The use of exact-length blanks in cloze tests of reading comprehension was investigated. Subjects were indigenous primary-school pupils in New Guinea for whom English was a foreign language. No significant difference was found between the mean scores of subjects on cloze tests using blanks of a uniform length and the mean scores of subjects on cloze tests using blanks of the same length as the deleted words. Both versions of cloze tests were equally valid as measures of general reading comprehension. The question of length of blank to use in cloze tests has practical as well as theoretical significance for teachers in constructing cloze tests when they wish to use photocopying facilities. The implication of these findings is that by the use of a photocopy procedure, such factors as size of print, illustrative material, and page layout may be included in the estimate of the ease or difficulty of printed material as measured by cloze tests. Tables and references are included. (Author/PR)
A Report of Research on Comprehension in Reading

Section B-1, Saturday, August 8, 1970
2:30 to 3:34

ABSTRACT

The purpose of this study was to investigate the use of exact-length blanks in close tests of reading comprehension. Subjects for the study were indigenous primary school pupils in New Guinea for whom English was a foreign language. No significant difference was found between the mean scores of subjects on close tests using blanks of a uniform length and the mean scores of subjects on close tests using blanks of the same length as the deleted words. Both versions of close tests were equally valid as measures of general reading comprehension. The implication of these findings is that such factors as size of print, illustrative material and page layout may be included in the estimate of the ease or difficulty of printed material as measured by close tests.
The present paper is a report of research on cloze procedure, a technique that researchers, reading specialists and classroom teachers are finding increasingly useful to measure the reading difficulty or readability of printed material.

Cloze procedure has been described by Taylor (11) who introduced it as a psychological tool for measuring the effectiveness of communication. Basically, the procedure consists of a set of rules for constructing cloze tests over samples of written (or spoken) materials, administering these tests to subjects and scoring them, and determining from the cloze scores the degree of comprehension of the materials. The construction of a cloze test simply requires words to be deleted through a passage according to some mechanical system. A subject is asked to replace the missing words and the number of words he correctly replaces is an index of his comprehension of the passage.

There is a sound theoretical basis for cloze scores as measures of reading comprehension. Fries (2) 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. Cloze procedure taps these three layers of language meanings for, as Koled (7) 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. (7, p.3.)

The set of rules for constructing cloze tests is quite well established. Comprehensive reports by Rankin (6, 2) and by Anderson (1)
reviewed studies on close procedure, studies carried out in the main in English speaking parts of the world with native speakers of English. There is, however, a growing body of research showing that close procedure may be more widely applied, for example, to English and to French learned as foreign languages (1, 4), to Japanese (10), and to Korean (12).

Previous studies on close procedure have invariably employed blanks of a standard length to denote words deleted from a passage following Taylor's (11) assumption that length of blank influences "guessing". Thus, Bormuth in his definition of close readability procedure states that "the deleted words are replaced by underlined blank spaces of a uniform length . . ." (2, p.1.) The reasoning behind this assumption is presumably that the length of the blank is a clue providing additional information about the deleted word. However, it is difficult to predict the way in which this clue might operate. If a subject is very familiar with the layer of meanings carried by the grammatical structures, with the layer of meanings carried by the lexical items, and with the layer of social-cultural meanings, this additional clue might be quite unnecessary. On the other hand, to the extent that a subject is not familiar with the various layers of meanings in written language, the additional clue might prove of little assistance. These considerations suggest the operation of at least two important variables: the reading difficulty of the written language and the comprehension ability of subjects.
Purpose

The purpose of the present study was to compare the use in cloze tests of standard or uniform length blanks with blanks of the same length as the word deleted. Comparisons were made using (a) passages varying in difficulty and (b) subjects differing in reading ability. The null hypotheses tested were:

1. There is no difference between the mean score for standard-length blanks and the mean score for exact-length blanks at different levels of passage difficulty and for different reading ability levels.

2. There are no interactions between length of blank, difficulty levels and reading ability levels.

The five per cent level of significance was required for the rejection of the hypotheses.

This study was part of a larger study conducted in New Guinea. The New Guinea setting provided an opportunity to test cloze procedure as a measuring instrument with subjects learning English as a foreign language. Further, it provided an opportunity to test cloze procedure against a multilingual background of great diversity and perplexity for it has been estimated that over five hundred distinct languages are spoken in New Guinea.

Test Measures

Cloze tests were constructed over passages drawn from three children's readers. The three readers were judged difficult, neither difficult nor easy, and easy reading for children in the upper primary levels of
J. Anderson

 Territory schools. There was a lead-in of ten words to each passage and thereafter every eighth word was systematically deleted through the passage. In half the cloze tests missing words were indicated by underlined blanks ten spaces in length (standard-length blanks); in the other half, blanks were the same length as the word deleted (exact-length blanks). Since the exact-length versions of the cloze tests did not allow sufficient space for subjects to print their answers, answer spaces were provided alongside the line in which the deletion appeared. The same procedure was adopted for the standard-length versions so that only length of blank varied. For example,

**Standard-length version**

When the tiger is too old to hunt for food _______ the jungle,

**Exact-length version**

When the tiger is too old to hunt for food __ the jungle,  

To control for initial reading ability, subjects were first ranked according to their scores on Watts' Test (13). This test was shown by Koleod (7) to correlate highly with cloze tests.

**Subjects**

All testing was carried out in one primary school. The predominant local language in the village where the school was situated was Pidgin while in the school the medium of instruction was English. All pupils in
the top three standards \((N=110)\) constituted the sample.

**Experimental Design**

An experimental design that allows the effect of length of blank to be tested at different levels of passage difficulty and for different levels of reading ability is three-dimensional analysis of variance. The advantage of matching subjects on a control variable (the Watts' Test) is that an important source of error is reduced thereby increasing the precision of the experiment.

The design called for two testing sessions. In the first, the Watts' Test was administered to each standard. On the basis of Watts' scores, an order of merit was made for the 110 subjects. High, average, and low reading ability groups were formed by dividing pupils into top quarter, middle half, and bottom quarter respectively. Within each ability group subjects were randomly allotted to standard- or exact-length versions of close tests at one of the three levels of passage difficulty. This reduced \(N\) to 96. To satisfy the condition of proportionality from row to row, high and low ability treatment cells each contained four subjects with eight subjects in each average ability treatment cell. Three days later the criterion close tests (standard- and exact-length) were administered under untimed conditions.

To minimize what Lindquist (6) calls Type 2 (or group) errors, "treatment(s)" were administered at the same time in class groups. Each treatment group was made up of approximately proportional numbers of pupils from each standard and order of class testing was randomized.
A further comparison between standard- and exact-length blanks was made by correlating scores on each of these forms of cloze tests with scores on the Watts' Test for each level of difficulty. This, in effect, gave an estimate of the concurrent validity of cloze tests as measures of general reading comprehension.

Results

The key assumptions underlying analysis of variance are homogeneity of variance and normality of distribution. The former was tested for the main comparison groups (length of blank and passage difficulties) using F test for differences between variances of independent samples and Hartley's Fmax test (2) respectively; and the latter was tested by inspection. There was no evidence to suggest either assumption was unreasonable.

Lindquist's (6) procedure for testing main and simple effects was followed. The summary of the analysis of variance is presented in Table 1. Means, standard deviations, and the number of subjects on which these were based are shown in Table 2.

---

Tables 1 and 2 here

---

Inspection of Table 1 shows that no interaction term was significant. Nor was there a significant difference between the two versions of cloze tests. Thus there was no basis for rejecting either Hypothesis 1 or Hypothesis 2.
The differences between passages was highly significant which was to be expected since this was the basis for selecting the passages. The purpose of dividing subjects into reading ability levels was to increase the precision of the experiment by matching subjects on a related variable.

The product moment correlation coefficients for each group of 16 subjects on the Watts' Test and the standard- and exact-length versions of cloze tests for each of the three passages appear in Table 3. All correlation coefficients were significant at the five per cent level. A single estimate of the correlation for each type of blank and the Watts' Test was made using Fisher's z test. The mean z value for the three passages was calculated and transformed to the corresponding r. The last row of Table 3 shows this average coefficient of correlation.

Discussion and Conclusions

The principal finding in this study concerned differences between the mean score for standard-length blanks and the mean score for exact-length blanks. Although the latter yielded slightly higher scores than the former for passages at three levels of difficulty, this difference was not significant. For the subjects and passages used in the present study it mattered little whether blanks were of a uniform length or the same length as the deleted word.

Significant correlations between the two versions of cloze tests and the Watts' Test showed that the use of exact-length blanks was no less
valid a procedure for measuring general reading comprehension than the use of standard-length blanks.

The question of length of blank to use in cloze tests has practical as well as theoretical significance. For example, if a teacher wishes to construct a cloze test for use with his pupils or if a reading centre wishes to calibrate a number of books, the usual procedure is to type the passage deleting words systematically through it (for example, every fifth or eighth word) and indicate deletions by lines of uniform length. However, with the general availability of photo-copying facilities it is simpler and more convenient to blank out words for deletion, provide a space for recording answers at the right hand side of the page, and run off multiple copies. This process necessitates having blanks of the same length as the word to be deleted.

The advantage of the procedure suggested above is that it allows such factors as size of print, illustrative material, and page layout to be incorporated in, not excluded from, the cloze estimate of readability. If the findings of this study are confirmed by other studies, cloze procedure may prove an even more powerful technique for measuring readability.
References


<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blanks (A)</td>
<td>1</td>
<td>555.844</td>
<td>555.844</td>
<td>2.629</td>
<td>NS</td>
</tr>
<tr>
<td>Passages (B)</td>
<td>2</td>
<td>26855.146</td>
<td>13427.573</td>
<td>63.498</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Levels (L)</td>
<td>2</td>
<td>23472.282</td>
<td>11736.141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction (AxB)</td>
<td>2</td>
<td>273.938</td>
<td>136.969</td>
<td>5.492</td>
<td>NS</td>
</tr>
<tr>
<td>Interaction (AxL)</td>
<td>2</td>
<td>35.113</td>
<td>17.557</td>
<td>0.061</td>
<td>NS</td>
</tr>
<tr>
<td>Interaction (BxL)</td>
<td>4</td>
<td>2865.062</td>
<td>716.266</td>
<td>2.498</td>
<td>NS</td>
</tr>
<tr>
<td>Interaction (AxBxL)</td>
<td>4</td>
<td>845.855</td>
<td>211.464</td>
<td>7.38</td>
<td>NS</td>
</tr>
<tr>
<td>(Cells)</td>
<td>17</td>
<td>(54903.240)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within (W)</td>
<td>78</td>
<td>22361.750</td>
<td>286.689</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>95</td>
<td>77264.990</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 2

Sample Means and Standard Deviations for three Passage Difficulties under Standard- and Exact-Length Blanks

<table>
<thead>
<tr>
<th>Passage Statistics</th>
<th>Standard-Length</th>
<th>Exact-Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Mean</td>
<td>58.19</td>
<td>40.13</td>
</tr>
<tr>
<td>N</td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>
### TABLE 3

Validity Coefficients at Three Levels of Difficulty

<table>
<thead>
<tr>
<th>Difficulty Levels</th>
<th>Watts and Standard-Length</th>
<th>Watts and Exact-Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passage A</td>
<td>.51*</td>
<td>.65</td>
</tr>
<tr>
<td>Passage B</td>
<td>.92</td>
<td>.73</td>
</tr>
<tr>
<td>Passage C</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>Mr</td>
<td>.61</td>
<td>.75</td>
</tr>
</tbody>
</table>

* A correlation coefficient of 0.426 is required for significance at the 0.05 level with 14df.