A VERBAL THINKING TEST FOR HIGH SCHOOL STUDENTS

by

MARY VIVIEN NICKELS

B. S., Kansas State College
of Agriculture and Applied Science, 1931

A THESIS

submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

KANSAS STATE COLLEGE
OF AGRICULTURE AND APPLIED SCIENCE

1932

TABLE OF CONTENTS

INTRODUCTION

This study has been undertaken to develop an inexpensive and short test for high school students which measures ability to think about fairly familiar materials.

The principle of the test used was first developed by J. C. Peterson in his Verbal Thinking Test (for college students). This test consists of sixty groups of four familiar words, each group constructed so that it contains one word which includes at least one meaning of each of the other three words. The problem is to find that word in each group.

The Verbal Thinking Test is only partially a vocabulary test. The words used are supposedly familiar, and emphasis is placed on the fact that one word may have more than one meaning.

In order to utilize the principle of Dr. Peterson's test to measure high school students, the number of words in each group was reduced to three, and allowance was made for the difference in vocabulary between high school and college students. The problem is to select the word in each group which includes a meaning of each of the other two words.

Each group constitutes a test item.

MATERIALS AND PROCEDURE

There were constructed, originally, 192 items, examples and arrangement of which are shown below.

1	2	3
box	trio	season
cuff	trinity	spring
chest	three	bound

The plan was to select the 100 best items and combine them into one test. In order to do this, the original items were divided by chance into three tests of 64 items each. These tests (called Forms A, B, and C merely for identification) were given to the 172 high school students of Onaga and Sharon Springs, Kansas. Of these 172 students, 54 took Form A, 60 took Form B, and 58 took Form C. Precautions were taken to insure equal opportunities, including giving the tests to all students of one high school at the same time, identical direction sheets for all supervisors, and allowing the same time, thirty minutes, in each case. Since these tests were not speed tests, the time allowed was sufficient for every student to complete the test. It was found that 20 minutes was sufficient time to complete the 64 item test.

Forms A, B, and C were scored for the number of correct responses. The scores were then distributed and percentile ranks calculated. The highest 25 per cent and the lowest 25 per cent of the papers were used to determine the value of

the items. The contrast between these groups of papers makes the differentiation between the high paper and the low paper more clear cut than would the contrast between the highest 50 per cent and the lowest 50 per cent. The plan followed in determining the value of the items was as follows. Each student's response to each item was tabulated, so as to show the number of correct and incorrect responses to each item. The criterion used to select the best items was "goodness", defined as the extent to which the item measured the difference between the high paper and the low paper. Two constants were calculated for each item. The first one was the number of correct responses divided by the number of incorrect responses, among the highest 25 per cent of the papers. The second one was the number of correct responses divided by the number of incorrect responses among the lowest 25 per cent of the papers. The ratio representing the "goodness" of an item was the first constant divided by the second constant, or the ratio showing the extent to which the item differentiated between the high paper and the low paper. The items having the largest ratios were selected as the best. Among the 100 best items these ratios varied from infinity (items which no student in the lowest 25 per cent answered correctly) to 16.566.

The 100 best items as selected by "goodness" were then arranged in order of increasing difficulty, the difficulty

of an item being defined numerically as the ratio of the number of correct responses among the highest 25 per cent and the lowest 25 per cent of the papers, to the total number of responses.

These 100 items were combined into one test which was given to 1150 students, 371 in West Junior High School, 291 in East Junior High School, and 488 in the Senior High School, of Parsons, Kansas. Opportunities were again equalized as far as possible by having all tests in Junior High Schools given during the same period, and all tests in Senior High School given during the same period. Supervisors were given direction sheets and asked to follow them absolutely; this was evidently done, with two exceptions to be noted later.

Three scores were recorded for each paper; namely, (1) the number of correct responses among the odd-numbered items, (2) the number of correct responses among the even-numbered items, and (3) the total number of correct responses.

The reliabilities were calculated from the first two scores by the odd-even method, keeping the Junior High Schools separate from the Senior High School. The reliabilities were stepped-up to the length of the test by the Spearman-Brown prophecy formula (3). The Spearman-Brown formula is given below.

$$r_{nn} = \frac{n\, r_{11}}{1 + (n-1)\, r_{11}},$$

in which r_{nn} is the stepped-up reliability, n is the number of times it is stepped-up, and r_{11} is the original reliability coefficient.

A percentile distribution was made of the scores in each class; percentile ranks were found and converted into percentile scores.

Mathematics and English grades were secured for the students, and validity coefficients obtained by correlating Verbal Thinking scores with these grades, separately and then combined. Four grades were recorded for each student-first semester Mathematics, second semester Mathematics, first semester English, and second semester English.

Jack Dunlap and Edward Cureton (1) have developed a formula for the correlation coefficient corrected for attenuation in the criterion, with its standard error. They have found that between a test Y_2 and a criterion X_∞ measured by two fallible scores X_1 and X_2, the correlation is found by the formula

$$r_{\infty 2} = \frac{r_{2(1+3)}}{\sqrt{\frac{2r_{13}}{1 + r_{13}}}}, \text{ in which } r_{2(1+3)} = \frac{r_{12} + r_{23}}{\sqrt{2 + 2r_{13}}}$$

The standard error of $r_{\infty 2}$ is;

$$\sigma_{r_{\infty 2}} = \frac{r_{\infty 2}}{\sqrt{2N}} \left[\frac{2(1-r^2_{2(1+3)})^2}{r^2_{2(1+3)}} + \frac{(1-r_{13})^2}{2r^2_{13}} - \frac{(1-r^2_{2(1+3)})\ (1-r_{13})}{r_{13}} \right]^{\frac{1}{2}}$$

Using first and second semester grades as the two measures of the criterion, a validity coefficient is secured

which has been corrected for attenuation in the criterion. There were three such coefficients found for each class, the correlation between Verbal Thinking and Mathematics, the correlation between Verbal Thinking and English, and the correlation between Verbal Thinking and combined Mathematics and English.

A comparison between the validity of boys' and girls' scores was made. These coefficients were secured by the method described above.

Terman Group Test Intelligence Quotients were available for most of the students in East Junior High School. A study was made in each class of the extent to which these quotients correlated with grades and with Verbal Thinking scores. Correlations were secured between Verbal Thinking scores and grades for the students who had Terman scores. Grades were then correlated with the Terman and Verbal Thinking scores by means of multiple correlations, as given by Kelly's (5) formula for finding multiple correlations when three variables are involved. The formula is as follows;

$$R_{1(23)} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}},$$

in which X_1 is the criterion, X_2 is Terman scores, and X_3 is Verbal Thinking Scores.

Grade norms were found for each grade, and age norms for the years 12 to 17. Each student's age was taken as that of his nearest birthday.

All of the errors calculated for coefficients in this study are standard errors.

RESULTS AND DISCUSSION

The reliability of the Verbal Thinking Test was found to be .961 ± .002 for the Senior High School, and .938 ± .003 for the Junior High Schools.

The validity coefficients are shown in tables 1 and 2. These coefficients are higher when the criterion used is the combined Mathematics and English grades than they are for either subject separately. The validity is found to vary from .378 ± .063 for the seventh grade to .520 ± .043 for the ninth grade. These coefficients could reasonably be expected to be higher for the composite of all grades. Freeman (1) found that the correlation between standardized tests and composite standings of the pupils could be said to lie usually between .40 and .60.

A comparison of the validity of the scores made by boys with those made by girls is shown in table 3. With one exception, the girls' scores show markedly higher correlations.

The comparison of Terman Intelligence Quotients and Verbal Thinking scores, as shown in table 4, must necessarily

Table 2. Validity Coefficients -- Senior High School

	Sophomore	Junior	Senior
Verbal Thinking and 1Math	.394 ± .070	.425 ± .074	.288 ± .075
Verbal Thinking and 2Math	.446 ± .066	.298 ± .083	.314 ± .074
1Math and 2Math	.855 ± .022	.827 ± .028	.761 ± .034
*Verbal Thinking and Math	.454 ± .070	.397 ± .081	.347 ± .068
Verbal Thinking and 1Eng	.440 ± .067	.421 ± .075	.470 ± .046
Verbal Thinking and 2Eng	.347 ± .073	.399 ± .076	.354 ± .072
1Eng and 2Eng	.756 ± .035	.796 ± .033	.777 ± .032
*Verbal Thinking and Eng	.451 ± .072	.459 ± .077	.467 ± .060
Verbal Thinking and (1Math + 1Eng)	.462 ± .065	.465 ± .071	.436 ± .067
Verbal Thinking and (2Math + 2Eng)	.450 ± .066	.393 ± .077	.376 ± .071
(1Math + 1Eng) and (2Math + 2Eng)	.866 ± .020	.884 ± .020	.846 ± .023
*Verbal Thinking and (Math + Eng)	.490 ± .064	.456 ± .074	.440 ± .068

*Coefficient corrected for attenuation in the criterion.

Table 3. Validity Coefficients -- Comparison of Boys and Girls

		V.T.* and (1Math+1Eng)	V.T. and (2Math+2Eng)	(1Math+1Eng) and (2Math+2Eng)	V.T. and (Math+Eng)**
Seventh	-All students	.345 ± .062	.354 ± .061	.854 ± .019	.378 ± .068
	Girls	.356 ± .085	.353 ± .085	.821 ± .031	.390 ± .089
	Boys	.264 ± .094	.295 ± .092	.843 ± .029	.304 ± .093
Eighth	-All students	.460 ± .061	.358 ± .068	.836 ± .023	.446 ± .066
	Girls	.440 ± .083	.349 ± .091	.831 ± .032	.432 ± .070
	Boys	.418 ± .097	.381 ± .101	.815 ± .039	.442 ± .101
Ninth	-All students	.463 ± .051	.498 ± .049	.852 ± .018	.520 ± .043
	Girls	.593 ± .057	.584 ± .058	.869 ± .021	.632 ± .055
	Boys	.350 ± .086	.422 ± .081	.824 ± .031	.423 ± .082
Sophomore	-All students	.462 ± .065	.450 ± .066	.866 ± .020	.490 ± .064
	Girls	.579 ± .078	.472 ± .092	.876 ± .027	.561 ± .085
	Boys	.400 ± .098	.374 ± .101	.840 ± .034	.421 ± .102
Junior	-All students	.465 ± .071	.393 ± .077	.884 ± .020	.456 ± .074
	Girls	.476 ± .095	.438 ± .100	.909 ± .021	.478 ± .095
	Boys	.462 ± .107	.345 ± .119	.816 ± .045	.446 ± .114
Senior	-All students	.436 ± .067	.376 ± .071	.846 ± .023	.440 ± .068
	Girls	.588 ± .079	.516 ± .089	.814 ± .041	.611 ± .082
	Boys	.363 ± .097	.305 ± .102	.826 ± .035	.364 ± .099

*Verbal Thinking

**Coefficient corrected for attenuation in the criterion.

Table 4. Comparison of Terman Intelligence Quotients and Verbal Thinking Scores, for East Junior High School

	Seventh Grade	Eighth Grade	Ninth Grade
Terman I.Q. and (1Math + 1Eng)	.586 ± .076	.707 ± .059	.616 ± .088
Terman I.Q. and (2Math + 2Eng)	.593 ± .075	.664 ± .066	.630 ± .086
(1Math + 1Eng) and (2Math + 2Eng)	.901 ± .022	.809 ± .041	.873 ± .033
*Terman I.Q. and (Math + Eng)	.620 ± .074	.761 ± .057	.666 ± .084
Verbal Thinking and (1Math + 1Eng)	.445 ± .093	.373 ± .102	.482 ± .109
Verbal Thinking and (2Math + 2Eng)	.436 ± .094	.312 ± .107	.564 ± .097
(1Math + 1Eng) and (2Math + 2Eng)	.901 ± .022	.809 ± .041	.873 ± .033
*Verbal Thinking and (Math + Eng)	.464 ± .094	.391 ± .100	.569 ± .010
Terman I.Q. and Verbal Thinking	.473 ± .089	.460 ± .093	.554 ± .094
(Math + Eng) and (Terman I.Q., Verbal Thinking)	.649	.762	.704

*Coefficient corrected for attenuation in the criterion.

be limited, for the numbers were small. There were 76 in the seventh, 71 in the eighth, and 54 in the ninth, grade.

The coefficients show that the Terman Intelligence Quotients correlate with grades to a much greater extent than do Verbal Thinking scores. It must be remembered, however, that the Terman scores were available before the grades were given and therefore the relationship may be to some extent causal. That fact may account for the high validity shown by the Terman Group Test for those students. Verbal Thinking scores were not available until after the grades were given.

The correlations between Terman scores and Verbal Thinking scores vary from .473 $\pm$ .089 to .554 $\pm$ .094. Therefore, the tests do not measure all of the same abilities. That they supplement each other is shown by the multiple correlations between grades and the combined Terman and Verbal Thinking scores, the multiples being higher than the validity coefficients for either test.

The reliability of the Terman Group Test is given by Kelley (4) as .89 for the ninth grade, to compare with a reliability of .938 for the Verbal Thinking Test in the Junior High School.

The norms, with the number of students they were based upon are shown in table 5.

Table 5. Age and Grade Norms

Age	N	Norm	:	Grade	N	Norm
12	72	43.51	:	Seventh	219	40.10
13	111	45.01	:	Eighth	195	45.45
14	197	47.34	:	Ninth	248	48.58
15	190	48.93	:	Sophomore	183	53.88
16	113	49.94	:	Junior	150	51.24
17	67	52.10	:	Senior	155	59.94
			:			

These norms show a continual increase for all ages, and for all grades except the Junior. In this case, exceptions from the prescribed procedure in the administration of the test were found which might explain the lower average. The exceptions were failures to explain the examples given on the direction sheet.

CONCLUSIONS

The Verbal Thinking Test for High School Students was designed originally for the use of Senior High School students. A comparison shows that the reliability and validity of the test are higher for Senior High School than for Junior High School.

The high reliability and the progressively higher norms indicate that the Verbal Thinking Test possesses the essentials of a good test. The validity coefficients are no lower than those usually found. The grades used as a criterion were not normally distributed, as evidenced by the

fact that in some cases there were as many as 29 A grades where there were only two failures. Such faulty distributions make grades a very unsatisfactory criterion.

The Verbal Thinking Test is not intended to be a general intelligence test. It is intended to be a test of ability to think. It is very possible that this ability is not included in the awarding of grades.

This study has shown that it will be worth while to validate the Verbal Thinking Test on the basis of such a criterion as pooled judgments of that ability which it is designed to measure. This will necessitate selection of competent judges and the formulation of a basis for measuring students as to their ability to think.

The Verbal Thinking Test will again be revised by means of the data secured from the Parsons students. The revised test will consist of approximately 60 items, and the time allowed will be 20 minutes. The test will then be standardized on high school students.

ACKNOWLEDGMENTS

I am indebted to Dr. J. C. Peterson for his encouragement and able supervision. I also wish to thank the principals of the Onaga, Sharon Springs, and Parsons schools for their fine cooperation.

LITERATURE CITED

(1) Dunlap, J., and Cureton, E. E.
The Correlation Corrected for Attenuation and Its' Standard Error. American Journal of Psychology 42: 405-407. July 1930.

(2) Freeman, F. N.
Mental Tests. Houghton Mifflin Co., 1926. P. 372.

(3) Holzinger, K. J.
Statistical Methods for Students in Education. Ginn and Co., 1928. P. 169.

(4) Kelley, T. E.
Interpretation of Educational Measurements. World Book Co., 1927. P. 299.

(5) Kelly, E. L.
The Relationship Between the Techniques of Partial Correlations and Path Coefficients. Journal of Educational Psychology 20: 119. February 1929.