DOCUMENT RESUME

ED 069 782 TM 002 267

AUTHOR Hopkins, Kenneth D.; Bracht, Glenn H.
TITLE A Longitudinal Study of Constancy of Reading Performance.
NOTE 12p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Achievement Tests; Comparative Analysis; Correlation; Educational Research; Elementary Grades; *Grade Equivalent Scores; Longitudinal Studies; Measurement Instruments; *Performance Tests; Reading Comprehension; *Reading Tests; Secondary Grades; *Standardized Tests; Statistical Analysis; Tables (Data); Vocabulary Skills
IDENTIFIERS Iowa Tests of Basic Skills; Iowa Tests of Educational Development; Metropolitan Achievement Tests

ABSTRACT

The stability of reading performance, as measured by the Metropolitan Achievement Tests, Iowa Tests of Basic Skills, and Iowa Tests of Educational Development, was studied using students in grades 1 through 7 and grades 9 and 11. A reading vocabulary test and a reading comprehension test are included in all three test batteries. The standard scores on the three tests were pooled to obtain a composite reading score for three independent samples of students. Sample I consisted of grades 3-6 and 9 and 11, the number of students varying from a low of 71 (grade 5) to a high of 1,116 (grade 9); Sample II was made up of students from grades 1-7 and grade 9, the number varying from 520 (grade 2) to 1,240 (grade 7); and Sample III contained students from grades 1-6, varying in number from 1,095 (grade 6) to 1,320 (grade 1). Results of the study showed that substantial long-term stability was reflected in both the vocabulary and comprehension tests; grade 1 scores correlated above .5 with all subsequent measures. By the end of the primary grades, students' scores correlated above .70 with all subsequent measures. When the coefficients were correlated for attenuation to allow an estimate of the relationships after errors of measurement on the test were removed, the values were about .10 higher. It is concluded that although reading does not represent temporary maturational status for most pupils, it does have substantial relationship with terminal achievement levels in both reading vocabulary and comprehension. (DB)

Dr. Kenneth D. Hopkins
Laboratory of Educational Research
University of Colorado
Boulder, Colorado 80302

Dr. Glenn H. Bracht
Department of Guidance and Educational Psychology
Southern Illinois University
Carbondale, Illinois 62901

A LONGITUDINAL STUDY OF CONSTANCY OF READING PERFORMANCE

Is early success in reading related to long-term reading competency? Are differences in initial reading success the result of age and maturational differences that dissipate with time? Are "slow starters" just immature pupils who eventually achieve normally or do they continue to have "ignition" trouble? Although there has been considerable research on the question of IQ constancy, it is surprising that so little attention has been devoted to the related issue of achievement stability, a topic of much greater educational and social importance.

Bloom's survey of research on the predictability of achievement data failed to locate any research that studied even short-term consequences of grade on reading performance. The published studies on the topic of achievement constancy in the area of reading revealed by the literature search are summarized in Table 1. The information in Table 1 reveals that most studies have utilized small N's (which yield unreliable stability estimates) and that all published studies have extended over a five-year interval or less.

Table 1

Summary of Studies on the Stability of Reading Performance.

TEST	N	GRADES	r	INVESTIGATOR
Metropolitan Reading	105	2 - 5	.76	Townsend (1944)
Standford Reading	47	2.9 - 6.9	.67	Hildredth (1936)
Stanford Paragraph Meaning	81	5 - 6	.77	
ITBS Reading	27	6.9 - 8.9	.75	Kvaraceus and Lanigan (1948)
Cooperative Reading Comprehension		7 - 12	.77	Traxler (1950)
		8 - 12	.76	
		9 - 12	.82	
		10 - 12	.82	
Nelson Denny Reading	517	13 - 14	.83	Silvey (1951)
ITBS and ITED	256	7 - 9	.83	
	251	7 - 11	.79	
ITBS Reading	900	3 - 8	.77	Merenda and Jackson (1965, 196
ITBS Vocabulary			.76	
ITBS Reading	9972	5 - 8	.79	Linn (1969)
ITBS Vocabulary			.83	

The purpose of this study was to investigate the stability of reading performance as measured by standardized tests at various intervals over the initial eleven grade levels.

Method

Reading tests from three popular achievement batteries were used in the study: Metropolitan Achievement Tests (MAT), Iowa Tests of Basic Skills (ITBS), and Iowa Tests of Educational Development (ITED). The use of different achievement tests is both a strength and a weakness. Varying the tests increases the generalizability of findings--showing that the results are not limited to a particular measuring instrument. At the same time the nature of the variables being measured differs somewhat among the tests, hence the stability estimates will be conservative in nature.

Tests were administered annually in grades 1 through 7, and also in grades 9 and 11. A reading vocabulary test is included in all three test batteries at each grade level. (In the MAT battery the test is called Word Knowledge.) A reading comprehension test is included in all three test batteries, in fact three reading comprehension tests are included in the ITED battery: Ability to Interpret Reading Materials in the Social Studies, Ability to Interpret Reading Materials in the Natural Sciences, and Ability to Interpret Literary Materials; hence the standard scores on the three tests were pooled to obtain a composite reading score for each student.

Three independent samples of students were included in the study: The means and standard deviations of the reading vocabulary and comprehension scores are reported for each grade level in Table 2, along with sample sizes and dates of testing. The standard deviations and means are

based on the scores of all students in a grade level. In order to determine whether the degree of variability in the samples differed from the population variability, the standard deviations of grade-equivalent scores were compared with corresponding values for the norming population reported in the test manuals. (MAT and ITED estimates were computed from s_e and r_{11}: $s = s_e\sqrt{1 - r_{11}}$.) A comparison of these values with the sample values revealed the sample had greater variability in 18 of 39 instances and less variability in the remaining 21 comparisons. In most cases the differences were small. Hence the variability in the samples appears to be quite representative.

The standard deviations were also computed for only the students present at the most extreme grade levels to assess potential selection effects on the variability within the samples. The average variability of this smaller sample differed only slightly (about .03σ in grade 1 and .02σ in grade 11) from the total sample within each grade level, hence the stability coefficients are not non-representative as a consequence of atypical sample variability.

Achievement Stability

The stability and generalizability coefficients for the reading vocabulary scores are reported in Table 3. Any student who was enrolled in any two or more of the grade levels during the specified years was included in the sample. The N on which the stability coefficient is based is given below the diagonal for each sample. For example, in Sample II the correlation between grade one and grade two reading vocabulary was .64, which was based on 415 pairs of scores. A factor which must be kept in mind in the interpretation of the stability coefficients is the change in test battery at grades 4 and 9. The slight but generally consistent decreases in stability at grades 4 are probably related to the change in test battery rather than a real change

Table 2

Grade Levels, Sample Sizes, Means and Standard Deviations, and Date of Testing for Samples I, II, and III.

SAMPLE			Vocabulary[a]		Comprehension[b]			
	Grade	N	$\bar{X}$[c]	s	$\bar{X}$[c]	s	Date of Testing	Test and Form
I.	3	461	4.64	1.37	4.79	1.49	3/60	MAT, Elem. A
	4	452	4.92	1.31	5.08	1.66	9/60	ITBS, 1
	5	71	6.36	1.57	6.32	1.79	9/61	ITBS, 2
	6	1024	7.37	1.73	7.31	1.53	9/62	ITBS, 1
	7	1065	8.56	1.62	8.39	1.58	10/63	ITBS, 2
	9	1116	16.36	5.52	15.33	5.30	11/65	ITED, X4
	11	919	19.83	5.42	18.77	5.83	12/67	ITED, Y4
II.	1	540	2.10	.50	2.12	.62	3/60	MAT, Prim. I, A
	2	520	3.35	.91	3.63	.85	3/61	MAT, Prim. II, B
	3	1115	4.65	1.27	4.72	1.38	3/62	MAT, Elem. B
	4	1115	4.92	1.14	5.07	1.52	9/62	ITBS, 1
	5	1185	6.30	1.51	6.28	1.76	10/63	ITBS, 2
	6	1195	6.91	1.48	6.90	1.47	10/64	ITBS, 3
	7	1240	8.14	1.57	7.81	1.56	12/65	ITBS, 4
	9	1050	16.06	5.22	15.30	5.30	12/67	ITED, X4
III.	1	1320	2.18	.53	2.19	.65	3/63	MAT, Prim. I, A
	2	1250	3.45	.92	3.63	.90	3/64	MAT, Prim. II, B
	3	1275	4.80	1.25	4.88	1.43	3/65	MAT, Elem. A
	4	1315	4.77	1.11	4.71	1.28	10/65	ITBS, 1
	5	1140	5.84	1.25	5.97	1.38	9/66	ITBS, 2
	6	1095	6.92	1.44	6.98	1.47	9/67	ITBS, 3

[a] "Word Knowledge" test on the MAT

[b] "Reading" tests of the MAT and ITBS; average of three reading comprehension tests of the ITED.

[c] Grade equivalent units on the MAT and ITBS; standard scores on the ITED

Table 3

Stability Coefficients (Above Diagonals) and Corresponding Ns (Below Diagonals) for Reading Vocabulary Scores for Samples I, II, and III.

		TESTS AND GRADE LEVEL									
		MAT			ITBS				ITED		
Sample I Grade	Test	1	2	3	4	5	6	7	9	11	r_{tt}[a]
3	MAT[a]				.82	.86	.81	.79	.79	.76	.95
4	ITBS			388		.80	.80	.76[b]	.78	.77	.91
5	ITBS			368	394		.89	.87	.85	.83	.91
6	ITBS			352	373	776		.88	.87	.85	.91
7	ITBS			341	362	707	891		.88	.87	.89
9	ITED			314	324	611	751	836		.91	.93
11	ITED			285	281	532	642	697	878		.95
Sample II											
1	MAT[a]		.64	.56	.56	.52	.55	.56	.51		.79
2	MAT[a]	415		.78	.72	.65	.66	.66	.59		.92
3	MAT[a]	400	435		.80	.79	.76	.74[b]	.67		.95
4	ITBS	375	400	975[b]		.82	.79	.76[b]	.71		.88
5	ITBS	355	375	885	990		.85	.83	.81		.91
6	ITBS	345	360	825	910	1030		.85	.81		.88
7	ITBS	320	335	765	840	930	1040		.83		.88
9	ITBS	300	305	685	740	795	870	1005			.93
Sample III											
1	MAT[a]		.70	.64	.55	.58	.56				.82
2	MAT[a]	1000		.80	.70	.69	.65				.93
3	MAT[a]	885	995		.78	.78	.76				.95
4	ITBS	815	895	1090		.80	.76				.87
5	ITBS	730	790	940	1070		.83				.86
6	ITBS	720	770	915	1025	1025					.88

[a]From test manuals, adapted to the variability of the sample using formula given by Guilford (1954, p. 392).

[b]Corresponding value from Merenda and Jackson (1969) was also .76.

[c]There was an increase in sample size at this grade level due to school district reorganization.

in the stability of reading vocabulary. The decreases are slight, however, indicating that the stability is not limited to intra-battery inferences but is quite general across competing achievement batteries.

Reading vocabulary scores were only moderately stable from grade 1 through grade 3 ($r \doteq .6$), which suggested considerable stability in pupils' scores on the vocabulary tests. The individual differences in pupils' reading vocabulary at grade one represents a definite lasting characteristic for the group. The correlation of .51 between vocabulary scores at grades one and nine indicates that, on the average, a pupil tended to be about half as far from the grade nine as he was from the grade one mean. Notice that the stability coefficients tend to increase as grade level is increased. Notice also that there is little loss in stability after the initial two or three years. The stability of reading vocabulary scores in grade two was considerably greater than that reflected in grade one performance, correlating about .6 with scores seven years later. Beginning at grade 3, the stability of reading vocabulary achievement was maintained at a high level through the secondary grades with an eight-year stability of .76 in Sample I. The stability of reading vocabulary achievement was extremely high for all groups by the beginning of grade five, with the stability coefficients approaching the tests' reliabilities. The generally higher stability coefficients in Sample I can be explained partially by the slightly greater variability of scores (cf. Table 2).

Since the correlation coefficients in Table 3 reflect true change and stability of reading vocabulary plus errors of measurement, the coefficients were corrected for attenuation to provide an estimate of the

stability of true scores in reading vocabulary. These disattentuated stability coefficients are given in Table 4.

The values in Table 4 provide estimates of the degree of relationship between reading vocabulary performance free from the contamination effects from errors of measurement, and hence address the theoretical issue of true stability better than the corresponding values found in Table 3. If the reliability of the MAT and ITBS reading vocabulary tests were to be increased to 1.0, the correlation between grade one with grade six scores would be expected to be .71 and .65 with grade nine scores, reflecting substantial long range implications of initial reading success. True reading vocabulary scores near the end of the primary cycle (2.7) were very highly related (r's = .83 - .88) to true scores in grade six and grade eleven (r = .80). The rank-order of pupils' true reading vocabularies change very little after four years of formal reading instruction (i.e. after 5.1), with disattenuated correlations with all measures thereafter approaching .9 or higher.

It can be concluded that reading vocabulary near the end of grade one gives a good indication of the reading vocabulary of pupils ten years later; the indication is excellent after the completion of grade four.

Reading Comprehension

The stability coefficients for reading comprehension tests are given in Table 5. Since three different standardized tests were employed, each one operationally defining reading somewhat differently, the coefficients must be viewed as conservative estimates. They are, however, generalizability coefficients (Cronbach, Rajaratnam, and Gleser, 1963) which have allowed both time and test battery to vary. The stability coefficients for reading comprehension

Table 4

Disattenuated Stability Coefficients for Reading Vocabulary for Samples I, II, and III.

Sample I		Grades								
Grade	Test	1	2	3	4	5	6	7	9	11
3	MAT[a]				.88	.92	.87	.84	.83	.80
4	ITBS					.88	.88	.84	.85	.83
5	ITBS						.98	.97	.92	.89
6	ITBS							.98	.95	.91
7	ITBS								.97	.95
9	ITED									.97
Sample II										
1	MAT		.75	.65	.67	.61	.66	.67	.60	
2	MAT			.83	.80	.71	.73	.73	.63	
3	MAT				.88	.85	.83	.81	.71	
4	ITBS					.92	.90	.86	.78	
5	ITBS						.95	.93	.88	
6	ITBS							.97	.89	
7	ITBS								.91	
Sample III										
1	MAT		.80	.73	.65	.69	.66			
2	MAT			.85	.78	.77	.72			
3	MAT				.86	.86	.83			
4	ITBS					.93	.87			
5	ITBS						.95			

Table 5

Stability Coefficients of Reading Comprehension[a] Before (Above Diagonals) and After (Below Diagonals) Correction for Attenuation Scores for Samples I, II, and III.

		TESTS AND GRADE LEVEL									
Sample I		MAT			ITBS			ITED			r_{tt}
Grade	Test	1	2	3	4	5	6	7	9	11	
3	MAT				.82	.79	.80	.77	.76	.72	.96
4	ITBS			.86		.79	.78	.78	.77	.73	.96
5	ITBS			.82	.82		.85	.82	.79	.78	.96
6	ITBS			.83	.81	.89		.87	.85	.83	.91
7	ITBS			.82	.83	.85	.91		.85	.85	.92
9	ITED			.80	.82	.84	.90	.92		.87	.93
11	ITED			.76	.77	.82	.87	.91	.94		.95
Sample II											
1	MAT		.59	.58	.59	.57	.50	.53	.53		.81
2	MAT	.68		.71	.66	.65	.65	.62	.58		.90
3	MAT	.66	.76		.77	.76	.75	.70	.66		.96
4	ITBS	.67	.71	.81		.83	.79	.74[b]	.71		.95
5	ITBS	.65	.70	.79	.87		.83	.78	.78		.96
6	ITBS	.58	.72	.80	.85	.89		.82	.80		.91
7	ITBS	.62	.68	.75	.79	.82	.90		.79		.92
9	ITED	.61	.63	.70	.76	.82	.87	.85			.95
Sample III											
1	MAT		.64	.62	.61	.61	.59				.83
2	MAT	.73		.72	.69	.71	.65				.92
3	MAT	.69	.77		.76	.77	.73				.96
4	ITBS	.69	.75	.80		.81	.76				.93
5	ITBS	.69	.77	.81	.87		.84				.93
6	ITBS	.67	.71	.78	.83	.91					.91

[a]Actual test titles: "Reading" for MAT and ITBS, and the average of three reading interpretation tests (tests 5-7) on the ITED.

[b]Corresponding value from Merenda and Jackson (1969) was .77.

for reading comprehension are very similar to corresponding values for vocabulary given in Table 3; the mean and mode vocabulary stability coefficients being .02 and .01 larger than corresponding comprehension values.

The disattenuated stability coefficients are given below the diagonal for each sample in Table 5. These values averaged about .05 less than corresponding values for vocabulary indicating that there is less true-score stability in comprehension than vocabulary, although both are very stable after grade three. (The stability coefficients in Tables 3 and 5 agree very closely with those from Linn (1969) (5) and Merenda and Jackson (1969) (6) who studied the grade 4 - 7 and 5 - 8 intervals, respectively using the ITBS.)

Summary

The stability of reading vocabulary and comprehension were studied over the grade one to grade eleven interval using three large samples of students. Substantial long-term stability was reflected in both types of tests; grade one scores correlated above .5 with all subsequent measures. By the end of the primary grades, students' scores correlated above .70 with all subsequent measures. When the coefficients were correlated for attenuation to allow an estimate of the relationships after errors of measurement on the test were removed, the values were about .10 higher.

Early performance in reading does not represent temporary maturational status for most pupils, but has substantial relationship with terminal achievement levels in both reading vocabulary and comprehension.

REFERENCES

1. Cronbach, L. J, Rajaratnam, N., and Gleser. G. C. "Theory of Generalizability: A Liberalization of Reliability Theory," The British Journal of Statistical Psychology, 16 (Nov. 1963), 137-163.

2. Hildredth, Gertrude. "Results of Repeated Measurement of Pupil Achievement," Journal of Educational Psychology, 21 (1936), 286-296.

3. Krantz, L. L. "The Relationship of Reading Abilities and Basic Skills of the Elementary School to Success in the Interpretation of the Content Materials in the High School," Journal of Experimental Education, 26 (1957), 97-114.

4. Kvaraceus, W. C., and Lanigan, M. A. "Pupil Performance on the Iowa Every Pupil Tests of Basic Skills Administered at Half-year Intervals in the Junior High School," Educational and Psychological Measurement, 8 (1948), 93-100.

5. Linn, R. L. "A Note on the Stability of the Iowa Tests of Basic Skills," Journal of Educational Measurement, 6 (1969), 29-30.

6. Merenda, P. F., and Jackson, R. M. "Relationship between Fourth Grade and Seventh Grade Performance on the Iowa Tests of Basic Skills," Journal of Educational Measurement, 5 (1968), 163-165. (cf. Erratum, 6 (1969), 31.)

7. Silvey, H. M. "Changes in Test Scores after Two Years in College," Educational and Psychological Measurement, 11 (1951), 494-502.

8. Townsend, Agatha. "Some Aspects of Testing in the Primary Grades," Educational Records Bulletin, 40 (1944), 51-54.

9. Traxler, A. E. "Reading Growth of Secondary School Pupils During a Five Year Period." Achievement Test Program in Independent Schools and Supplement Studies, ERB Bulletin, 54 (1950).