TEST-RETEST RELIABILITY OF THE RAVEN PROGRESSIVE MATRICES TEST (FORM 1938) AND THE CALIFORNIA TEST OF MENTAL MATURITY, LEVEL 4 (S-F 1963)[1]

G. Emerson Tully
Florida Board of Regents

In schools generally throughout the United States, verbal tests of general ability have clearly won approval from school personnel over non-verbal tests designed to assess general ability. This state of affairs in testing is not a clear gain in measurement. The verbal test of general ability, when used above the primary grades, is one of the best predictors of the future academic success of students, and the preference shown by school administrators and guidance workers for the verbal measure is understandable. The verbal test, however, with its items that closely resemble classroom tasks, must be looked upon as a measure of school-developed general ability, and not as an inclusive measure of the full spectrum of general ability. By its nature and design, the verbal measure of general ability may bring about a degree of inaccuracy in the assessment of all components of general ability for some students at the same time that it serves as a sound basis for predicting the subsequent school success for other students.

The way in which the verbal test introduces an element of inaccuracy in the appraisal of general ability is apparent. General ability tests are intended to elicit maximum effort on the part of the examinee. The student with poorly developed reading and computational skills, in contrast to the student who is proficient in these classroom-developed skills, cannot be expected to display the heightened motivation that should characterize examinee behavior on a test of maximum performance. The inability of the poor student to deal successfully with the verbal item not only lowers his motivation, but may prevent him from making a good showing even though he desires to earn his best score.

If the preference of schools for verbal tests is bringing about some inaccuracy in estimating the general ability of students, especially those students who are "disadvantaged" in that they do not come from home environments in which classroom success is encouraged and supplemented, then the measurement specialist may well inquire if there is a dimension in mental measurement that may be better

[1]This investigation was supported by a grant from the Research Council, Florida State University.

assessed by non-verbal tests than by verbal tests. One of the first steps to be taken in such a course of action is for the testing specialist to ascertain the comparative technical characteristics of both the verbal and non-verbal tests. Although the literature of testing contains some information of this sort, reliability and validity reports of non-verbal tests are far outnumbered by reports dealing with verbal measures.

The principal aim of this study was to establish the test-retest reliability of the Raven Progressive Matrices Test (PM) 1938, and the California Test of Mental Maturity (CTMM) Level 4, S-F, 1963. A related purpose was to determine the degree of overlap in the scores of subjects on the tests. The PM is almost wholly a non-verbal test, in contrast to the CTMM, which yields a non-language, language, and a total score. Knowledge of the extent to which the two instruments measure the same abilities may be helpful to the counselor seeking to evaluate the two tests as instruments for assessing the general ability of students with differing school-developed skills.

Although the central purpose of this study was intended to deal with the technical aspects of the tests, the investigator also sought to observe the pattern of scores of the subjects on the PM and the CTMM for 1963 and 1964 and to relate this pattern to differences in schools.

Procedure

In October, 1963, the PM, the CTMM and the School and College Ability Test, (SCAT) 3B were administered to a sample of tenth-grade subjects in four Florida high schools. A year later, the PM and the CTMM were given to the same subjects in the four schools who were then enrolled in grade eleven.[1] The four schools in which the tests were given were selected to represent four types of schools that for many years characterized Florida secondary schools: an urban school enrolling white students, an urban Negro school, a white rural school, and a Negro rural school. At the time of the study, there were no integrated schools in Northwest Florida where the study was being conducted.

In 1963, a section of the tenth-grade selected randomly from the tenth grade sections of each school was given the three tests in this order: PM, CTMM, and SCAT. In 1964,

[1]The SCAT was not given to the subjects of this study in 1964. Data for establishing the test-retest reliability of the SCAT were obtained in 1962 and 1963. Although some reference to the SCAT is made in the present study, the full report of the SCAT study made by Tully and Hall appears elsewhere (1965).

the subjects were tested with the PM first and the CTMM as the second and final test. Product moment correlations between the 1963 and 1964 scores on the PM and CTMM were computed to yield test-retest reliabilities. Produce moment correlations between PM and CTMM, PM and SCAT, and SCAT and CTMM were also determined to provide indices of the degree to which the tests were measuring similar mental functions.

Mean scores on the PM and the CTMM for each school for 1963 and 1964 were established. The four schools were ranked according to the quality of their instructional programs. This ranking was a highly subjective procedure and was based on the observations made by the investigator during four half-day visits to the schools during which he talked with principals, teachers and students, and observed classes in session.

Findings and Discussion

Test-retest correlations for the PM when a year intervenes between the testing sessions appear in Table 1. Except for the r's obtained for Schools 2 and 3, in which the subjects were relatively homogeneous in general ability as revealed by the PM SD's, the test-retest correlations compared favorably with a correlation of .67 established by Eysenck with a sample of normal adults after an interval of four weeks (1944).

Test-retest correlations for the CTMM ranged somewhat higher than for the PM, except for School 4, in which the performance of the subjects on the CTMM was uniformly low. Presumably, these homogeneously grouped low scorers resorted to considerable guessing during testing which seemingly accounted for the relatively low correlation of .50.

SCAT test-retest correlations for the total group of 98 subjects were as follows: V .90, Q .89, and T .93. Tully and Hall found that test-retest correlations for the SCAT were not affected by the order of placement of the SCAT in the series of tests (1965). In the present study, no attempt was made to gauge the impact of order on the stability of PM or CTMM scores.

The degree to which the three instruments of general ability may have been measuring the same mental abilities is revealed by the correlations in Table 2. A correlation of .81 between total scores of the verbal SCAT and less verbal CTMM suggested that generally the students exhibited some-

Table 1

Test-Retest Reliabilities for the Progressive Matrices Test (1938) and the California Test of Mental Maturity, Level 4 (S-F 1963)

		PROGRESSIVE MATRICES					CALIF. TEST OF MEN. MAT. (TOTAL SCORE)				
		1963		1964			1963		1964		
	N	MEAN	S.D.	MEAN	S.D.	r	MEAN	S.D.	MEAN	S.D.	r
School 1	30	38.0	10.35	41.6	9.53	.75	46.1	13.72	54.8	16.18	.87
School 2	21	45.2	5.51	47.8	5.23	.55	59.2	16.58	70.1	13.04	.78
School 3	21	43.6	7.13	44.8	6.84	.57	61.0	13.38	66.8	13.28	.92
School 4	26	27.0	11.54	30.8	12.99	.80	30.8	9.34	33.9	8.06	.50
GIRLS	41	37.5	10.85	40.3	10.62	.84	46.2	16.94	54.2	17.77	.92
BOYS	57	38.1	12.16	41.1	11.86	.81	49.4	18.50	55.7	20.10	.87
All subjects	98	37.8	11.63	40.8	11.36	.82	48.0	17.91	55.1	19.17	.89

Table 2

Intercorrelations of the Progressive Matrices Test (1938), California Test of Mental Maturity, Level 4, (S-F) 1963, and School and College Ability Test (3B)

	PM	CTMM(NON-L)	CTMM(LANG)	CTMM(TOTAL)	SCAT(V)	SCAT(Q)	SCAT(T)
PM		.65	.56	.65	.57	.56	.61
CTMM (NON-L)			.72	.93	.65	.63	.70
CTMM (LANG)				.92	.83	.63	.81
CTMM (TOTAL)					.80	.68	.81
SCAT (V)						.63	.92
SCAT (Q)							.86
SCAT (T)							

what the same performance on these two measures. Correlations of less magnitude between the PM and CTMM(T) and SCAT (T) of .65 and .61 respectively indicated differential performance by students on the PM and the two more verbal measures of general ability.

In the judgment of the investigator, School 2 (urban white) had the most effective overall instructional program. School 3 (rural white) the next most effective program, School 1 (urban Negro) the third most effective and School 4 (rural Negro) perhaps the least effective program of the four schools. If this subjective ranking of the instructional programs of the four schools was valid, then the patterns of scores yielded in 1963 and 1964 provided some evidence that the educational experiences of a particular school setting throughout the year were more related to gains on the CTMM than on the PM. Figures 1 and 2 show the fairly uniform pattern of gains on the PM in contrast to the less uniform pattern for the CTMM.

Summary

The Raven Progressive Matrices (Form 1938), a nonverbal measure of general ability, and the California Test of Mental Maturity Level 4 (S-F, 1963) were administered to 98 students in four high schools in Florida to determine (a) the test-retest reliabilities of the two instruments (b) inter-correlations of the tests, and (c) the relative performance subjects in contrasting school settings.

The PM yielded a coefficient of stability of .82 for the total sample when the period intervening between tests and retests was one year. When coefficients of stability were computed for each school, the range of scores was restricted bringing about substantially lower correlations for each school than for the total sample. The test-retest reliability coefficients for the two schools in which the subjects were more homogeneous in general ability as measured by the PM than the subjects on the other two schools were lower than the correlations obtained in the schools where the range of ability was greater.

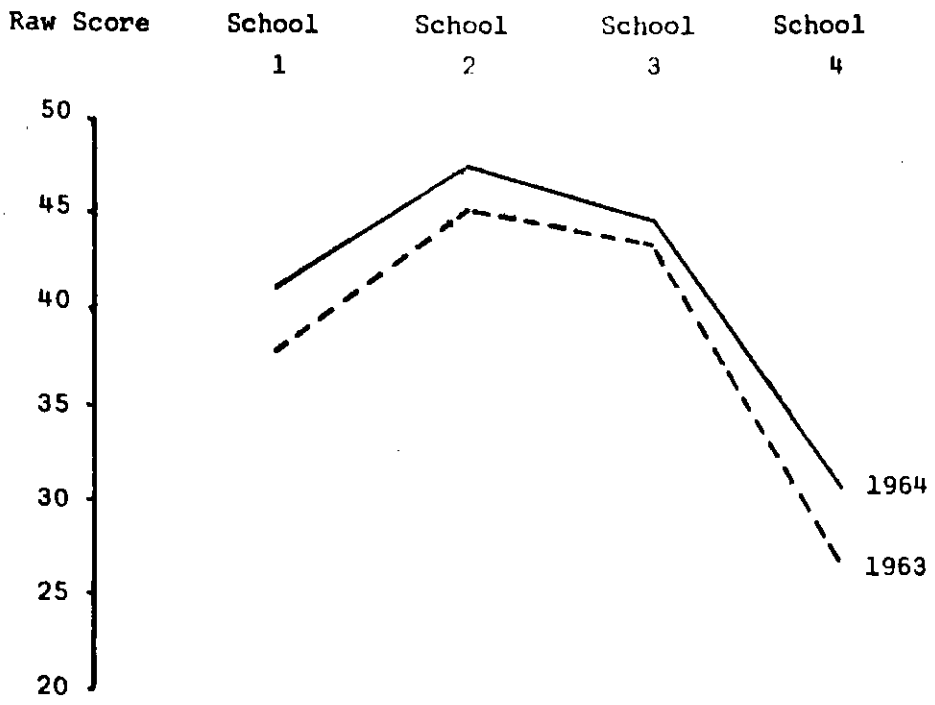

Fig. 1.--Mean total score on the PM (form 1938) obtained in four Florida schools in 1963 and a year later in 1964

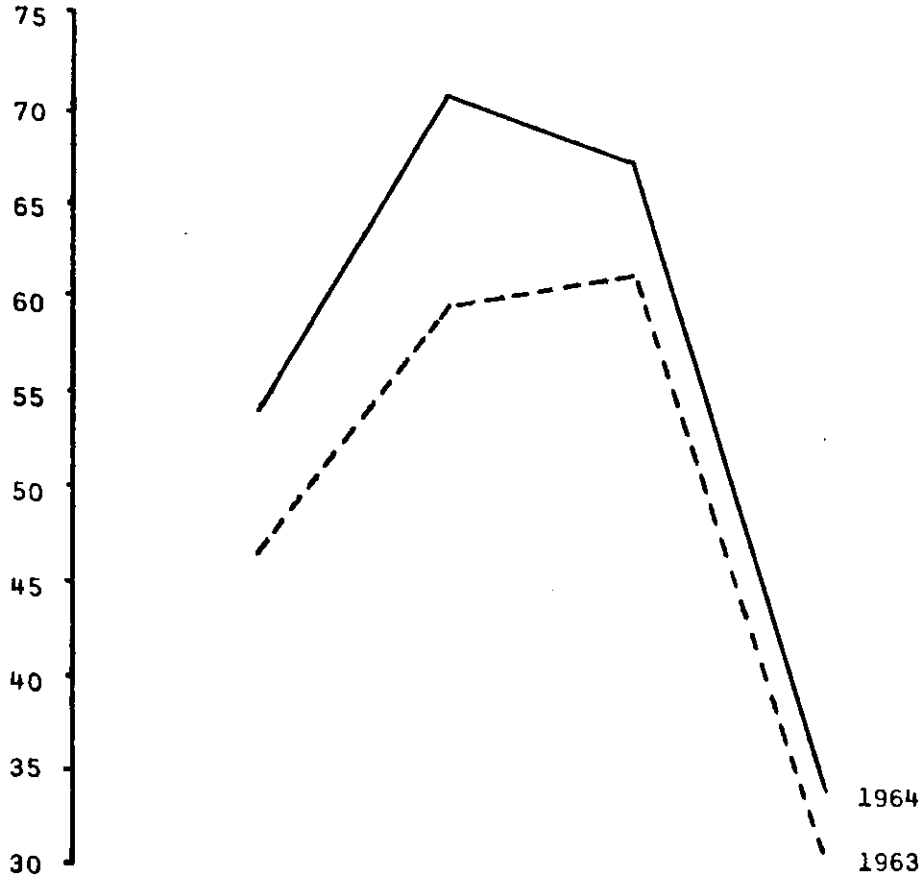

Fig. 2.--Mean total score on the CTMM (level 4, 5-4, 1963) in four Florida schools in 1963 and a year later in 1964

The test-retest reliability for the total score of the CTMM for the total sample (N=98) was .89. Correlations for three of the four schools were generally high except for one school (School 1, .87; School 2, .78; School 3, .92 and School 4, .50). The relatively low correlation for School 4 may have stemmed from guessing among the uniformly low scorers in that school.

The PM and the CTMM are stable measures of general ability, even when the interval between testing is as long as a year. The PM and the CTMM have substantial overlap, but this overlap between the non-verbal PM and the more verbal CTMM is not as great as the overlap between the two verbal measures.

The combined pattern of scores from verbal and non-verbal measures contributed to a clearer picture of the general ability possessed by the subjects in the four schools than either test used singly, despite evidence of overlap. Further study that seeks to explore the relationship between classroom learning experiences and performance on the PM and CTMM, as well as other measures of general ability, is needed.

References

Eysenck, H. J. "The effect of incentives on neurotics and the variability of neurotics as compared with normals," British Journal of Medical Psychology, 1944, Vol. 20, pp. 100-103.

Tully, G. E., and Hall, B. W. "Test-retest reliability of the School and College Test." Journal of Educational Measurement, 1965, Vol. 2, No. 1, p. 129.