

## Concordance And Consensus Indicators Of Four Measures Of Readability

Tan Soon Hock  
Pusat Bahasa  
Universiti Malaya

### Rationale

The effect of modern approaches in many subjects is to put a higher premium than ever on the ability to read. There is increasing use of assignment cards and worksheets. All too often these and the tasks they prescribe make no allowance for individual differences in reading ability, and the advice given to subject departments should include a concern for readability levels in the materials being used

(Department of Education and Science, England, 1975).

The short quotation above, though with reference to a foreign setting, has touched on factors that are equally applicable to the Malaysian education environment, which forms the prime focus of this paper, namely,

- (i) the importance of the reading skill,
- (ii) the significance of being aware of individual differences and how to cater for them in reading;
- (iii) by implication, the indispensable role of the language teacher in choosing suitable reading materials with or without the help of subject specialists;
- (iv) the role of content specialists in selecting materials of suitable readability levels.

The importance of the reading skills cannot be overstated. It is an indispensable component to any language course, if not a specialised need by itself. Its position stretches the entire length of the education. At the highest end of the continuum it assumes an all-important role of forging the key to open the door to a whole vista of knowledge crucial to the individual disciplines. It is in dealing with such a vast expanse of levels and an interesting mix of individuals with their own preferences, levels of ability and interest that a teacher, regardless of the educational level he fits into, needs to cater to individual differences. Considerations for this purpose would, therefore, not only have to take into account the individual but also the text itself. For the individual, factors like entry level of reading proficiency, age, background, interest, amount of prior knowledge he brings to the text as an aid to text understanding, need to be examined. A lack of these considerations might have helped to contribute to a situation in which as Aukerman (1965) found, at least three million young people in Grades 7-12 in America were given American Literature, English Literature and World Literature textbooks that they could not read. Chall (1958) suggests that elements relating to the text that would affect reading difficulty would be vocabulary load, sentence structure,

idea density, human interest or directness of approach. The English language teacher often bears the responsibility of selecting texts which cater to these two main considerations. He has to constantly ask: Are the materials suitable for my pupils? Are they appropriate to their abilities? In assuming this role, the decision is a difficult one, as the intertwine of content and language raises a problematic and often controversial issue. If the content matter is right, is the linguistic level suitable and vice-versa? If the attention were to be placed on linguistic suitability, would this be at the expense of content aptness?

The situation associated with the above problem is highlighted in an ESP reading program that teaches reading through the content area. Einstein noted that 'reading is the most complex task that man has ever devised for himself' (Dechant, E.V & Smith, H.P 1977:333). It involves the interpretation of printed symbols and the making of discriminative reactions to the ideas expressed by them. Quoting Davis (1960:34) '...it is the experience of teacher after teacher that reading processes can be taught and learned only in the context of the ideas and the context of the reading materials themselves.' In such a case the language teacher would best work in close consultation with the subject specialist in choosing the 'appropriate' material so as to lend a touch of authenticity, to provide a purpose for reading and to arrive at the right linguistic level of the students.

It is the main interest of this investigation, therefore, to probe how far in this collaboration, do the two specialist parties agree with each other in viewing a text. A commonly used instrument to judge readability of texts has also been used concurrently with these two measures, namely, the Fog and Fry formulae.

### **Research Questions**

The study aims to answer the following questions

- (i) What is the degree of concordance between the four measures of readability for the six Geography passages? The four measures of readability were yielded by the following groups.
  - a. the content specialists, that is, Geography lecturers
  - b. the language specialists
  - c. the Fog readability index
  - d. the Fry readability index
- (ii) What is the degree of concordance between the four measures of readability for the eight Science passages? The four measures of readability were:
  - a. the content specialists, that is Science lecturers
  - b. the language specialists
  - c. the Fog readability index
  - d. the Fry readability index
- (iii) What is the degree of consensus between types of measures namely:
  - a. the content specialist raters for Science and Geography passages
  - b. the language specialists raters for Science and Geography passages
  - c. the two readability formulae for Science and Geography passages

## Method

### (i) *Text Selection*

Six Geography passages were selected from texts used for Forms 1-5 in national-type secondary schools in Malaysia. The same source, namely, school Science texts for Forms 1-5 were used for choosing the eight Science passages. Within the range of levels it was hoped that a similar range of text difficulty in terms of concept and language would be available.

### (ii) *Instruction to Specialist raters*

A team of 4 raters was assigned to each set of passages. the Geography lecturers rated the Geography passages, while the Science lecturers rated the Science passages. The language specialists, however, were required to rate both Geography and Science passages. Content specialists were specifically instructed to gauge the conceptual difficulty of each text with reference to a Form 5 target group along a 3-point continuum: Easy, Average and Difficult. Language specialists were asked to gauge the linguistic levels of each of the text.

Hence, the two groups would be paying attention to their area of specialisation: content for the content specialists and language for the language specialists.

## *Readability Formulae*

One of the more commonly-used approaches to assessing the comprehensibility of reading material is the use of readability formulae. A large number of them have been developed. The formulae usually involved a selection of samples of texts followed by the counting of some objective characteristic of the samples. Betts (1949) suggested the major determinants of readability as

- a. average number of words per sentence
- b. number of simple sentences
- c. number of prepositional phrases
- d. the percentage of different words
- e. the number of uncommon words
- f. the number of words beginning with certain letters
- g. the number of polysyllabic words
- h. the number of adjectives, adverbs, personal pronouns and other words having a personal reference.

Sochor (1954) suggested that the difficulty of materials depended largely on the number and unusualness of facts that are presented, the vocabulary or terminology, the context of language setting in which they are presented. Peterson (1954), after reviewing research on readability, suggested that the following factors deserved further thinking:

- a. density of quantity of facts presented in a limited space
- b. degree of directness with which ideas are presented

- c interest appeal
- d difficulties of ideas caused by remoteness of ideas from reader's experience and by lack of explanation
- e abstractness of treatment
- f use of verbal or pictorial illustrations
- g patterns of organization

Dale and Chall (1949) said that readability formulae do not directly measure conceptual difficulty, organisation or abstractness of subject matter though these factors are known to affect comprehensibility. In fact, at times they have been found to yield different results. This is not to say, however, that it does not have use for giving relative estimates of the difficulty of books and texts.

The two formulae used in this study are the Fry graph and the Gunning's FOG

The graphical form that Fry offers makes it a convenient, straightforward way of obtaining a readability index. It shows the approximate reading grade level of printed materials by measuring two factors, word length and sentence length. Directions for using the Fry Readability Graph are as follows.

- a. Select three hundred-word samples of continuous print, one from near the beginning, one about the middle and one near the end of the book. Do not count numbers. Count proper nouns.
- b. Count the total number of sentences in each hundred-word sample, estimating to the nearest tenth of a sentence. Find the average of the three total numbers of the three samples.
- c. Count the total number of syllables in each hundred-word sample. Or, for convenience, count every syllable over one per word and then add a hundred. Find the average of the three total numbers of the three samples.
- d. Plot the average number of sentence length and of syllables of the three hundred-word samples on the Readability graph.
- e. If there is large variability between the three hundred-word samples, in either sentence length or in number of syllables, select several more hundred-word samples and average them before plotting them on the graph.

The FOG index of readability concentrates on the type of words and the number of words per sentence in the text. Its instructions are:

- a. Count the average number of words per sentence ( $w$ ) using several samples.
- b. Count the percentage of words with three or more syllables ( $s$ ), excluding proper nouns, combinations of short easy words like 'book-keeper', the verb form syllables 'ed' and 'es'
- c. The FOG index is  $(w + s) \times 0.4$ .

The Index provides a rough guide to the number of years of schooling needed for the reader to understand clearly what has been written, for example,

Fog Index	Reader
17+	Not easily read by anyone
16	Graduates
13	'A' level
11	'O' level
9	Non-qualified school leavers

#### Consensus Indicators

Tables 6, 7, 8, 9 indicate the degree of consensus that exists within each group of specialist raters for a specific set of texts.

Each set of texts was rated by four specialists in the same content area. For example, Geography texts were rated by four Geography lecturers. The consensus indicator showed the degree of agreement amongst the raters on each text when judging conceptual difficulty. If three out of four of the raters agreed that the text in question was easy the consensus indicator would be  $\frac{3}{4}$ . A unanimous decision would score  $\frac{4}{4}$ . Both these values would be regarded as strong consensus. In a case where two raters agreed on one point and the other two on another, then the consensus would be  $\frac{2}{4}$  (refer Table 6, Text 2). This is considered as an average consensus. Weak consensus would arise when two raters agreed on a particular point with the other two differing in extremes (Table 6, Text 3 — two raters agreed that the text was of average difficulty while the third considered it as easy and the fourth rater found it difficult).

#### Findings

Tables 1, 2, 3, 4 show the degree of concordance among the four measures of readability for different sets of texts. Some significant results were obtained.

- (i) Generally, there is low correlation between the four measures of readability with reference to Science texts. The highest degree of concordance is between the two readability formulae, Fog and Fry (0.673). The language teachers had different correlations with the two formulae, a higher correlation of 0.524 with the Fog and 0.102 with the Fry. Science specialists and language teachers only correlated at 0.173 while discordance occurs between the ratings of the Science lecturers and the Fog formula ( $-0.244$ ) — refer Table 1

	1 Science Specialists	2 Language Teachers	3 Fog	4 Fry
1 Science Specialists		0.173	-0.244	0.12
2 Language Teachers			0.524	0.102
3 Fog				0.673
4 Fry				

*Table 1* Spearman's Correlation Coefficient of Four Measures of Readability for Science Texts

- (ii) The language teachers and the readability formulae have a high consensus regarding the difficulty of Geography texts. The two readability formulae have a strikingly high correlation coefficient here (0.815). However, the content specialists have low correlation with the three linguistic measures (0.258, 0.2 and 0.3), similar to the findings on the Science texts (Table 9)

	1 Geography Specialists	2 Language Teachers	3 Fog	4 Fry
1 Geography Specialists		0.258	0.2	0.3
2 Language Teachers			0.886	0.757
3 Fog				0.815
4 Fry				

*Table 2.* Spearman's Correlation Coefficient of Four Measures of Readability for Geography Texts.

- (iii) Table 3 shows the rank ordering of Science texts by the four measuring devices. Kendall's coefficient of concordance was  $W = 0.3273$  which is not significant at the 0.05 level. Therefore, the null hypothesis that there is no relationship among the four measures of readability is accepted.

Science Texts	Science Specialists	Language Teachers	Fog	Fry	$\Sigma$	$\Sigma^2$
1	2.5	7	8	4	21.5	462.25
2	6.5	3	7	7.5	24	576
3	2.5	5	5.5	4	17	289
4	6.5	8	3	4	21.5	462.25
5	2.5	3	5.5	7.5	18.5	342.25
6	2.5	1	1	1	5.5	30.25
7	8	3	4	4	19	361
8	5	6	2	4	17	289
	36	36	36	36	144	2812

Table 3. Rank Ordering of Science Texts by Four Measures of Readability

$$\text{Kendall's Coefficient of Concordance: } W = \frac{12s}{K^2(n^3 - n)} = 0.3273$$

- (iv) Similar kinds of information laid out in Table 3 is found in Table 4 for Geography texts. Kendall's  $W = 0.5589$  is significant at the 0.05 level.

Geography Texts	Geography Specialists	Language Teachers	Fog	Fry	$\Sigma$	$\Sigma^2$
1	2	6	5	4.5	17.5	306.25
2	5	2	1	1	9	81
3	2	2	2	2.2	8	64
4	5	4	4	6	19	361
5	5	5	6	4.5	20.5	420.25
6	2	2	3	3	10	100

Table 4. Rank Ordering of Geography Texts by Four Measures of Readability

$$\text{Kendall's Coefficient of Concordance: } W = \frac{12s}{K^2(n^3 - n)}$$

$$W = 0.5589 \text{ (Significant } \alpha = .05)$$

- (v) In rating Science texts, the Kendall's coefficient of concordance among the three linguistic measures of readability (language teachers, Fog and Fry) is 0.3902. (Not significant at 0.05 level).

For Geography texts, however, the three linguistic measures yielded a coefficient of concordance  $W = 0.8317$  (Significant at the 0.05 level)

Language Teachers, Fog, Fry	Science Texts	Geography Texts
	$W = 0.3902$	$W = 0.8317$

*Table 5. Concordance Among Three Linguistic Measures of Readability*

- (vi) Table 6, 7, 8 and 9 examine in greater detail the degree of consensus within each set of raters with reference to particular discipline texts.

In rating Science texts, Science specialists showed strong agreement in rating difficulty of content. Seven texts had 3/4 or 4/4 as consensus indicators (Table 6).

Among the language teachers, however, (Table 7), there was only strong consensus for three texts (texts 1, 4 and 6). It is interesting to note that while 3 out of 4 Science lecturers considered Text 3 as easy in content, two of the language teachers thought it of average linguistic difficulty. One teacher even considered it difficult. Text 8 too, received extreme judgement from the two groups of raters, namely, Science specialists and language teachers. While the Science raters unanimously agreed that Text 8 was of average conceptual difficulty for fifth formers, language teachers considered it both easy and difficult. On the whole, there seems to be greater variance in estimation of linguistic difficulty of Science texts among language teachers.

Tables 8 and 9, refer in the same way as Tables 6 and 7, to Geography texts. Among Geography specialists, texts 4 and 5 attracted strong consensus, text 1, 3 and 6 average consensus while 2 was judged as being easy to difficult in content for fifth formers. (Table 8). By comparison, the language teachers did not show much more consistency than the subject specialists. Four of the texts (2, 3, 5, 6) had average consensus while only text 1 scored a strong consensus (3/4). Text 4 serves as an interesting contrast between Tables 8 and 9. While the Geography lecturers thought it to be of average difficulty, the language teachers were divided in opinion. In other words, though concept might be thought to be of average difficulty, there might be differing consensus regarding linguistic levels.



Geography Texts	Geography Lecturers				Consensus Indicators
	1	2	3	4	
1	Easy	Easy	Aver	Aver	2/4
2	Easy	Diff	Aver.	Easy	2/4 R
3	Easy	Aver	Aver.	Easy	2/4
4	Easy	Aver	Aver	Aver	3/4
5	Aver	Easy	Aver.	Aver	3/4
6	Easy	Aver	Aver	Easy	2/4

Consensus Indicator	f
3/4 and 4/4 — Strong Consensus	2
2/4 — Average Consensus	3
2/4 R — Weak Consensus	1

*Table 8: Consensus Indicators for Geography Texts Among Geography Lecturers*

Geography Texts	Language Teachers				Consensus Indicators
	1	2	3	4	
1	Diff	Diff	Aver	Diff.	3/4
2	Easy	Easy	Aver	Aver.	2/4
3	Easy	Aver.	Easy	Aver	2/4
4	Aver	Aver	Easy	Diff.	2/4 R
5	Diff	Diff.	Aver	Aver	2/4
6	Aver	Easy	Easy	Aver	2/4

Consensus Indicator	f
3/4 and 4/4 — Strong Consensus	1
2/4 — Average Consensus	4
2/4 R — Weak Consensus	1

*Table 9: Consensus Indicators for Geography Texts Among Language Teachers*

### Discussion

In both content areas, content specialists do not seem to concur with linguistic measures in their judgement of text difficulty (Tables 1 and 2). At times there are even extreme cases of judgement for the same text. The highest correlation is noted between the ratings of Geography specialists and Fry. This seems to be supporting the traditional dichotomy of concept and language in ESP teaching — that what is familiar and easy to the content specialists need not be so to the language teachers. Indeed, it might even be dauntingly difficult for some of the latter. This might lend further support to observations that simple topics might be couched in difficult language or, conversely, that simple language need not necessarily be dealing with simple concepts. This has an important bearing on choosing or writing teaching materials. Which would be accorded priority? Would conceptual grading supersede linguistic difficulty or vice-versa? If so, would understanding of concepts, however systematically graded, be hampered by syntactic, lexical, semantic difficulties? A realignment of linguistic and conceptual gradation seems to suggest itself.

Correlation coefficients between language teachers and the two formulae are much higher for Geography texts (0.886 and 0.752) than for Science texts (0.524 and 0.102). The disparity could have been influenced by the nature of the lexical items in the texts. A mono-syllabic word, which would be regarded as easy by the readability formulae, might have the reverse effect on the language teacher. Sentence length, a factor in the formulae, might not have been too overbearing for the teachers. Hence, the discrepancy. The ratings of the language teachers for the Science texts fluctuated as compared with the more consistent ratings for the Geography texts. The latter ratings are also indicators of weak consensus.

The above situation might be attributed, in part, to the nature of the language raters who are graduates of the social sciences. As such, it was possible that their linguistic ratings could have been tempered by their lack of familiarity with the Science content. This raises the perennial problem not only in judging readability but also in ESP teaching. Teachers who need to grapple with the matter of choosing suitable teaching materials for, for example, Law students, are frequently perplexed by their inability to separate language from content. The fact that it is difficult to do so may help account for a disparity in ratings between individual raters as well as between sets of raters.

### Reference

- Aukerman, R. (1960). 'Readability of Secondary School Literature textbooks. A First Report', *English Journal*, 54, 1533-40.
- Betts, E.A. (1949). 'Readability: Its Application to the Elementary School,' *Journal of Educational Research*, 42, 438-59.
- Chall, J.S. (1958). *Readability: An Appraisal of Research and Applications*, Bureau of Educational Research Monographs, Ohio State University, No 34.
- Dale, E. and Chall, J.S. (1949). 'The Concept of Readability', *Elementary English*, 26, 19-26.

Davis, S.E. (1960) 'High School and College Instructors Can't Teach Reading? Nonsense' *North Central Association Quarterly* 34, 295-99

Dechant, E.V and Smith, H.P (1977). *Psychology in Teaching Reading* Englewood Cliffs, N Jersey' Prentice Hall, Inc.

Klare, G (1974). 'Assessing Readability', *Reading Research Quarterly*, X, 62-102.

Mackay, R. and Mountford, Alan (1978) — *English for Specific Purposes, A Case Study Approach*, London. Longmans.

Peterson, E.M. (1954). *Aspects of Readability in the Social Studies*. New York: Bureau of Publications, Teachers College, Columbia University

Siegel, Sydney (1956) *Non-parametric Statistics* New York. McGraw Hill.

Sochor, E.E. (1954). 'Readiness and the Development of Reading Ability at all School Levels' *Education*, 74, pp. 555-60.