

# A PSYCHOLINGUISTIC APPROACH TO THE MEASUREMENT OF READING MATERIALS COMPREHENSION

Paper read at the Third Annual Congress of the Linguistic Society of Papua–New Guinea

Jonathon Anderson  
University of New England

The contribution of psycholinguistics to the problem of reading comprehension measurement is examined and a relatively new technique for measuring reading comprehension, cloze procedure, is described within the framework of psycholinguistics. The paper presents the main findings of a study carried out in Papua–New Guinea in which a test of reading comprehension was constructed based on psycholinguistic theory and using cloze procedure.

It is probably true to say that no subject in the curriculum has received as much attention from researchers as reading. This holds for all levels of education, primary through to tertiary. It has been estimated, for example, that in the United States alone more than 500 studies into reading are conducted each year. What are the results of all this research activity into reading? On the measurement side, one result has been the production of a great number of standardized instruments designed to measure reading or reading comprehension (the two terms are often used synonymously).

In many ways the profusion of reading tests has posed a real difficulty for practising teachers. Which reading test should be used? This question gives rise to the more general question: what do reading tests measure? There is need for careful evaluation. Giving a test a name; for example, calling it a test of reading comprehension or a test of ability to see relationships, is not evidence that it measures these skills. This applies to sub-scores as well as to the test as a whole. Another weakness of many reading tests is that there is lacking any firm theoretical basis.

David (1944a) pointed to the difficulty faced by test-constructors in the field of reading, namely that there was no widespread agreement as to exactly what reading ability was. A quarter of a century later there is still no widespread agreement as to the nature of reading ability. Kingston (1961) likened the situation to some of the difficulties formerly faced in describing and measuring intelligence. Adopting a similar solution to that adopted by some psychologists, reading ability could perhaps be described as what is

measured by a particular reading test.

Although researchers have not reached exact agreement as to what reading ability is, there is fairly general agreement that reading consists of a collection of different skills. The disagreement has been in that some authorities have postulated a large number of skills while other authorities have preferred to define reading in terms of a relatively small number of factors. For example, Gray (1937) concluded that the process of reading involved "...such habits as accuracy in recognizing the words that make up a passage, span of recognition, rate at which words and phrases are recognised, rhythmical progress of perceptions along the lines, and accurate return sweep of the eye from the end of one line to the next." Other authorities (e.g. Strang, McCullough, and Traxler, 1967) have put forward similar lists of skills but Burkhart (1945) probably took the prize when he postulated 214 separate abilities.

In contrast to the experts who have postulated many skills, factor analysis has been used by others to determine a few underlying factors. There has been a great number of factor analytic studies of reading ability and just some of these at the tertiary, secondary, and primary level are listed here. Langsam (1941) using college students identified five factors among a battery of reading tests - verbal, perceptual, word, number, and relationships. Davis (1944b) adopted a somewhat different approach from previous studies in that he first identified the comprehension skills agreed upon by most reading experts, then he constructed tests to measure these skills, and finally he subjected the test results to factor analysis. He isolated nine factors including knowledge of word meanings, ability to follow the organization of a passage, ability to identify the main theme, ability to draw inferences, and knowledge of stated facts. At the high school level Anderson (1949) identified three factors: vocabulary, analysis - synthesis or reasoning, and intelligence. At the primary level, studies by Gans (1940) and Jay (1950) revealed generally similar findings to the above.

However, despite such refined statistical techniques as factor analysis, there has been no dramatic break-through in the measurement of reading in the past forty to fifty years. Kingston (1961) makes the point that in the same way that reading cannot be defined by listing its observable attributes, nor can it be explained by using complex and abstract terms. For progress to be made in the field, he affirms, "...bold new conceptual frameworks and theoretical designs are required."

The purpose of the present paper is, first, to examine briefly the contribution of psycholinguistics to the problem of reading comprehension measurement, second, to

meaning (Fries, 1952).

Sequential constraint refers to the predictability of elements in a message by virtue of their statistical characteristics. Taking words as the elements of a message, certain sequences are more probable than others in a given situational context. 'For breakfast we had bacon and \_\_\_\_\_' is more likely to be completed in most western parts of the world by 'eggs' than by 'porridge' or by 'yams'.

In encoding the writer is restricted to some extent in his word choices by his particular language habits. Similarly, in decoding, the reader anticipates certain patterns or sequences at various choice points because of his language habits. Osgood (1960) described these dispositional mechanisms and their operation as follows:-

Language habits, like habits in general, appear to be organized into hierarchies of alternatives. At both lexical and structural choice points, to the extent that there is choice, certain alternatives will be most probable, others less probable, and others very improbable. Grammatical and syntactical regularities or redundancies place severe restrictions on structural choices, and frequency-of-usage factors have a similar effect on lexical choices (both of these factors contributing to the norm in any encoding situation), but there is still room at many points for modification of the probability structure of hierarchies as functions of situational and individual variables.

(Osgood, 1960, p.296)

### Cloze Procedure as a Measure of Reading Comprehension

Cloze procedure was defined by Taylor (1953) when he introduced the technique as follows:

A method of intercepting a message from a 'transmitter' (writer or speaker), mutilating its language patterns by deleting parts, and so administering it to 'receivers' (readers or listeners) that their attempts to make the patterns whole again potentially yield a considerable number of cloze units.

(Taylor, 1953, p.416)

If reading comprehension is defined, as it was above, as the correspondence between the semantic and grammatical habit systems of encoders and decoders using the same language it can be seen how cloze procedure indexes this correspondence or communality between the language systems of different individuals. Reading comprehension is thus

defined operationally. A message from an encoder has its language patterns mutilated by some mechanical procedure. Blanks are inserted to denote words that are deleted and subjects then attempt to replace the blanks with the missing words. That is, choice points are created which intercept a reader's sequential progression across the page and the cloze task requires him to make the most likely replacement in the light of his language system and the cues that are available. If the reader's system of language habits corresponds closely to that of the writer, he is generally successful in his attempts at replacement. His cloze score (number of correct replacements) is taken as a measure of the amount of meaning he has obtained from the writer's message or his degree of comprehension.

There is support from linguistics for cloze procedure as a measure of reading comprehension. Fries (1963) identified three layers of language meanings: the layer of meanings carried by the grammatical structures, the layer of meanings carried by the lexical items, and the layer of social-cultural meanings. Fries' first two layers of language meanings correspond approximately to what were distinguished as syntactic and semantic redundancy and the third layer of language meaning is included in sequential constraint. Cloze procedure taps these three layers of language meanings for, as McLeod (1965), stated, to successfully reconstruct a message

requires a familiarity with the grammatical structure of English, an understanding of lexical meaning and, if the passages selected are concerned with a variety of experiences familiar in a given culture, they reflect to some extent 'social-cultural' meaning. (McLeod, 1965 p.3)

There is some theoretical basis, then, for believing that cloze scores index reading comprehension.

#### A Reading Comprehension Test for Papua and New Guinea

##### Origin

Eighteen short, self-contained passages were selected to cover the range of reading difficulty appropriate for the middle and upper levels in primary school. In an attempt to counter Kingston's (1960) criticism that few test writers justify the selection of test materials on theoretical grounds, the criteria for selecting the passages were that they should be from readers or reading materials used in different parts of the world where English is taught as a foreign language, that they should not as far as could be judged deal with concepts unfamiliar to indigenous children in Papua and New Guinea, and thirdly, that they should deal with general topics of a non-technical nature.

### Try-Out Testing

A slightly novel feature of the try-out testing was that the sample used did not consist of primary school subjects for whom the test was ultimately intended but of 'efficient' and 'competent' readers of English as a foreign language. The total intake of Preliminary Year students at the University of Papua and New Guinea (N=122), all First and Second year students attending Goroka Secondary Teachers' College (N=128) and Stage Two students at the Administrative College in Port Moresby (N=46) constituted the try-out sample.

Eight versions of each test passage had cloze procedure applied to them. In the first version the eighth and every subsequent eighth word were deleted; in the second version the ninth and every subsequent eighth word were deleted; and so on. In this way every word in each passage was systematically deleted after allowing a sufficient lead-in to the passage. The eight versions of each passage were administered to randomly selected groups of about 37 subjects each with instructions to replace the missing words. For each passage an estimate of reading difficulty was obtained by averaging the number of correct responses over the eight versions. The form of item-analysis adopted was to find those deletions which competent and efficient readers showed greatest unanimity in their responses.

A previous experiment had established that blanks separated by at least four words could be regarded as statistically independent. The deletions that satisfied the criteria of maximum redundancy and statistical independence were identified within each passage. A previous experiment had also indicated that a test containing about 50 items gave a satisfactory degree of reliability. Accordingly, selection was made among the passages for those which were approximately evenly spaced along the continuum of reading difficulty and which contained the required number of items. In this way nine passages containing a total of 50 items were chosen for the first form of the Cloze Reading Comprehension test.

### Administration of Tests in a Primary "T" School

The first form of the Cloze Reading Comprehension test was administered to a random sample of 150 pupils in Standards 4, 5, and 6 of one Primary "T" school in Port Moresby such that there were 50 pupils from each standard.

Two other tests were also administered to these same children. One was Watts' Reading Comprehension Test and the second, a 'traditional-type' comprehension test made over the same passages as the modified cloze tests. That is, multiple-choice questions were constructed which were designed to measure six comprehension skills most frequently listed in the literature, namely, knowledge of vocabulary used in the passages, knowledge

of stated facts, ability to perceive sequence of events, to see relationships, to identify the main theme, and to make inferences. These comprehension tests had previously been pre-tested on a sample of 90 children in Standards 4 to 6 in one other Port Moresby Primary "T" school. As a result of this pre-testing certain non-discriminatory items had been deleted as had certain items which contained ambiguities. The final comprehension test contained a total of 36 vocabulary items, 18 fact items, and nine items for each of sequences, relationships, main theme, and inferences - an equal number of each type of item from each passage.

The Cloze Reading Comprehension and Watts' tests were administered on the one occasion and one week later the Comprehension tests. For the latter, pupils were presented with the passage and were then asked to answer questions about what they had read without referring back to the passage. With the exception of Watts' test, tests were given under untimed conditions. The Cloze Reading Comprehension test was administered in three parts, A, B, and C. Part A contained passages 1, 5, and 9; Part B contained passages 2, 6, and 7; and Part C contained passages 3, 4, and 8. Reasons for this were to guard against test fatigue, to control order effects, and for the requirements of the validation study to follow.

#### Validation Study

All tests were scored by computer and this ensured marker reliability. For each subject 10 scores were available: three cloze test results, six comprehension measures, and the Watts' comprehension score. In addition there were two composite measures: total cloze score and total comprehension score.

Reliabilities of all measures were calculated using coefficient alpha. This coefficient is generally regarded as providing a good estimate of reliability since a major source of measurement error is due to content sampling. Means, standard deviations, standard errors of the mean, maximum scores and reliability coefficients are presented in Table 1. These data show the cloze tests to be highly reliable. Spearman-Brown estimates indicate that even the comprehension sub-tests would have very satisfactory reliabilities if lengthened. Reliabilities for the three cloze tests combined (i.e. the Cloze Reading Comprehension test) and for the total Comprehension test were both 0.95.

describe a relatively new technique for measuring reading comprehension - cloze procedure, and, third, to present the main findings of a study carried out in Papua and New Guinea in which a test of reading comprehension was constructed based on psycholinguistic theory and using cloze procedure.

#### A Psycholinguistic View of Reading Comprehension.

From a psycholinguistic viewpoint, the reading comprehension process begins with the written language encoded by the writer. The graphic symbols are seen by the reader as sequences of visual stimuli. The message is decoded or interpreted when the reader, utilizing all his relevant past experiences, his previous learning, and his language development takes advantage of the semantic and syntactic cues present in the written language (Goodman, 1966). Reading comprehension, then, is defined as the correspondence between the way in which the message is encoded by the writer and the way in which it is decoded by the reader.

The psycholinguistic view of reading comprehension is largely based on Osgood's (1952, 1953) learning theory of communication. A key concept in this theory is the development of what Osgood calls 'dispositional mechanisms'. These are complex verbal patterns or habits which in the learning of a language become more or less automatic. In acquiring these verbal habits, there are individual differences and there are situational differences. The learning of English as a foreign language in Papua and New Guinea is an example of a situational deviation but even among learners in New Guinea there are variations about the situational norm. Despite individual and situational variations, however, there is some correspondence between the dispositional mechanisms of subjects using the same language, otherwise communication would be impossible.

The development of dispositional mechanisms reflects two characteristics of language: redundancy and sequential constraint.

Redundancy is the excess of rules or syntax in a language that minimises the possibility of errors in reception (Cherry, 1957). Linguists, however, generally distinguish two aspects of redundancy: semantic redundancy or the repetition or association of the meaning of words; and syntactic redundancy or the excess of rules in language conveying the same bits of information. Syntactic redundancy is conveyed in English by a relatively small number of high frequency words (structure words) which have little or no semantic meaning and whose function is largely to mark nouns or verbs or indicate negatives. For signal transmission, meaning is of no importance but for an utterance or a message to be understandable, there must be both structural (or syntactic) meaning and lexical (or semantic)

TABLE I

Means, Standard Deviations, Standard Errors of Means,  
Maximum Scores, and Reliability Coefficients of  
Cloze Tests, Comprehension Tests, and Watts' Test

| <u>Test Measures</u> | <u>Mean</u> | <u>Standard Deviation</u> | <u>S. E. of Mean</u> | <u>Max. Score</u> | <u>Coefficient Alpha</u> |
|----------------------|-------------|---------------------------|----------------------|-------------------|--------------------------|
| Cloze Test A         | 9.17        | 4.41                      | .36                  | 17                | .88                      |
| Cloze Test B         | 9.72        | 4.10                      | .34                  | 16                | .87                      |
| Cloze Test C         | 11.01       | 4.46                      | .37                  | 17                | .89                      |
| Vocabulary           | 18.09       | 10.62                     | .87                  | 36                | .96                      |
| Facts                | 13.93       | 3.08                      | .25                  | 18                | .78                      |
| Organization         | 5.32        | 1.80                      | .15                  | 9                 | .44                      |
| Relationships        | 5.97        | 1.74                      | .14                  | 9                 | .53                      |
| Main Theme           | 4.76        | 1.67                      | .14                  | 9                 | .43                      |
| Inferences           | 4.99        | 1.74                      | .14                  | 9                 | .41                      |
| Watts' Test          | 12.75       | 5.78                      | .47                  | 35                | .90                      |

The Technical Recommendations of the National Education Association (1955) distinguished between four types of validity - content validity, predictive validity, concurrent validity, and construct validity. Predictive validity was not estimated in this study. An attempt to ensure some degree of content validity was made by selecting test passages from reading materials used in different parts of the world where English is taught as a foreign language.

In estimating the concurrent validity of cloze tests, Rankin (1957, 1959) made the important distinction between general reading comprehension and specific reading comprehension. The former is gauged by correlation with a standardized measure of reading comprehension while the latter is gauged by correlation with comprehension tests made over the same passages as the cloze tests. Thus, to obtain a concurrent validity estimate of general reading comprehension, total cloze score was correlated with Watts' score. The obtained correlation was 0.78. Similarly, an estimate of concurrent validity of specific reading comprehension was obtained by correlating total cloze score with total Comprehension score ( $r=0.85$ ). Both validity estimates approach the highest



correlation that the respective test reliabilities permit.

To determine the construct or factor validity of the Cloze Reading Comprehension test intercorrelations between the 10 scores available for each subject were calculated for the total sample. The intercorrelation matrix appears in Table 2. A principal axis analysis was carried out. Since only one eigenvalue greater than unity resulted, the factor matrix was not rotated. Factor loadings and eigenvalues are presented in Table 3. Interpretation of the factor matrix was straight-forward. The three cloze tests were included in a battery of measures purporting to measure reading comprehension. Only one factor emerged and this factor accounted for 65 per cent of the variance. It seems clear that this factor relates to reading comprehension. Furthermore, the three cloze tests have highest loadings on this factor.

TABLE 2  
Intercorrelation Matrix of Cloze Tests, Comprehension  
Tests, and Watts' Test

| Test Measures    | Test Measures |    |    |    |    |    |    |    |    |
|------------------|---------------|----|----|----|----|----|----|----|----|
|                  | 1             | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
| 1. Cloze Test A  | -             |    |    |    |    |    |    |    |    |
| 2. Cloze Test B  | 86            | -  |    |    |    |    |    |    |    |
| 3. Cloze Test C  | 85            | 85 | -  |    |    |    |    |    |    |
| 4. Vocabulary    | 78            | 72 | 76 | -  |    |    |    |    |    |
| 5. Facts         | 72            | 69 | 68 | 66 | -  |    |    |    |    |
| 6. Organization  | 55            | 52 | 47 | 46 | 54 | -  |    |    |    |
| 7. Relationships | 70            | 62 | 58 | 63 | 69 | 55 | -  |    |    |
| 8. Main Theme    | 54            | 51 | 51 | 54 | 54 | 50 | 61 | -  |    |
| 9. Inferences    | 55            | 52 | 46 | 52 | 56 | 38 | 52 | 44 | -  |
| 10. Watts' Test  | 78            | 74 | 71 | 71 | 62 | 48 | 61 | 55 | 56 |

Decimal points have been omitted.

All correlations were significant beyond the 0.01 level.

TABLE 3

Factor Loadings on Cloze Tests, Comprehension Tests,  
And Watts' Test

---

|     | Test Measures | Loadings<br>I |
|-----|---------------|---------------|
| 1.  | Cloze Test A  | .92           |
| 2.  | Cloze Test B  | .88           |
| 3.  | Cloze Test C  | .87           |
| 4.  | Vocabulary    | .85           |
| 5.  | Facts         | .83           |
| 6.  | Organization  | .66           |
| 7.  | Relationships | .81           |
| 8.  | Main Theme    | .70           |
| 9.  | Inferences    | .67           |
| 10. | Watts' Test   | .85           |

Eigenvalue  
6.54

65.43 per cent of variance extracted by Factor I

---

Conclusions

Both the Cloze Reading Comprehension test and the underlying theory are validated simultaneously in this study. Evidence is presented that the Cloze Reading Comprehension test is a highly reliable measuring instrument and, further, that it is a valid measure of specific as well as of general reading comprehension.

## REFERENCES

- Anderson, C. C. 1949. "A Factorial Analysis of Reading", British Journal of Educational Psychology. 19, 220-221.
- Burkhardt, K. H. 1945. "An Analysis of Reading Abilities", Journal of Educational Research. 38, 430-439.
- Cherry, C. 1957. On Human Communication. New York. Wiley.
- Davis, F. B. 1944a. "What Do Reading Tests Really Measure?" English Journal. 33, 180-187.
- Davis, F. B. 1944b. "Fundamental Factors of Comprehension in Reading", Psychometrika. 9, 185-197.
- Fries, C. C. 1952. The Structure of English. New York. Harcourt, Brace.
- Fries, C. C. 1963. Linguistics and Reading. New York. Holt, Rinehart & Winston.
- Gans, Roma. 1940. "A Study of Critical Reading Comprehension in the Intermediate Grades", Contributions to Education, No. 811. New York. Teachers College, Columbia University. 1-135.
- Goodman, K. S. 1966. "A Psycholinguistic View of Reading Comprehension", Fifteenth Yearbook of the National Reading Conference. 15, 188-196.
- Gray, W. S. "The Nature and Types of Reading", Thirty-Sixth Yearbook of the National Society for the Study of Education, Part 1. Bloomington, Illinois. Public School Publishing Co. 23-38.
- Jay, Edith, S. 1950. A Factor Study of Reading Tasks. Unpublished doctoral dissertation. University of Chicago.
- Kingston, A. J., Jr. 1960. "The Measurement of Reading Comprehension", Ninth Yearbook of the National Reading Conference. 9, 88-93.
- Kingston, A. J., Jr. 1961. "A Conceptual Model of Reading Comprehension", Tenth Yearbook of the National Reading Conference. 10, 100-107.
- Langsam, Rosalind, S. 1941. "A Factorial Analysis of Reading Ability", Journal of Experimental Education. 10, 57-63.

- McLeod, J. 1965. Gap Reading Comprehension Test Manual. Melbourne. Heinemann.
- Osgood, C. E. 1952. "The Nature and Measurement of Meaning", Psychological Bulletin. 49, 197-237.
- Osgood, C. E. 1953. Method and Theory in Experimental Psychology. New York. Oxford University Press.
- Osgood, C. E. 1960. "Some Effects of Motivation on Style of Encoding", in (Sebeok, T.A. (ed.). 1960). 293-306.
- Rankin, E. F., Jr. 1957. An Evaluation of the Cloze Procedure as a Technique for Measuring Reading Comprehension. Unpublished doctoral dissertation, University of Michigan.
- Rankin, E. F. Jr. 1959. "The Cloze Procedure - Its Validity and Utility", Eighth Yearbook of the National Reading Conference. 8, 131-144.
- Sebeok, T. A. (ed.). 1960. Style in Language. London. Wiley.
- Strang, Ruth, McCullough, Constance, M., and Traxler, A. E. 1967. The Improvement of Reading. 4th ed. New York. McGraw-Hill.
- Taylor, W. L. 1953. "Cloze Procedure: A New Tool for Measuring Readability", Journalism Quarterly. 30, 415-433.
- Technical Recommendations for Achievement Tests. 1955. Washington. National Education Association.
- Watts, A. F. 1957. Sentence Reading Test I. London National Foundation for Educational Research.