68 for a non-representative group of ninth grade students at the University of Chicago High School retest one, two, and three years later on the Otis Self-Administering Test.

Seagoe (1934) studied pupils who were tested in grade one (Detroit First-Grade), three (Detroit Primary), five (National Intelligence) and seven (Terman Group). Unexpectedly, the correlations between scores in grades one and seven (.77) was higher than for scores in grades one and three (.64). Other correlations ranged from .70 (3 with 5) to .87 (5 with 7).

Pinter and Stanton (1937) studied 59 students who took Thorndike's CAVD tests in grades three, four, and five. No data on means or variability was given, hence the representativeness of the sample is uncertain, but the stability coefficient reported were r = .68 (3 with 4), r = .76 (3 with 5), and .82 (4 with 5).

Goldin and Rothchild (1942) investigated IQ stability on 174 pupils tested in grades one, and four on the Pitner-Cunningham and the Henmon-Nelson. The correlation was only .46. Additional Henmon-Nelson test scores were available on 54-60 of the same students in grades six and eight. The correlation coefficients between IQ scores in grades four and six and six and eight were in the .81-.85 range. Other correlations were not reported.

Knezevich (1946) found a correlation of .70 between Henmon-Nelson IQ scores on 113 sophomores in a rural high school with corresponding scores two years later. The group was atypically homogeneous (standard deviation was only 8.1 points), hence the stability coefficient is a conservative estimate.

Doppelt and Bennett (1951) studies the grade 9-12 stability of the various tests of the Differential Aptitude Tests for 140 boys and girls. Stability coefficients ranged .59-.78 for all tests except Verbal Reasoning, which had higher coefficients in the .84-.86 range. No descriptive data were given for the sample.

Layton (1954) studied 2,169 students who were tested in grades 9 and 12 with the American Council on Education (ACE) test, and reported a correlation of.80. This test is usually considered primarily an achievement, rather than an aptitude test, and bears only indirectly on IQ stability.

Collman and Newlyn (1958) studied 182 English pupils, ages 6-15, in "schools for the subnormal" who took the Terman-Merrill Intelligence Test twice. The testing interval was one year. A correlation of .93 was reported, unfortunately it was not made clear whether IQ scores or Mental Ages were correlated -- a serious defect in view of the age range in the sample. An average IQ of 68 was reported for the group, with a 42-89 range.

Tyler (1958) studied 83 pupils who were tested with Primary Mental Abilities Tests (PMA) in grades one and four. Low stability coefficients were .48, .42, .51, .53 for the Verbal, Spatial, Numerical and Perceptual Tests respectively, but a correlational .70 was reported for the total score. She also studied 127 pupils tested with the PMA in grades 4 and 8. Correlations of .65, .82, .41, .62, and .66 were reported for the Verbal, Numerical, Spatial, Reasoning, and Total scores respectively. Note the low stability of the Spatial test. No descriptive data on the means on variance were reported.

Meyer and Bendig (1961) studied 100 eighth grade students who took the Primary Mental Abilities Tests (PMA) in grades 8 and 12. Stability coefficients of .81, .75, .73, .66 and .43 were reported for the Verbal, Reasoning, Numerical, Spatial and Word Fluency Tests respectively. The total scores had a stability coefficient of .82. Note the very low stability of the Word Fluency test. Notice also that these values are higher than those found by Tyler with the same time interval but at a younger age.

Hopkins and Bibelheimer (1963) studied 404 students with representative means and variability who took the California Test of Mental Maturity (CTMM) (S-Form) in grades 3, 5, and 7. A parallel study of 314 pupils who took the CTMM was reported subsequently by Finley, Thompson, and Cognata (1966). Both studies found low stability coefficients (.50 or below) between Language and Non-language IQ scores at grades three with corresponding scores in grades five and seven. The grade-five-to-seven Language IQ stability was higher (r's of .71-.78) than Non-language IQs (r's of .47-.52).

In a subsequent study Hopkins and Bibelheimer (1971) obtained CTMM IQ data on 354 pupils from the same sample and found that grade eight IQs had about the same stability coefficients as the grade seven reported above, and that Language IQs tended to be more stable over time than Non-language IQs.

Bradshaw (1964) investigated IQ stability on 58 public school children tested in grades 2 and 4 with the California Test of Mental Maturity (CTMM) and found coefficients of .65, .50, and .66 for Language, Non-language, and Total IQs respectively. No data on group variability were reported.

Tulley and Hall (1965) investigated the one-year stability of the School and College Ability Test over the grade 9-10 period. Very high stability coefficients of .92, .88, and .94 were reported for the Verbal, Quantitative, and Total Scores respectively. This test does not yield IQ scores and is usually considered to be more heavily loaded with academic achievement content than are most "intelligence" tests and may have greater stability accordingly, although little is known about achievement stability.

Eagle (1966) studied 115 boys and 150 girls who were tested with the Lorge-Thorndike Intelligence Test in grades 3 or 4 and retested in grades 7 or 8. He reported correlations of .78 and .67 for Verbal and Non-verbal IQs for girls, and corresponding lower values of .59 and .49 for boys. Unfortunately, no data were reported on group variability.

Lorge, Thorndike, and Hager (1966) reported several stability coefficients for certain groups of pupils for one to three-year intervals on the Lorge-Thorndike Intelligence Test. The stability coefficients were larger for the higher grades. For example, grades 3-6 correlations were about .5 for both Verbal and Non-verbal IQs, but the corresponding values were .82 and .74 for the grade 6-9 correlations.

Summary of the Literature Review.

Several general conclusions can be drawn from the literature survey. Seven of the 21 studies were conducted and reported prior to 1938 and pertain, for the most part, to obsolete tests. Five other utilized tests, although cognitive in nature, are not intelligence tests in the conventional sense. Six studies employed currently used intelligence tests, but the largest time interval investigated was five years.

Several of the studies failed to report data on the sample's mean and variability. The latter being especially critical since it has a substantial affect onthe magnitude of the stability coefficient -- other things being equal, the more heterogeneous the sample, the larger the stability coefficient. Even more serious is the fact that in some of the studies, it is not clear whether IQ scores, raw score or Mental Ages were correlated. The latter will have a much larger stability coefficient when the sample is not homogeneous in age. This factor probably accounts for the abnormally high coefficients reported by Garrison and Robinson (1925), Hirsh (1936), Collmann and Newlyn (1958).

Another confounding factor in analyzing previous research results from the fact that all IQ stability studies on group tests prior to 1966 investigated the stability of ratio IQ's. Almost all currently used tests use standard score deviation IQs. With ratio IQs one rarely can determine whether a sample's variability is representative or not. For example, prior to 1966 when the Stanford-Binet used ratio IQ's, the standard deviation of the IQ scores was 12 points at age 6 but 20 points at age 12 (on Form L). This type of normative data is not available on the group tests, hence one cannot make adjustments for atypical variability. For example, Knezevich (1946) reported a standard deviation of only 8.1 points for the twelve-grade Henmon-Nelson IQ scores, in his sample, yet Hirsh's (1930) sample had standard deviation of 19.8 points on Otis IQ scores at grade six. It is unfortunate that only six or so studies were adequately conducted and reported to be of value as background data against which those from the present study can be related.

METHODS

The entire student population for two graduating classes ('69 and '72) from a school district of approximately 30 schools and 20,000 enrollment in Boulder County, Colorado were employed in this study. This district's testing program has prescribed group intelligence tests be administered to all pupils in grades 1, 2, 4, 7, 9, and 11 for more than a decade. All testing was conducted in September or October -- the mean age of the initial (grade one) testing was 78.2 months.

Initially the results for the two graduating classes ('69 and '72) were analyzed separately and gave parallel findings, but both groups were subsequently combined to give greater stability and generalizability to the findings.

Table 1 gives specific descriptive data for the two classes and the tests employed. Scores for any students who was present for at least two of the testing periods were retained in the analysis. The means and standard deviations of the Verbal (V), Non-verbal (NV), and Total (T) IQ scores of all pupils retained in the sample are given in Table 2. Although it is evident the sample is above the national average in central tendency, this difference would not be expected to seriously affect the stability coefficients since the sample's variability is comparable (slightly less) than the national parameters.

The standard deviations in Table 2 are for all students in the sample at each grade level. The means and standard deviation were also computed separately for the 186 students present at each of the six testings. Table 3 gives these data. Notice that in most cases the corresponding standard deviations differby less than 1 IQ point, hence the resulting stability coefficients would not be expected to be seriously affected by differential variability resulting from student attrition and mobility.

Table 1

Year and Month of Testing, Mean Age in Months, and Test Administered at Each Grade Level in Classes of '69 and '72

Grade	Class	Testing (Mo./Year)	Mean Age (Mo.)	Test Used
1[b]	'69	10/57	78	CTMM[a], 1950 S-Form, Primary Level
2	'69	10/58	92	CTMM, 1957 S-Form, Primary Level
	'72	9/60	91	CTMM, 1957 S-Form, Primary Level
4	'69	10/60	116	CTTM, 1957 S-Form, Elementary Level
	'72	9/62	114	L-T[c], Form A, Level 3
7	'69	9/63	150	L-T, Form A, Level 4
	'72	10/65	151	L-T, Form A, Level 4
9	'69	9/65	174	L-T, Form A, Level 5
	'72	10/67	175	L-T, Form A, Level 5
11	'69	10/67	199	L-T, Form B, Level 5
	'72	10/69	199	L-T, Form B, Level 5

[a]California Test of Mental Maturity.

[b]First-grade IQ scores were not available for Class of '72.

[c]Lorge-Thorndike Intelligence Test.

Table 2

Means and Standard Deviations for the Verbal, Non-Verbal and Total IQ Scores for the Sample at Various Grade Levels

Grade	Test	N		Verbal IQ	Non-Verbal IQ	Total IQ
1	CTMM[a]	446	$\bar{X}$	114.24	105.68	109.87
			s	12.56	12.09	10.45[b]
2	CTMM	995	$\bar{X}$	113.77	110.95	112.40
			s	17.98	17.83	15.20
4	CTMM and L-T	1629	$\bar{X}$	111.88	111.71	111.84
			s	16.08	16.87	14.74
7	L-T	2284	$\bar{X}$	112.64	113.42	113.03
			s	14.14	14.88	13.28
9	L-T	2166	$\bar{X}$	114.34	118.99	116.66
			s	13.43	14.49	12.73
11	L-T	1859	$\bar{X}$	113.71	119.91	116.82
			s	13.50	14.20	12.47

[a]Since the CTMM used a ratio IQ, the value of the standard deviation is not fixed or constant across grade levels.

[b]The variance of the average of two variables, each with equal variances, will always be less than the variance of either variable except when $\rho = 1.0$, in which case they will be equal (see Hopkins and Bibelheimer, 1971 for proof).

Table 3

Means and Standard Deviations of Verbal, Non-Verbal, and Total IQ Scores for Students with a Complete Set of Scores across Grades 1-11 (N = 186)

Grade	Test	Verbal IQ	Non-Verbal IQ	Total IQ
1	CTMM $\bar{X}$	115.61	106.58	111.03
	s	11.14	11.87	9.44
2	CTMM $\bar{X}$	117.45	116.04	116.88
	s	15.76	17.32	13.20
4	CTMM $\bar{X}$	120.92	120.98	121.06
	s	16.27	17.74	14.41
7	L-T $\bar{X}$	118.32	117.48	117.95
	s	12.77	14.25	12.20
9	L-T $\bar{X}$	118.97	123.74	121.37
	s	12.59	11.83	10.74
11	L-T $\bar{X}$	117.72	122.98	120.30
	s	13.86	13.70	12.31

FINDINGS

Stability of Verbal IQs

The stability and generalizability (over time and tests) coefficients for Verbal IQs on the total sample are given in Table 4; the corresponding N's on which the r's are based are given below the diagonal. For example, the correlation of .51 between Verbal IQs in grades one and two was based on 334 pupils. Notice that the stability coefficients are quite low at grades one and two, even for a one or two-year interval. Conversely, the stability of the grade four verbal IQ scores are quite high for all subsequent tests thereafter. Since the sizes of the samples on which the correlation coefficients are based are large, the resulting standard errors of the r's are small, ranging from less than .01 (for N = 1813) to .05 (for N = 236).

The stability coefficients in Table 4 are reduced slightly in magnitude by the fact that the sample is somewhat less variable than a representative national sample (i.e., the standard deviations are less than 16). To estimate the coefficients for a normative sample, the coefficients were corrected for terminal variability (σ = 16) at grades 7, 9, and 11 (as done by Bloom, 1964, p. 56). Table 5 gives these corrected stability coefficients for the Verbal IQ scores. Notice that the values are .03 to .09 higher than corresponding values in Table 4. Note also that the sequence of coefficients over time becomes somewhat more orderly when the coefficients are viewed in relation to a fixed degree of terminal variability. The grade 11 column depicts the most relevant data since these correlations indicate the relationship with a more-or-less final adult level of intellectual performance. Note that the grade four Verbal IQ scores correlated very highly with all tests thereafter.

Table 4

Stability and Generalizability Coefficients of Verbal IQ Scores for the Total Sample. Corresponding N's are given below diagonal.

Verbal Grade	1	2	4	7	9	11
1		.51	.52	.50	.44	.50[a]
2	334		.46	.53	.51	.54
4	263	705		.79	.75	.73
7	260	620	1194		.86[b]	.80
9	253	586	1071	1813		.85
11	236	517	927	1527	1709	

[a]Value with largest standard error; .95 confidence interval is .40 - .59.

[b]Value with smallest standard error; .95 confidence interval is .847 - .873.

Table 5

Stability and Generalizability Coefficients of Verbal IQs Corrected[a] for Terminal Variability

Grade Level	7	9	11
1	.59	.53	.55
2	.61	.60	.62
4	.84	.82	.79
7		.90	.84
9			.88

[a]Corrected to a representative terminal variability (σ = 16), for grades 7, 9, and 11. Correction cannot be made on earlier tests since ratio IQ was involved and no accurate estimates of the population standard deviations available for the various grade levels.

The degree of stable variances in IQ scores is affected by both changes in individuals' true scores and errors of measurement. In order to remove the influence of measurement errors on the stability coefficients, hence, estimate the relationship between true Verbal IQ scores over various time intervals, the stability coefficients in Table 4 were corrected for attenuation (errors of measurement). However, since reliability estimates were not available in the technical manuals for certain of the tests, it was assumed that the geometric mean of the reliability coefficients associated with each coefficient was .90. This value was selected as it would provide realistic, yet somewhat conservative dissattenuated coefficients since .90 was above the median reliability coefficient reported in the tests' technical manuals. These disattenuated correlation coefficients appear in Table 6.

It is evident from Table 6 that less than half the variance in true Verbal IQ scores at grades 1 and 2 on group intelligence tests is lasting and general. It is important that any interpretation of IQ scores at grade 1 and 2 be very tentative and reflect the fact that substantial changes in true IQ will occur in many cases. By grade four however, the individual differences in true IQ scores have stabilized greatly and have about 80 percent common variance with subsequent IQ scores thereafter.

Stability of Non-Verbal IQs

The stability coefficients for Non-verbal IQs for the total sample are given in Table 7. (The corresponding N's on which the r's are based were given in Table 4.) Notice that the coefficients tend to be less than corresponding values for Verbal IQs shown in Table 4.

Table 6

Stability and Generalizability Coefficients for Verbal IQ Scores Corrected for Terminal Variability and Attenuation[a]

Grade Level	7	9	11
1	.65	.59	.61
2	.68	.67	.69
4	.93	.91	.88
7		1.00	.94
9			.98

[a]Assuming geometric mean of reliability coefficients is .90.

Table 7

Stability Coefficient of Non-Verbal IQ Scores for Total Sample

Grade	2	4	7	9	11
1	.36	.29	.36	.26	.29[a]
2		.47	.49	.45	.47
4			.61	.59	.56
7				.78[b]	.73
9					.76

[a]Value with the largest standard error; .95 confidence interval is .17 - .40

[b]Value with the smallest standard error; .95 confidence interval is .761 - .797

Olkin's (1967, p. 113-115) procedure to test for significance of the difference between the non-independent stability coefficients was used to test for significance of difference between corresponding stability coefficients for Verbal and Non-verbal IQs. The z-ratios were all significant ($p < .05$) except for coefficient involving grade two. All comparisons involving grade four or above were significant beyond the .001 level. The irregularity at grade two resulted from the fact that, whereas Non-verbal IQs at grade two were considerably more stable than those at grade one, grade two Verbal IQs were not more stable than first-grade scores.

The data support the general conclusions that observed Non-verbal IQ scores show even greater change over time than do Verbal IQ scores on group tests. This is particularly surprising in view of the fact that a logical study of the items suggests that they should be far less influenced by environmental factors.

The Non-verbal stability coefficients in Table 7 were corrected for their slightly-less-than-typical variability. Estimates of corresponding representative stability coefficients ($\sigma = 16$) are given in Table 8 (analogous to Table 5 for Verbal IQs). Corresponding disattenuated coefficients (analogous to Table 6 for Verbal IQs) are shown in Table 9.

Table 8 indicates that the greater stability coefficients of the Verbal IQ scores is not due to differential heterogeneity. Table 9 suggests that there is less lasting general variance in the true Non-verbal IQ scores than in the Verbal IQ scores. Not until grade 7 was the common variance in true scores greater than fifty percent. In other words, even greater caution is required with Non-verbal IQs than for Verbal IQs in any long term guidance or other uses. Both the obtained IQ scores and true IQ scores show greater change over time than for Verbal tests. A study of other previous research indicated that this finding is generally consistent across other tests (Hopkins, 1963; Finley et al., 1966; Lorge et al., 1966; Tyler, 1958; Meyer and Bendig, 1961; Eagle, 1966; Bradshaw, 1964; Hopkins, 1971).

The Stability of Total IQ Scores.

Since Total IQ scores on the CTMM and L-T result from averaging the Verbal and Non-verbal IQ scores, they tend to have slightly less associated measurement error. Consequently, their values, as shown in Table 10, tend to be slightly, but not meaningfully, higher than that found for Verbal IQs. The stability and generalizability coefficients for Total IQ scores, and the corresponding values corrected for terminal variability are given in Table 11. Table 12 presents the associated disattenuated values.

Table 8

Stability and Generalizability Coefficients[a] for Non-verbal IQ Scores[b] Corrected for Terminal Variability

Grade Level	7	9	11
1	.40	.34	.33
2	.55	.56	.53
4	.66	.67	.61
7		.82	.77
9			.79

[a]See Table 4 for associated sample sizes.

[b]σ = 16; (see footnote a, Table 5).

Table 9

Stability and Generalizability Coefficients for Non-verbal IQ Scores Corrected for Terminal Variability and Attenuation[a]

Grade Level	7	9	11
1	.44	.38	.37
2	.61	.62	.59
4	.74	.74	.68
7		.91	.85
9			.88

[a]Assuming geometric mean of reliability coefficients is .90.

Table 10

Stability and Generalizability for Total IQ Scores

Grade Level	2	4	7	9	11
1	.57	.51	.50	.45	.50
2		.63	.62	.62	.63
4			.77	.74	.73
7				.88	.83
9					.86

Table 11

Stability and Generalizability Coefficients for Total IQs Corrected to a Common Variability[a]

Grade Level	7	9	11
1	.57	.55	.56
2	.69	.72	.70
4	.82	.81	.78
7		.91	.86
9			.89

[a]Corrected to representative variability according to normative data

Table 12

Stability and Generalizability Coefficients for Total IQ scores Corrected for Terminal Variability and Attenuation[a]

Grade Level	7	9	11
1	.63	.61	.62
2	.77	.80	.77
4	.91	.90	.87
7		1.00	.96
9			.99

[a]Assuming geometric mean of reliability coefficients is .90.

Graphic Representation of Stability Coefficients

The findings pertaining to the stability and change of Verbal, Non-verbal, and Total IQs are presented in Figure 1, which graphically depicts the correlation of IQs at each grade level with terminal (Grade 11) IQ scores (corrected to a common terminal variability, $\sigma = 16$). Corresponding data from Bloom (1964, p. 56) on an individual verbal test (Stanford-Binet) are also reproduced in Figure 1 to illustrate the contrast in IQ stability between individual vs. group tests.

General Conclusions regarding IQ Stability.

Several general conclusions are evident from Figure 1 and associated Tables.

1. Verbal IQs are much more stable than Non-verbal IQ scores, especially prior to junior high school.

2. At grades one and two Verbal and Non-verbal IQs reflect relatively little lasting variance, even over short intervals.

3. By grade four, Verbal IQs show very high relationship (about .8) with corresponding IQs thereafter.

4. Non-verbal IQs at grade seven reflect about the same degree of stability as Verbal IQs at grade four.

5. Total IQs have a pattern of stability highly similar to that of Verbal IQ scores.

6. IQ scores from group tests to be considerably less stable than Stanford-Binet IQs prior to grade seven, after which there is little differences in stability.

7. In the main, there is considerable change in true Verbal IQ scores prior to grade four, but little change thereafter.

8. In the main, there is considerable change in true Non-verbal IQ scores in elementary schools, but only minor changes thereafter.

Changes in IQ Scores.

Changes in IQ scores are reported in a different form in Tables 13-15 where the means and standard deviations for both algebraic and absolute differences between IQs are given for each pair of grade levels. The average amount of change per student, without regard for the direction of the change, is given by the mean of the absolute differences. Since the distribution of absolute differences is positively skewed, the normal curve model should not be applied to estimate the proportion of students whose IQ changed by some specified amount. (Tchebycheff's inequality (Hays, 1963, p. 188) can

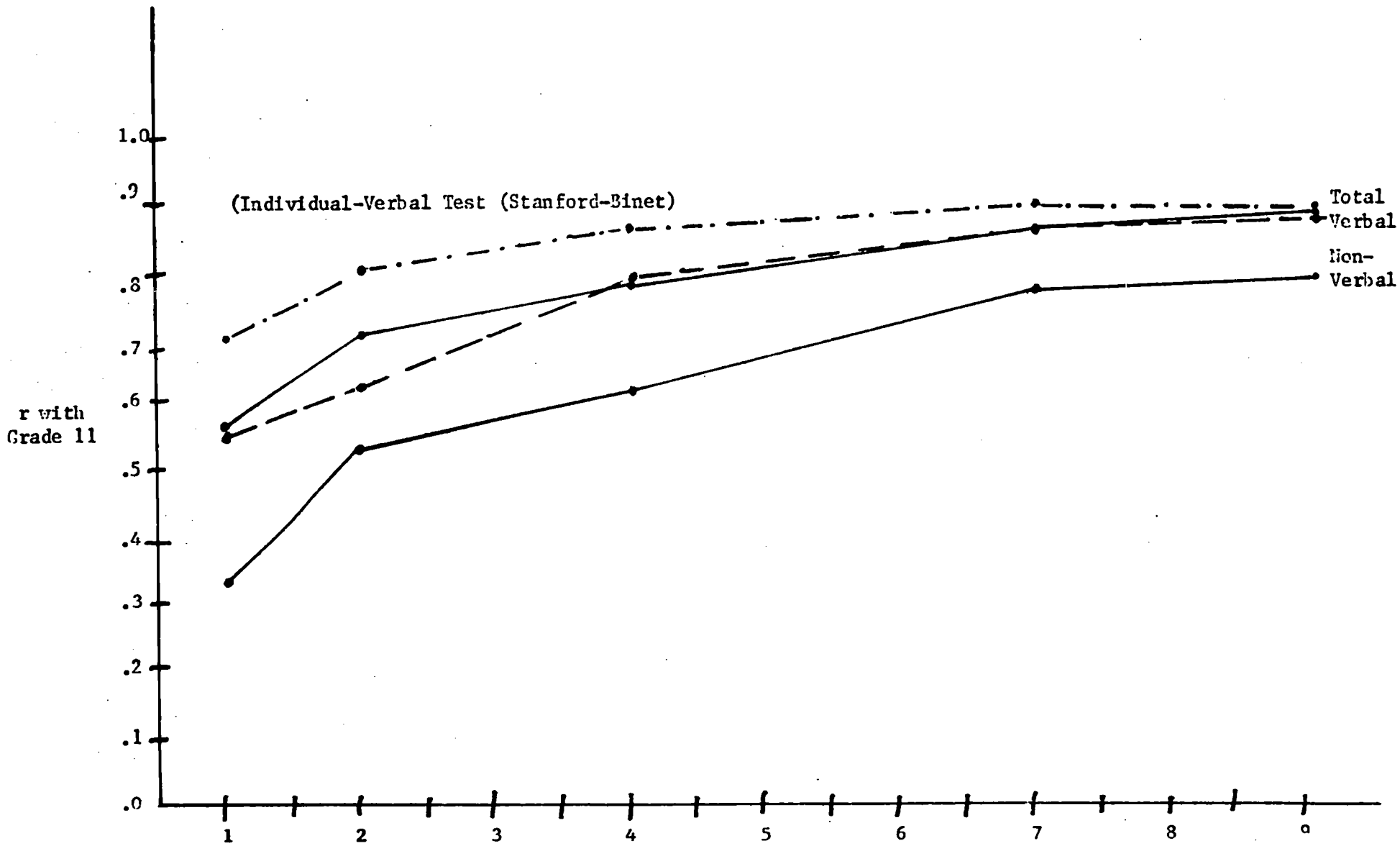

Figure 1. Graphic Representation of the Stability Coefficients for Verbal, Non-verbal, and Total IQs at Certain Grade Levels with IQ at Grade 11 (corrected to a common variability, $\sigma = 16$). Corresponding data from Individual Test also shown for comparison.

be used to obtain crude, conservative estimates of such proportions. For example, the proportion having an IQ change which is greater than two standard deviations of absolute change from the mean absolute change is less than .25.)

The absolute differences between IQs are influenced by both the instability of scores and the difference between test means at the two grade levels. To remove the effect of differences between test means, which may reflect norming artifacts, the distribution of algebraic differences between IQs was obtained. The standard deviation of algebraic differences indicates the degree of IQ change independent of norming imperfections and/or changes in the group's intellectual elevation relative to the norming population.

The mean and standard deviation of the distribution of algebraic differences allow an estimate of the proportion of students whose IQ changed by any specific amount. Previous research (Hopkins and Bibelheimer, 1971) has indicated that the distributions of algebraic differences in IQ scores are normally distributed; hence the normal curve model is applicable. For example, Table 13 reveals that from grade one to grade two, about one-sixth of the students had Verbal IQ scores that decreased by 15 or more points, and about one-sixth had scores that increased by 15 or more points. The mean change, irrespective of direction (absolute difference) was about 12 (11.79) points. As another example, Table 13 shows that between grades 7 and 11, about two-thirds of the students have Verbal IQ scores that differ by 9 (8.42) points or less with the mean change being about 7 points (6.75). The Total IQ show less IQ change than either Verbal or Non-verbal IQs largely because σ is less than 16. (See Table 2, especially footnote b.) The mean Total IQ change between grades 9 and 11 was only 5 points and two-thirds of the sample had IQs that changed less than 7 points during that period (see Table 15).

Comparative IQ Stability for Boys and Girls.

Relatively little attention has been devoted to the comparative IQ stability as related to sex. Eagle (1966) reported much less four-year stability for boys than girls on both Lorge-Thorndike Verbal and Non-verbal Tests. Hopkins and Bibelheimer (1963) and Finley, et al. (1966) found no significant sex differences in IQ stability on the California Tests of Mental Maturity. Although other investigations have not explored the question statistically a perusal of the findings of Meyer and Bendig (1960, 1961), Doppelt and Bennett (1951) and Layton (1958) reveals no consistent pattern of differential stability by sex on other cognitive tests.

Table 16 presents means, and standard deviations for boys and girls in the total sample.

Notice that boys and girls are highly similar, both in means and variability. Corresponding stability coefficients for Verbal, Non-verbal, and Total IQs for each sex are given in Table 17.

Table 13

Means and Standard Deviations Algebraic and Absolute Differences Between Verbal IQs for Each Pair of Grade Levels

Grade		Algebraic Differences 2	4	7	9	11
1	$\bar{X}$	- .04	-4.94	-2.14	-3.07	-2.23
	s	15.08	14.44	12.39	12.72	12.75
2	$\bar{X}$		.52	.52	- .72	.66
	s		17.18	15.23	15.49	14.89
4	$\bar{X}$			- .63	-2.09	- .42
	s			9.96	10.53	10.78
7	$\bar{X}$				-1.16	.53
	s				7.28	8.62
9	$\bar{X}$					1.75
	s					7.33

Grade		Absolute Differences 2	4	7	9	11
1	$\bar{X}$	11.79	12.21	9.98	10.33	10.26
	s	9.38	9.14	7.62	8.00	7.87
2	$\bar{X}$		13.67	12.03	12.39	11.93
	s		10.40	9.34	9.32	8.91
4	$\bar{X}$			7.83	8.43	8.54
	s			6.18	6.64	6.59
7	$\bar{X}$				5.76	6.75
	s				4.60	5.38
9	$\bar{X}$					5.89
	s					4.70

Table 14

Means and Standard Deviations of Algebraic and Absolute Differences Between Non-Verbal IQs for Each Pair of Grade Levels

Grade		Algebraic Differences 2	4	7	9	11
1	$\bar{X}$	-8.42	-13.64	-10.04	-16.73	-16.73
	s	17.73	18.74	15.60	14.77	15.25
2	$\bar{X}$		-3.28	-4.29	-9.84	-8.90
	s		17.30	16.00	16.09	15.91
4	$\bar{X}$			-1.47	-6.65	-5.81
	s			14.23	14.16	14.27
7	$\bar{X}$				-5.24	-4.07
	s				9.66	10.47
9	$\bar{X}$					.16
	s					9.73

Grade		Absolute Differences 2	4	7	9	11
1	$\bar{X}$	15.61	18.81	15.01	18.80	18.50
	s	11.88	13.52	10.89	12.01	13.03
2	$\bar{X}$		13.81	13.25	15.26	14.74
	s		10.91	9.93	11.08	10.64
4	$\bar{X}$			10.98	12.38	12.09
	s			9.16	9.56	9.55
7	$\bar{X}$				8.62	9.02
	s				6.81	7.36
9	$\bar{X}$					7.66
	s					5.99

Table 15

Means and Standard Deviations of Algebraic and Absolute Differences Between Total IQs for Each Pair of Grade Levels

Grade		Algebraic Differences 2	4	7	9	11
1	$\bar{X}$	-4.36	-9.46	-6.23	-10.08	-9.52
	s	12.10	12.71	11.29	10.80	11.13
2	$\bar{X}$		-1.38	-1.84	-5.25	-4.08
	s		12.35	11.73	11.56	11.33
4	$\bar{X}$			-1.00	-4.31	-3.09
	s			9.50	9.85	9.77
7	$\bar{X}$				-3.21	-2.29
	s				6.33	7.28
9	$\bar{X}$					.95
	s					6.50

Grade		Absolute Differences 2	4	7	9	11
1	$\bar{X}$	10.24	12.60	10.43	12.24	11.54
	s	7.76	9.61	7.57	8.27	9.02
2	$\bar{X}$		9.60	9.50	10.24	9.64
	s		7.89	7.12	7.51	7.20
4	$\bar{X}$			7.42	8.54	8.10
	s			6.01	6.53	6.27
7	$\bar{X}$				5.56	5.84
	s				4.41	4.91
9	$\bar{X}$					5.15
	s					4.08

Table 16

Means and Standard Deviations for Each Sex at Each Grade Level[a]

	Verbal				Non-Verbal			
	Boys		Girls		Boys		Girls	
Grade	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
1	114.6	10.7	116.6	11.5	105.6	12.1	107.6	11.7
2	117.8	16.4	116.2	17.2	113.0	16.9	113.4	16.5
4	112.5	16.1	114.9	15.2	113.7	17.5	114.8	15.3
7	112.9	13.7	115.7	13.2	115.0	15.1	115.2	13.4
9	114.8	13.0	115.9	13.0	120.5	14.2	119.8	13.4
11	113.9	13.5	113.3	13.7	121.3	14.8	118.6	13.5

[a]Sample sizes are given in Table 17

Table 17

Stability Coefficients for Verbal, Non-Verbal and Total IQ Scores by Sex

Grade		N	2 V	2 NV	2 T	4 V	4 NV	4 T	7 V	7 NV	7 T	9 V	9 NV	9 T	11 V	11 NV	11 T
1	B	93	.51	.32	.50	.39	.21	.41	.32*	.20	.34*	.23*	.19	.29**	.36	.27	.36**
	G	93	.38	.35	.50	.52	.36	.56	.54*	.37	.55*	.51*	.33	.54**	.55	.38	.59**
2	B	225				.43	.43	.58	.47	.46	.58	.44	.43	.56	.49	.45	.59
	G	215				.50	.42	.62	.59	.54	.66	.53	.45	.62	.55	.51	.64
4	B	432							.78	.56	.75	.76	.55	.73	.75	.53**	.72
	G	404							.76	.63	.79	.74	.62	.75	.72	.62**	.76
7	B	751										.85	.76	.87	.81	.71**	.83
	G	705										.86	.77	.88	.80	.76**	.85
9	B	888													.84**	.76	.85**
	G	821													.87**	.76	.88**

* p < .10

** p < .05

Only 9 of the 45 comparisons in r's differed by enough to reject the null hypothesis at the .10 level (whereas about 5 would have been expected to by chance). In every instance where the difference was significant, greater stability was reflected by the girls.

Using a binomial statistical model which is not definitive due to the lack of independence in the data, the girls showed greater stability coefficients on 9 of the 15 (60%) Verbal IQ comparisons, 13 of 14 (93%) comparisons in Non-verbal comparison ($p < .01$), and 14 of 14 comparisons in Total IQ score ($p < .01$). It seems clear that there is a trend for slightly greater stability in the IQ scores of girls. The differences is so small, however, that it is of negligible practical significance. The differences were much less than those reported by Eagle (1966), even though the same tests (L-T) and grade levels (4-7) were employed.

Reliability of Verbal vs. Non-verbal IQ Differences.

Interpretations of performance on Verbal and Non-verbal intelligence tests often consider the difference in IQ scores, as well as the specific Verbal and Non-verbal IQ scores. Clinical and educational significance is often attached to these differences (see Hopkins, 1964). If, these differences are unreliable over time, however, no lasting significance can validly be attached to a Verbal-Non-verbal IQ difference. Table 18 gives the correlations of Verbal IQ - Non-verbal IQ differences across the various time intervals. Notice that the reliability of the difference in scores is extremely low for grade two differences, quite low for grade four differences, and becomes moderately stable at grade seven. Hence no lasting significance should be attached to differences in Verbal and Non-verbal IQs below grade seven, unless, of course, the difference is very great (e.g., 20 or more points at grade four).

The difference in Verbal and Non-verbal IQs at junior high school does show considerable stability (.50 or higher) and could be of diagnostic value in assessing intra-individual strengths and weaknesses. One should also bear in mind the fact that if the coefficients in Table 18 were dissattenuated to indicate the stability of differences in true Verbal and true Non-verbal IQs, they would increase substantially. The meager lasting stable variance in Non-verbal IQ scores prior to grade seven no doubt accounts for the lack of lasting stable Verbal - Non-verbal difference prior to grade seven.

Associated Changes in Verbal and Non-verbal IQs.

If maturational, environmental, or situational differences account for much of the individual differences in both Verbal and Non-verbal IQ scores, changes in Verbal IQ scores would be expected to correlate with changes in Non-verbal IQ scores. The correlation coefficients in Table 19 reflect the relationship among the associated changes in Verbal IQ scores with Non-verbal IQ scores. Notice that all the correlations are positive and most are statistically significant,

yet low in magnitude. It can be concluded therefore, that, although there is a definite, but slight trend for increases or decreases in Verbal IQs to accompany increases and decreases in Non-verbal IQ scores. The magnitude of the relationship is too small to be of any real predictive value, yet it is consistent with theoretical expectations.

Table 18

Stability Coefficients for Verbal - Non-verbal IQ Differences for Students with a Complete set of Scores in Grades 2-11. (n = 440)

Grade Level	4	7	9	11
2	.08	.23	.13	.19
4		.32	.41	.34
7			.53	.50
9				.55

Table 19

Correlation Coefficients Between Changes in Verbal IQ Scores Across Time with Associated Changes in Non-Verbal IQ Scores. (n = 440)

Grade Level	4	7	9	11
2	.02	.12**	.08*	.08*
4		.18**	.25**	.22**
7			.13**	.17**
9				.18**

* p < .05 (directional test)
** p < .01 (directional test)

Predictive Value of Multiple IQ Scores.

To what extent are IQ scores other than the most recent score an indicator of eventual intellectual status. In practice, often only the most recent score is used and all prior IQ scores are ignored. In a few situations, all available IQ scores are averaged, with the assumption that errors of measurement will be reduced. The averaging procedure implicitly assumes constancy of true IQ scores over time. A third strategy would be to employ a multiple regression model which would optimally weight all available scores to maximize the prediction of a subsequent score. From a practical perspective, the multiple regression approach suffers from a lack of simplicity and would require a convincing demonstration of superiority in order for it to offset its practicability drawbacks.

The Multiple Regression Approach. With grade eleven Verbal IQ as the dependent variable, the correlation of the most recent score (grade nine, Verbal IQ) rose from .827 to .887 when all the 15 variables from grades 1, 2, 4, 7, and 9 were weighted optimally. Accordingly the standard error of estimate declined from 7.8 to 6.6. The R-value increased .05 with the addition of the second variable (grade two, Total IQ), and an additional .02 with the addition of a third variable (grade four, Verbal IQ). The fourth predictor (grade seven, Verbal) increased R by only .01. All other subsequent increases were negligible (.003 or less).

Using grade eleven Non-verbal IQ as the dependent variable, the correlation of the most recent Non-verbal IQ (grade nine), r = .67, rose to .753, .764, and .769 when grade seven Non-verbal, grade four Total, and grade two Non-verbal were employed sequentially as predictors. All other variables increased R negligibly.

Using Total IQ as the dependent variable, the correlation of .829 for the most recent Total IQ score increased to .88 with the addition of the best four additional predictors.

In conclusion, scores in addition to the most recent scores produced an increase in the predictability of grade eleven IQ scores, but the increases were not substantial and appear to not be justified for most educational purposes. This pattern may have resulted from the fact that the most recent score (grade nine) correlated very highly with the criterion, hence making improvement difficult. Had the question been addressed eliminating the grade nine scores, the degree of improvement would no doubt have increased considerably. Nevertheless appropriate intellectual assessment should entail retesting every two or three years, hence the question was addressed in a realistic manner. In addition, the issue could have been addressed employing a different criterion, yet the grade eleven performance is the most definitive, educationally significant criterion.

The Use of the Average of Prior Scores as Predictors.

Optimal weighting strategies has little promise of being used generally, because of the idiosyncratic nature of the results. However, an average of previous scores can easily be implemented, if found to be fruitful means of identifying eventual intellectual status.

Table 20 given the correlation coefficients of the average of all previous tests of same type (Verbal, Non-verbal, or Total IQ) with various criteria. In parenthesis the correlation of the most recent test with the criterion is given for comparative purposes.

Table 20 reveals a strong trend toward the most recent score to being better for predicting subsequent IQ scores than the average of prior scores. The sole exceptions shows the average of IQs from grades one and two to be better than grade two IQs alone in both the Verbal and Total IQ areas. In all other instances the difference is negligible or the average IQ from previous is poorer than the use of only the most recent score. In some instance, the difference was substantial. For example, Total grade 7 IQs alone correlated .88 with Total grade nine IQs. Yet when the average of previous Total IQ scores from grades 1, 2, 4, and 7 were used, the correlation dropped to .76!

It can be concluded that the averaging of other scores with the most recent score is definitely not recommended (except in the rare situation where scores are available at both grades one and two). The most recent score is the best indicator of future performance. If other scores are to be used, a multiple regression equation will be needed so that the prediction can be improved slightly rather than attenuated.

Factorial Composition of Verbal vs. Non-Verbal IQ.

A factor analysis of Verbal and Non-verbal IQs extending over grade levels 1-11 was performed on the class of '69 data in order to obtain a better understanding of the conceptual meaning of the two types of IQ scores. Within the correlation matrix are two subsets of correlations which approached simplex structure, the inter-correlations of Verbal IQs and the intercorrelations of Non-verbal IQs. Cronbach (1967) has cautioned that a factor analysis of a correlation matrix with simplex structure should be viewed as only heuristic strategy, not as a method for discovering underlying developmental processes.

A principal-axis solution was performed with squared multiple correlations as communality estimates. The Harris-Kaiser method of oblique factor transformation was selected as it appears to have greatest promise for approximating simple structure with factorially complex data (Hakstian, 1971). The independent clusters solution was attempted initially, but failed to give results as good as those derived from the A'A proportional to L primary factor solution (in which

Table 20

Correlation of the Average of Previous Tests Compared to the Correlation of Only the Most Recent Test in the Set (in parenthesis).[a]

Grade Levels Averaged	Criterion Grade 4	7	9	11
Verbal IQ				
1 + 2	.61 (.46)	.60 (.53)	.53 (.51)	.61 (.54)
1 + 2 + 4		.73 (.79)	.68 (.75)	.74 (.73)
2 + 4		.74 (.79)	.70 (.75)	.72 (.73)
1 + 2 + 4 + 7			.77 (.86)	.80 (.80)
2 + 4 + 7			.80 (.86)	.79 (.80)
1 + 2 + 4 + 7 + 9				.84 (.85)
2 + 4 + 7 + 9				.84 (.85)
Non-Verbal IQ				
1 + 2	.41 (.47)	.51 (.49)	.39 (.45)	.48 (.47)
1 + 2 + 4		.59 (.61)	.49 (.49)	.56 (.56)
2 + 4		.60 (.61)	.58 (.59)	.51 (.56)
1 + 2 + 4 + 7			.61 (.78)	.67 (.73)
2 + 4 + 7			.69 (.78)	.69 (.73)
1 + 2 + 4 + 7 + 9				.71 (.76)
2 + 4 + 7 + 9				.74 (.76)
Total IQ				
1 + 2	.65 (.63)	.65 (.62)	.58 (.62)	.66 (.63)
1 + 2 + 4		.75 (.77)	.67 (.74)	.74 (.73)
2 + 4		.75 (.77)	.72 (.74)	.73 (.73)
1 + 2 + 4 + 7			.76 (.88)	.81 (.83)
2 + 4 + 7			.81 (.88)	.81 (.83)
1 + 2 + 4 + 7 + 9				.85 (.86)
2 + 4 + 7 + 9				.84 (.86)

A refers to the primary factors). The initial A'A proportional to L primary factor-pattern solution was converted to the corresponding reference structure solution to facilitate the interpretability of the factors. (The solution reported in Table 21 gives the correlations of the various Verbal-Non-verbal IQ scores with the references axis.)

Factor I in Table 21 is identified as a verbal factor which becomes strongly evident beginning in grade four. Factor II is a Non-verbal factor which is substantially defined by the Non-verbal IQs at grade seven and subsequently. Factor III appears to be a temporary grade one, and to a lesser extent grade two, effect. It may reflect transitory maturational or test-taking abilities. Factor IV appears to reflect a transitory non-verbal factor that is evident in grade two, and four, and to a slight extent, in grade seven.

The results in Table 21 suggests that separate Verbal and Non-verbal IQs should not be reported prior to grade four, since the related IQ scores represent different factors than do later IQ scores bearing the same label. (A separate factor analysis using orthogonal (varimax) protrayed an identical pattern, yet without the degree of simple structure depicted in Table 21.)

CONCLUSIONS AND RECOMMENDATIONS

Conclusions.

The study provides support for the following generalizations:

1. Contrary to most textbook treatments of the topic and prevailing professional impression, there is little or no information on the stability of IQ scores from group tests especially (1) at the primary grades, or (2) for time intervals greater than three years, or (3) on currently used tests, or (4) on Non-verbal tests.

2. Verbal IQs are much more stable than Non-verbal IQ scores, especially prior to junior high school.

3. At grades one and two Verbal and Non-verbal IQs reflect relatively little lasting variance, even over short intervals.

4. By grade four, Verbal IQs show a very high relationship (about .8) with corresponding IQs thereafter.

5. Non-verbal IQs at grade seven reflect about the same degree of stability as Verbal IQs at grade four.

6. Total IQs have a pattern of stability highly similar to that of Verbal IQ scores.

7. IQ scores from group tests to be considerably less stable than Stanford-Binet IQs prior to grade seven, after which there is little differences in stability.

Table 21

Correlations of Verbal and Non-verbal IQ Scores with Reference Axes.

Sample I	Reference Axis I[a]	Reference Axis II[a]	Reference Axis III[a]	Reference Axis IV[a]
Gr. 1 Lang - CTMM	.15	.06	.42	.05
Gr. 1 N-Lang- CTMM	-.02	.03	.42	.09
Gr. 2 Lang - CTMM	.21	.07	.24	.07
Gr. 2 N-Lang- CTMM	.06	.13	.16	.30
Gr. 4 Lang - CTMM	.58	-.01	.11	.17
Gr. 4 N-Lang- CTMM	.08	.19	.06	.35
Gr. 7 Verbal - L-T	.63	.12	.02	.08
Gr. 7 N-Verbal-L-T	.21	.43	.05	.19
Gr. 9 Verbal - L-T	.65	.19	.00	-.01
Gr. 9 N-Verbal-L-T	.14	.58	.03	.05
Gr.11 Verbal - L-T	.55	.17	.11	.05
Gr.11 N-Verbal-L-T	.10	.48	.06	.10

[a]Reflected

8. In the main, there is considerable change in true Verbal IQ scores prior to grade four, but little change thereafter.

9. In the main, there is considerable change in true Non-verbal IQ scores in elementary schools, but only minor changes thereafter.

10. Differences in Verbal and Non-verbal IQs show meager stability over time prior to grade seven, after which there is considerable stability.

11. The use of IQ scores in addition to the most recent score offers little promise for improving the determination of eventual intellectual status. Multiple regression can contribute slight improvement in prediction, wherever the averaging of prior performances lessened the accuracy in prediction.

12. The factor analytic study confirmed informal observations made from a study of the general correlation matrix -- Verbal IQ variance become stable at about grade four, Non-verbal IQ variance stabilizes at about grade seven.

Recommendations.

1. Group intelligence tests not be routinely administered in grades one and two unless the users and interpreters of these scores have received extensive training in the meaning and limitations of these scores. The long-term meaning of these scores is far below that of individually-administered verbal intelligence tests, yet the typical teacher lacks the training necessary to properly differentially interpret -- hence an "IQ score is an IQ score." A reading readiness test is more useful for identifying prospective success and failure in grade one (Hopkins and Sitkei, 1969), and the consequences of misinterpretation is far less critical.

2. Non-language and Non-verbal IQ scores must be interpreted even more cautiously than Verbal IQ scores.

3. Teachers and other users of IQ scores from group tests need to be made aware of the fact that even when there is considerable overall stability, the scores of a few individuals will vary greatly.

4. No meaning should be attached to differences in Verbal and Non-verbal IQ scores prior to grade seven.

References

Anastasi, A. Differential Psychology (3rd ed.). New York: The Macmillan Co., 1960.

Bloom, B. S. Testing cognitive abilities and achievement. In N. L. Gage (ed.) Handbook of Research on Testing. Chicago: Rand McNally, 1963.

Bloom, B. S. Stability and Change in Human Characteristics. New York: John Wiley and Sons, Inc., 1964.

Bradshaw, D. H. Stability of California Test of Mental Maturity IQ's from the second to the fourth grade. Educational and Psychological Measurement, 1964, 24, 935-939.

Collmann, R. D., & D. Hewlyn. Changes in Terman-Merrill IQs of mentally retarded children. Am. J. Ment. Def., 1958, 63, 307-311.

Cronbach, L. J. Year-to-year correlations of mental tests: A review of the Hofstaeffer analysis. Technical Report #1. Project on Individual Differences in Learning Ability as a Function of Instructional Variables, Stanford University, June, 1967.

Doppelt, J. E. & G. K. Bennett. A longitudinal study of the differential aptitude tests. Educational and Psychological Measurement, 1951, 11, 228-237.

Eagle, N. The stability of Lorge-Thorndike IQ scores between grades three and four and grade eight. Journal of Educational Research, 1966, 60, 164-165.

Finley, C. J., J. M. Thompson, & A. Cognata. Stability of the California Short Form Test of Mental Maturity: Grades 3, 5, and 7. Calif. Journal of Educational Research, 1966, 17, 157-168.

Garrison, S. C., & M. S. Robinson. A study of retests. Journal of Educational Research, 1925, 11, 190-196.

Goldin, M. R. & S. Rothschild. Stability of intelligence quotients of metropolitan children of foreign-born parentage. Elementary School Journal, 1942, 42, 673-676.

Hakstian, A. R. A comparative evaluation of several prominent methods of oblique factor transformation. Psychometrika, 1971, 36, 175-194.

Hays, W. L. Statistics for Psychologists. New York: Holt, Rinehart, & Winston, 1963.

Hirsh, N. D. M. Experimental study upon 300 school children over a six-year period. Genet. Psych. Monogr., 1930, 6, 487-549.

Hopkins, K. D. An empirical analysis of the efficacy of the WISC in the diagnosis of organicity in children of normal intelligence. Journal of Genetic Psychology, 1964, 105, 163-172.

Hopkins, K. D., & M. H. Bibelheimer. Two and four-year constancy of CTMM language and nonlanguage IQs. Toward A Professional Identity in School Psychology. Published by the California Association of School Psychologists and Psychometrists, March, 1963.

Hopkins, K. D., & M. H. Bibelheimer. Five-year stability of intelligence quotients from language and nonlanguage group tests. Child Development, 1971 42, 645-649.

Hopkins, K. D., & Sitkei, E. G. Predicting grade one reading performance: Intelligence vs. Reading Readiness Tests. Journal of Experimental Education , 1969, 37, 31-33.

Horrocks, J. E. Assessment of behavior. Columbus; Ohio: Charles E. Merrill Books, Inc., 1964.

Knezevich, S. The constancy of the IQ of the secondary school pupil. Journal of Educational Research, 1946, 39, 506-516.

Lauderbach, J. C. & E. Hause. On the reliability and validity of derived scores yielded by the McCall multi-mental scale. Journal of Applied Psychology, 1932, 16, 322-323.

Layton, W. L. The relation of ninth grade test scores to twelfth grade test scores and high school rank. Journal of Applied Psychology, 1954, 38, 10-11.

Layton, W. L., & E. O. Swanson. Relationship of ninth grade differential aptitude test scores to eleventh grade test scores and high school rank. Journal of Educational Psychology, 1958, 49, 153-155.

Lincoln, E. A., & V. L. Wadleigh. Change and constancy in group test IQ's. School and Society, 1932, 35, 606-608.

Lorge, I., R. L. Thorndike, & E. Hagen. The Lorge-Thorndike Intelligence Tests: Technical Manual. Revised edition. Boston: Houghton Mifflin Co., 1966.

Meyer, W. J. The stability of patterns of Primary Mental Abilities among junior high and senior high school students. Educational and Psychological Measurement, 1960, 20, 795-800.

Meyer, W. J., & A. W. Bendig. A longitudinal study of the Primary Mental Abilities Test. Journal of Educational Psychology, 1961, 52, 50-60.

Miller, W. S. Variation of IQ's obtained from group tests. Journal of Educational Psychology, 1933, 24, 468-474.

Olkin, I. Correlations revisited. In J. C. Stanley (ed.), Improving Experimental Design and Statistical Analysis. Chicago: Rand McNally, 1967.

Pintner, R., & M. Stanton. Repeated tests with the CAVD scale. Journal of Educational Psychology, 1937, 28, 494-500.

Seagoe, M. V. An evaluation of certain intelligence tests. Journal of Applied Psychology, 1934, 18, 432-436.

Thorndike, R. L. The effect of interval between tests and retest on the constancy of the IQ. Journal of Educational Psychology, 1933, 24, 543-549.

Thorndike, R. L. "Constancy" of the IQ. Psychological Bulletin, 1940, 37, 167-186.

Traxler, A. Reliability, constancy, and validity of the Otis IQ. Journal of Applied Psychology, 1936, 18, 241-251.

Tully, G. E., & B. W. Hall. Note: Test-retest reliability of the School and College Ability Test. Journal of Educational Measurement, 1965, 2,

Tyler, L. E. The stability of patterns of primary abilities among grade school children. Educational and Psychological Measurement, 1958, 18, 769-774.