**Proceedings of the Iowa Academy of Science**

Volume 56 | Annual Issue Article 43

1949

# Reliability and Relationship of Certain Measures of Reading Ability

David K. Trites
*Iowa State College*

# Reliability and Relationship of Certain Measures of Reading Ability

By David K. Trites

## INTRODUCTION

A number of studies (1) have been published concerning various methods of measuring reading rate and comprehension. The ophthalmograph has been described (2) as one of the more objective, semi-laboratory adaptations of these methods. The ophthalmograph is a portable, binocular eye-movement camera. The reading graph, or eye-movement photograph, obtained while the subject is reading a standard selection, provides objective evidence concerning reading speed, fixations, regressions, and other eye movements. By testing the subject on the material read a measure of his comprehension is obtained.

## PROBLEM

Some doubt has arisen concerning the reliability of the results secured through the use of the instrument. Stone (3), in a study of 64 college freshmen, obtained results indicating that a selection of at least 300 words should be used, and that to secure an average record of the reading ability of a college student the reading graph of the second or third 100 words should be taken. Since the standard card for use with the machine is only 50 words in length, it was felt that this warranted further investigation with different types of students and in a different locale. Therefore, this study was undertaken to determine the reliability coefficients of two consecutive ophthalmograph records and the correlations between each of the two sets of records and the results obtained on the Iowa State College Silent Reading Test and the Diagnostic Tests, (4) Survey Section, Form A.

## PROCEDURE

Eye movement photographs were obtained for all students enrolled in the Remedial Reading Classes at Iowa State College during the fall and winter quarters of the 1948-1949 academic year. There were 54 subjects in the group studied, some records being discarded because of defective film. The majority of these subjects were classified as freshmen and sophomores with a few juniors and seniors and one graduate student. In each case two records were made. The subject read one of the standard 50-word cards while

being photographed and was then given the standard test of ten true-false questions over the material read. A second record, using a second standard card of equivalent length and difficulty to the first, was made immediately after the first test. The test covering the material in the second card was then administered.

Each record was analyzed for reading speed in words per minute, percent comprehension, fixations for each 50 words, and regressions for each 50 words. Correlations were obtained between these four factors on the two sets of records.

Results on the Iowa State College Silent Reading Test were available for 44 of the 54 subjects. For the last three years all undergraduate students entering Iowa State College have taken this test as a regular part of the college entrance test program. The test yields two scores which may be converted to reading rate in words per minute and percent comprehension. The test of comprehension consists of 35 true-false questions.

Correlations between the speed and comprehension on each set of ophthalmograph records and the scores on the ISC Silent Reading Test were calculated.

Records on the Diagnostic Reading Test were available for 31 of the original 54 subjects. This test was given immediately after the subject had been photographed with the ophthalmograph.

In addition to these 31 subjects with double ophthalmograph records, five subjects were available who had only one ophthalmograph record and who had taken the Diagnostic Test. Hence, in analyzing the relationship between the first ophthalmograph records and the Diagnostic Test, 36 subjects were used, whereas with the second record, only 31 subjects were used.

The Diagnostic Test yields, in addition to other indices, scores which may be changed to speed of reading and a percent comprehension. The average reliability of these scores, as reported in the test manual (4) are: Rate of Reading, $r=.80$; Story Comprehension, $r=.72$. The Test of General Reading Comprehension is composed of 20 multiple choice questions.

## RESULTS AND DISCUSSION

The means, fiducial limits, and standard deviations of all the groups and all the tests are presented in Table 1. It will be noted that the sub-groups within each test have comparable means and that the fiducial limits overlap extensively. Therefore, from inspection, it was concluded that despite the varying number of cases comprising the sub-groups, the sub-groups within a single test were from

Table 1. Means, Fiducial Limits,[1] and Standard Deviations of Ophthalmograph, I.S.C., and Diagnostic Tests.

| | | Speed | | Comprehension | |
|---|---|---|---|---|---|
| | | Mean | Std. Dev. | Mean | Std. Dev. |
| Opht. 1 | N=54 | 282.5±22.01 | 82.5 | 64.8±4.96 | 18.6 |
| | N=44 | 274.9±22.48 | 76.1 | 67.7±4.88 | 16.5 |
| | N=36 | 264.7±22.29 | 68.2 | 63.9±5.72 | 17.5 |
| Opht. 2 | N=54 | 283.1±17.74 | 66.5 | 82.8±3.25 | 12.2 |
| | N=44 | 278.6±18.09 | 61.2 | 84.1±3.31 | 11.2 |
| | N=31 | 268.8±20.91 | 59.4 | 82.3±4.26 | 12.1 |
| ISC. | N=44 | 201.0±11.89 | 40.2 | 72.7±3.19 | 11.2 |
| Diag. | N=36 | 220.6±15.65 | 47.9 | 77.5±4.05 | 12.4 |
| | N=31 | 221.4±17.14 | 48.7 | 77.1±4.40 | 12.5 |

[1]The fiducial limits were computed at the 5 percent level by multiplying the standard error of the mean by 1.96

the same population. This similarity was particularly striking among the comprehension means.

To determine the significance of the differences between the ophthalmograph means and corresponding means on the other tests, the t-test of significance was used. The t-values are presented in Table 2. All but three of the t-values computed were highly sig-

Table 2. t-values for differences between means.

| | Speed | | | Comprehension | | |
|---|---|---|---|---|---|---|
| | Opht. 1 | ISC | Diag. | Opht. 1 | ISC | Diag. |
| Opht. 2 | .0697 | 11.8113**[a] | 3.8537** | 8.3830** | 5.4286 | 1.8978 |
| ISC | 7.0381** | | | 1.5674 | | |
| Diag. | 4.0091** | | | 3.9420** | | |

[a]The ** indicates significance at the 1 percent level of confidence.

nificant. The two ophthalmograph records did not show a significant difference in speed of reading; there was not a significant difference between the first ophthalmograph records and the Iowa State Test on percent comprehension; and the difference between the second ophthalmograph records and the Diagnostic Test was not significant. The latter, however, almost reached significance at the 5 percent level.

On the basis of these results, it seemed fairly conclusive that, with the exception of the three cases mentioned previously, the scores on both sets of ophthalmograph records were not similar to scores on the Iowa State Test and the Diagnostic Test.

The three means which apparently did not have a significant difference were also of interest. This apparent similarity in the groups was not borne out by the following correlations which are presented below.

The correlations between reading speeds on the two groups of ophthalmograph records and reading speeds on the Iowa State Test and the Diagnostic Test are presented in Table 3. Correlations between percent comprehension on the ophthalmograph records and

Table 3. Correlations between Reading Speeds on Ophthalmograph, ISC, and Diagnostic Tests.

| | Ophthal. 2 | ISC | Diagnostic | |
|---|---|---|---|---|
| | N=54 | N=44 | N=36 | N=31 |
| Ophthal. 1 | .658** | .423** | .392* | |
| Ophthal. 2 | | .609** | | .215 |

** Indicates a significant difference from zero at the 1 percent confidence level.
* Indicates a significant difference from zero at the 5 percent confidence level.

the other instruments are given in Table 4. All correlations were tested for significance.

The greatest r among reading speeds is that between the first and second ophthalmograph records, r=.658. This is not as great as would be expected from the non-significance of the difference between the means of the two groups.

Reading speed on the Iowa State Test was correlated more highly with the speed on the second ophthalmograph record, r=.609, than with speed on the first ophthalmograph record, r=.423. The reverse was true of the Diagnostic Test. The Diagnostic Test was correlated more highly with the speed on the first ophthalmograph record, r=.392, than with the speed of the second, r=.215. All correlations except the last were found to be significant.

Table 4. Correlations between Percent Comprehension on Ophthalmograph and ISC and Diagnostic Tests.

| | Ophthal. 2 | ISC | Diagnostic | |
|---|---|---|---|---|
| | N=54 | N=44 | N=36 | N=31 |
| Ophthal. 1 | .143[a] | —.136 | .070 | |
| Ophthal. 2 | | .224 | | .229 |

[a] None of the correlation coefficients is significantly different from zero.

The correlations in Table 4 between percent comprehension on the various tests are all of low magnitude. A negative correlation appeared between percent comprehension on the Iowa State Test and comprehension on the first ophthalmograph records r=—.136. However, none of the correlations was significantly different from zero.

Thus the evidence indicated that when the standard 50-word card was used with the ophthalmograph, the speed scores on the records could be considered as having a reliability sufficient to permit only

group differentiation. Furthermore, the ophthalmograph speed score may be considered as being only moderately indicative of a subject's score on the Iowa State Reading Test and the Diagnostic Test. The ophthalmograph comprehension score apparently has little reliability and little relationship to comprehension scores on the Iowa State and Diagnostic Tests. To the extent that the Iowa State and Diagnostic Tests are representative of all paper and pencil tests of reading ability, these conclusions may be generalized to all such measures of reading skill.

In addition to the preceding analysis, the reliability of the number of regressions and fixations on the ophthalmograph records per 50 words was investigated. The means, standard deviations, and correlations between the first and second sets of records are ar-

Table 5. Means and Standard Deviations of Ophthalmograph Regressions and Fixations.

| | Ophthal. 1 | | Ophthal. 2 | |
|---|---|---|---|---|
| | Fix. | Reg. | Fix. | Reg. |
| Mean | 35.0 | 8.39 | 34.8 | 9.17 |
| Standard Dev. | 7.24 | 4.62 | 8.93 | 5.06 |

$r_{fix}$=.639** $r_{reg}$=.544** Both r's are significant at the 1 percent confidence level.

ranged in Table 5. Since the determination of the number of regressions and fixations for a definite number of words is one of the more unique and useful features of the ophthalmograph, these correlations are of especial interest. According to Mursell (5) a test with a reliability coefficient above .5 but under .94 would be satisfactory for rough group differentiation but inadequate for individual classification.

It was hoped that this study would provide some evidence concerning the optimal length of reading selections for use with the eye-movement camera. If the correlation of the alternate forms of the reading cards used with the instrument can be considered as being equivalent to a split half reliability coefficient with regard to difficulty and if the errors of measurement are not correlated, then the Spearman-Brown prophecy formula may be used to determine the length of the reading selection necessary to achieve a reliability coefficient of .94. Using the computed reliability of .658, the Spearman-Brown formula indicated that the reading selection must be approximately eight times its present length to achieve the reliability of .94. This would mean that the selection would have to be approximately 400 words in length.

Unfortunately the assumption that the alternate forms of the

reading cards are comparable in difficulty has not been verified. Nor does the assumption that the errors of measurement were not correlated seem warranted. In obtaining the initial record on the eye-movement camera the period of time spent in adjusting the instrument was not constant. In all probability the length of time that the subject sat before the machine while it was being adjusted was correlated with the results obtained. Usually the second record was obtained in a much shorter time because the apparatus had already been adjusted for the subject. This error would probably have been reduced and the correlation between the consecutive ophthalmograph records increased if one card had been used for practice and two other cards administered in rapid succession to obtain the photographs used in the correlation. This would have had the effect of reducing the length of the record necessary for the .94 reliability coefficient as computed by the Spearman-Brown formula.

The correlations of the sets of ophthalmograph records with the two paper and pencil tests of reading ability provided no clear-cut results regarding the optimal length of the reading selection. The evidence indicated that the present 50-word card used with the camera did not provide reliability sufficient for individual classification or prediction. Apparently the speed and comprehension scores obtained through the use of the ophthalmograph are of a different order than corresponding scores secured from paper and pencil tests. If the paper and pencil test scores can be considered as yielding a truer index of a student's reading speed than does the ophthalmograph, the difference appeared to be an overestimation of this ability when the ophthalmograph was used. Making the same assumption with respect to comprehension scores on the paper and pencil tests, and judging by the means, the first ophthalmograph record seemed to underestimate the student's ability, whereas the second ophthalmograph record overestimated this ability. However, the correlation coefficients between the ophthalmograph and the other tests indicated that no significant relationship existed between these scores.

With respect to the reliability of the various measures of reading ability obtained through the use of the ophthalmograph, the results were fairly conclusive. However, further investigation is needed to substantiate these findings, and to provide more definite evidence regarding the optimal length of the reading selection for use with the machine.

## SUMMARY AND CONCLUSIONS

This investigation was undertaken to determine the reliability of speed of reading, comprehension, number of fixations, and number of regressions as measured by the ophthalmograph, or eye-movement camera. Evidence was sought concerning the most satisfactory and reliable length of the reading selection for use with the machine. The relationship between measures of reading ability obtained from paper and pencil tests and those obtained by the use of the eye-movement camera was investigated. Fifty-four college students provided data for the study.

On the basis of the evidence obtained from this investigation, the principal conclusions were:

(1) Speed of reading as measured by the ophthalmograph using the standard 50-word card did not have a comparable-form correlation coefficient high enough to warrant individual classification or prediction, r=.658. It was considered to be sufficient for group prediction.

(2) Reading comprehension as measured by the ophthalmograph had a comparable-form correlation which was not significantly different from zero, r=.143. This would seem to indicate little, if any, reliability for this measure as might be expected from the form and length of the test.

(3) The reliability of the number of regressions and fixations per 50 words, as measured by the ophthalmograph, was not great enough to be useful in individual prediction or classification, but was sufficient for group differentiation. The reliability coefficient for fixations was .639; for regressions it was .544.

(4) Speed of reading obtained from the eye-movement photographs was not highly correlated with speed scores obtained from the Iowa State Silent Reading Test and the Diagnostic Reading Test. The correlations ranged from .215 between the Diagnostic Test and the second set of ophthalmograph records to .609 between the Iowa State Test and the second set of ophthalmograph records. This would seem to indicate that the photographic test and paper and pencil tests are not determining the same aspects of reading ability.

(5) Reading comprehension as determined from ophthalmograph records did not have correlations significantly different from zero with comprehension scores in the Iowa State Reading Test and the Diagnostic Test.

(6) The evidence regarding the optimal length of the ophthal-

mograph reading selection, required in order to obtain a reliable measure of reading speed, was not conclusive. Further research is needed to clarify this point.

**References**

1. Porter, R. B., Shafer, H. & Monroe, E. 1946. Research in Reading During the War Years. *Language and the Fine Arts. Rev. Ed. Res.* XVI, 102-115.
Tinker, Miles A. 1946. The Study of Eye Movements in Reading. *Psych. Bull.* XLIII, 93-120.
2. 1937. The Ophthalm-O-Graph, The Metron-O-Scope, Manual for Controlled Reading. Edited by The Bureau of Visual Science, American Optical Company.
3. Stone, Lewis G. 1942. Reading Reactions for Varied Types of Subject Matter; An Analytical Study of the Eye-movements of College Freshmen. *J. Exper. Ed.* 10, No. 1.
4. Diagnostic Reading Tests, Survey Section, Form A. 1947. Committee on Diagnostic Reading Tests. Distrib. by Educ. Records Bur., New York, N.Y.
5. Mursell, James L. 1947. Psychological Testing. 1st ed. p. 51. Longmans, Green and Co., Inc. New York, London, Toronto.

DEPARTMENT OF PSYCHOLOGY
IOWA STATE COLLEGE