

New Readability Formula for Kannada Language

B. S. SWETHA AND N. NANJAPPA

Regional Horticultural Research and Extension Centre, UHS Campus, GKVK, Bangalore - 560 065

E-mail : shwethackmbs@gmail.com

ABSTRACT

The present study was conducted to develop and standardize a readability formula for Kannada by making use of word list developed for the study. As many as 155 sample criterion passages have been analyzed with reference to four important measures of readability, three related to words factor *viz.*, word familiarity, word length, technical words and one related to sentence factor *i.e.*, average sentence length (ASL). The two readability variables *viz.*, ASL ($r=0.34$) with respect to sentence factor and PNF_W ($r=0.83$) with respect to word factor had higher correlation coefficient values with the grade level and hence, these two variables were considered in the development of readability formula. The Readability formula obtained on the basis of regression analysis is $GL = 1.2967 + 0.2133 x_1 + 0.1177 x_2$. A readability chart has also been developed to facilitate quicker assessment of the readability level of the material studied. The readability formula developed has exhibited high reliability and has high content validity as well as experimental validity.

Keywords : Readability, criterion passages, sentence factor

A readability formula is a predictive device to determine the extent of readability of any selected written / published material *i.e.*, agricultural publications in particular or other publication in general. According to Willson and Gallup (1956) unless written material is readable and understandable, it cannot be an effective teaching device. Hence they suggested that most extension publication should be aimed at readers reading level. A piece of writing is readable if it could be read and understood by the readers for whom it is intended (Anon., 1963). The present day writers have a challenge of presenting the information in the most simplest and understandable form to reach the readers. The classic readability studies started as early in the late 19th century (DuBay, 2004) and the first formula to measure readability was published in 1923 by Lively and Pressy. Since then, more than 200 different readability formulas and more than 1000 studies in the field have been published in English (DuBay, 2004). However, of these formulas, only 12, at the most, are widely used in English (Gunning, 2003).

Attempts have also been made to develop readability formula in Indian languages for Hindi by Kaur (1978) and Sinha *et al.* (2012); for Malayalam, Rajan (1982); for Tamil, Oliver (1985); for Kannada, Nanjappa (1992) and Nomesh Kumar (2002).

However, it was only a beginning made in this direction and much needs to be explored. Earlier formulas in Kannada used word factors such as word complexity and personal words as variables. Since, knowledge of words has always been a strong measure of a reader's development, reading comprehension and verbal intelligence, hence word list developed for Kannada language has been used for measuring the variable non familiarity in developing readability formula. With these aforesaid aspects in view, the present study was conducted mainly to develop and standardize a readability formula for Kannada by making use of word list developed for the study.

METHODOLOGY

The relevancy of readability factors influencing readability formulae was discussed with specialists in the field of mass communication and journalism and only most relevant variables amenable for precise measurement were selected. Also based on review of literature three variables related to words and one variable related to sentence was selected as follows, *viz.*, 1. Word length, WL (McClusky, 1934); 2. Per cent technical words, PTW (Dale and Tyler, 1934 and Jacobson, 1965); 3. Percentage non familiar words, PNF_W (Dale and Chall, 1948; Forbes and Cottle, 1953); 4. Average sentence length in words, ASL

(Flesch, 1943; Oliver (1985); Nanjappa, 1992; Nomesh Kumar, 2002).

In the present study used the school text books prescribed by the government of Karnataka from 5th to 10th standards. To ascertain whether the text books selected for the study were comprehensible by the pupils of the corresponding standard, was confirmed by developing and conducting a comprehension test in each of the standard.

From the standard 5th to 10th subjects, science was selected comprising of 12 text books. In each of the text books considered, every 5th page was selected as sample for the purpose of enumerating the independent variables. In case of 5th page covering more than half a page of figures, bullet points and study questions etc., the nearest page was selected. In the sample passages following has been excluded from analysis; such as, headlines, names introducing a person's turn in dialogues, Captions to illustrations, Author names and signatures (often with time and place for letters), Additional information included in pictures or within, squares, tasks and exercises for, and instructions to, the reader (circled question marks), short facts and lists of items, tables and other illustrations, extra good stuff. Finally, 155 criterion passages were selected from 5th to 10th standard.

Further, these 155 sample criterion passages were analyzed with reference to four important measures of readability, three related to words factor *viz.*, word familiarity, word length, technical words and one related to sentence factor *i.e.*, average sentence length (ASL). At the first instance the data were subjected to correlation analysis. The two independent variables showed significant and positive relationship with the dependent variable was used for fitting the regression equation.

The multiple linear regression analysis was done by using the following model.

$$\hat{Y} = \alpha + b_1x_1 + b_2x_2 + \dots + b_nx_n + e$$

Where, \hat{Y} is the dependent variable $x_1, x_2, x_3, \dots, x_n$ are the independent variables b_1, b_2, \dots, b_n are the regression coefficients α is the ordinate or constant and e is the error, follows normal distribution with mean zero and constant variance.

An attempt has been made to develop readability formula using one of this variable percentage non-familiar words which is a strong predictor of text difficulty. In addition, another important readability variable representing the sentence factor in the form of Average sentence length (ASL) was considered to measure the readability.

Directions for the use of readability formula

Readability formula may be used by variety of people in different fields of communication and education. In this situation it is essential to use guidelines or the directions for the use of the formula. There should be clear guidelines for measurement of variables, or steps involved in calculating the readability scores. Without this, it would become difficult to apply the formula. Hence, the directions for use of the formula are explained as follows.

Selection of samples

While selecting samples, take about 100 words from every tenth page in case of books. For articles, select two to three 100 words samples per 1000 words. Never begin or end a sample in the middle of a sentence.

Measurement of variables: For analysing the readability variables in a selected passages and calculation of grade level scores of that passage, one can make use of the work sheet presented in Table I. The Table can be used to analyse 2-3 samples together.

Counting of words: Number of words occurred in the passage can be counted by following the procedure explained in the methodology chapter and enter the number of words under item 1 of the work sheet.

Counting of sentences: Compute the number of sentences in the sample as explained in the methodology chapter and record the same in item 2 of the worksheet.

Counting of non-familiar words: Count the number of Non-familiar words in the sample as elucidated in the methodology part and enter the same in item 3 of the work sheet.

TABLE I

A worksheet for analysis of samples for determining grade level difficulty

Article....	Sample 1	Sample 1	Sample 1
Author....	Page no:	Page no:	Page no:
Number of words in the sample			
Number of sentences in the sample			
Number of Non familiar words			
Percentage of Non-Familiar words (Divide item 3 by item 1, and multiply by 100)			
Average Sentence Length in words (Divide item 1 by item 2)			
Multiply ASL by 0.2133 (b_1x_1)			
Multiply PNFW by 0.1177 (b_2x_2)			
Constant (a)	1.2967	1.2967	1.2967
Formula score (add 6, 7 and 8) and round of the fractions to nearest whole number ($a + b_1x_1 + b_2x_2$)	_____	_____	_____
	Analysed by _____		
	Date _____		

Word familiarity per 100 words: Percentage of Non-familiar words is computed in item 4 it obtained by dividing the number of Non-familiar words in the sample by the number of words in the sample and multiplied by 100.

Average sentence length in words: The average sentence length (ASL) in words is worked out in item 5 by dividing the number of words in the sample by the number of sentences in the sample.

Calculation of grade level of passages: After analysing the readability variables occurred in the passages, follow through item 6, 7 and 8 on the work sheet. Wherein, item 6 obtained by multiplying ASL by 0.2133, item 7 attained by multiplying PNFW by 0.1177 and note the constant 1.2967 in item 7. Further, add items 6, 7 and 8 and round off the fractional value to the nearest whole number to attain the grade level of the sample passages in item 9 of the work sheet.

Reliability and validity

Reliability of the readability formula has been established using test – retest split-half methods to answer, can an analyst, analyzing the same samples at different times, agree with himself in his counts and another method to answer if the different analysts, analysing the same sample, agree with each other's counts. The validity of the present readability formula was ascertained by content validity and experimental validity using reading speed test.

Test re-test method

In this method ten selected reading passages from 5th to 10th grade were got analysed by ten judges at two intervals of time. Correlation between grade levels obtained by the judges at two intervals of time.

Split –half method

Consistency of the two different analysts in their judgement of the same material is the second aspect of reliability. To know the relationship between the grade level scores obtained by two groups of judges, six identical passages were got analysed by two groups and scores obtained were analysed by using Pearson's Product - moment correlation.

The present readability formula was found to be reliable on both these counts. The reliability of the readability formula rely upon on at least three major factors, viz., 1) Complexity of the formula, 2) Motivation of the analyst and 3) Experience of the analyst. It may be understood that the formula is not a difficult one to apply and does not necessitate high degree of skill for its application. Hence, the formula developed in this study is reliable.

While analysing the reliability of readability formula, Flesh (1943), Farr *et al.* (1951), Nanjappa (1992) and Nomesk Kumar (2002) reported high correlation value, which indicated that the formula possess the reliability.

Validity of the readability formula

The present readability formula was found to have content validity as well as experimental validity. The experimental validity was ensured by using an outside criteria i.e., reading speed. The independent variable had very high correlation with the dependent

variable and hence, it is presumed that the readability formula has an inbuilt validity.

Content validity: The readability variables were selected based on the extensive review of literature and discussion with experts. Further they were found significantly related to dependent variable as revealed by the significant value of correlation coefficient. Hence, the present readability formula is confirmed to have content validity.

Experimental validity

This type of validity concerns the ability of the formula to predict an 'outside criterion' of readability. In the present study experimental validity was established with an outside criteria *i.e.*, Reading speed. Grade level scores of 5th to 10th standard passages as measured by present readability formula was correlated with the reading speed of these passages with the help of Pearson's product-moment correlation.

RESULTS AND DISCUSSION

The values of correlation coefficients of four readability variables with grade level is presented in

Table II. Among the four variables, *viz.*, Average sentence length ($r = 0.34$), percentage Non-familiar word ($r = 0.83$), percentage of technical words ($r = 0.43$) and word length ($r = 0.09$). Further, there was a significant correlation at five per cent between Average sentence length and percentage Non-familiar words. The first two variables were selected for development of readability formula, as they exhibited higher 'r' values. The average sentence length as a variable was used by the earlier researchers like Bormuth (1966), Oliver (1985), Nanjappa (1992) and Nomesh Kumar (2002).

Higher value of multiple correlation (Table III) in the present study indicates the strength of relationship between observed and predicted variable and strong linear relationship. This implied that 73 per cent of variation ($R^2 = 0.73$) in the grade level was explained by the two independent variables, *viz.*, average sentence length and percentage non-familiar words. Further, extent of variation predicted in the present study was in conformity with that of earlier studies by Dale and Chall (1948) and Nanjappa (1992).

TABLE II
Correlation matrix between grade level and four Readability variables

	Grade level	ASL	WL	PTW	PNFW	
Grade level	1					
ASL	0.34 **	1				
WL	0.09 NS	-0.05	1			
PTW	0.43 **	-0.11	0.07 NS	1		
PNFW	0.83 **	0.18 *	0.07 NS	0.48 **	1	

* - Significant at 0.05 level of probability

** - Significant at 0.01 level of probability, NS – Non significant

TABLE III
Regression analysis of grade level with two independent variables

(n=155)

Variable	Regression Co efficient (b)	SE of Regression Co efficient (Seb)	t' value	A	R2	F
ASL	0.2133	0.046	4.63 **	1.2967	0.73	209.29
PNFW	0.1177	0.006	18.78 **			

Regression weights presented in Table II were utilized for fitting regression equation:

$$a = 1.2967, b_1 = 0.2133 \text{ and } b_2 = 0.1177$$

Using these values, the following readability formula was obtained:

$$GL = 1.2967 + 0.2133 x_1 + 0.1177 x_2$$

Where, GL= Grade level of a passage, x_1 =Average Sentence Length in words (ASL), x_2 =Percentage Non-Familiar words (NFW).

The present readability formula developed by making use of regression analysis. Which utilized the percentage non-familiar words as word factor and average sentence length in words as sentence factor for assessing the grade level of sample passages by readers. The variable word familiarity has been a strong measure of a reader’s reading comprehension and verbal intelligence.

The formula developed has been standardized. The selected criterion passages were tested for their indication of that particular standard difficulty by using school children and conducting comprehension tests.

In order to attain at the grade level scores of passages, there is a need to measure the ASL and PNFW of the intended passages and substitute the figures in the formula. Grade Level increases as the values of ASL and PNFW increase. Normal probability plot of regression equation is depicted in Fig.1.

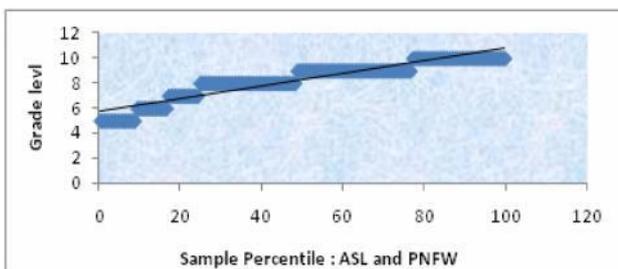


Fig. 1 : Normal Probability Plot of Regression equation

Each number on the formula scale corresponds to one grade level. This relationship is proper only up to the tenth standard. Further than that formula underrates grade level. However, it may extrapolated till the 11th and 12th standard and even higher standards.

Reliability of the readability formula

The results shown that the present readability formula developed had exhibited high analyst reliability both on test retest method and independent judging method (Table IV and V).

Test Re-test method

Correlation between grade levels obtained by the judges at two intervals of time : Scores obtained by the judges at the two intervals of time (Table IV) were correlated using Pearson’s Product–moment correlation. The value of ‘r’ found to be 0.99, which was significant at 1 per cent level of probability, which indicated the reliability of the formula developed.

TABLE IV

Correlation between grade levels obtained by the judges at two intervals of time

Judges	Grade level score obtained		'r' value
	I Time	II Time	
1	5.90	5.80	
2	6.65	6.53	
3	7.19	7.07	
4	5.02	5.13	0.99 **
5	4.43	4.32	
6	7.27	7.17	
7	5.35	5.47	
8	5.46	5.57	
9	5.35	5.45	
10	4.49	4.38	

** Significant at 0.01 level of probability

Split–half method

Correlation between grade level scores of the two groups of judges on identical passages : The results of this test are given Table V revealed a correlation coefficient of 0.98, which was significant at 0.01 level of probability, thereby indicating the reliability of the present formula.

Experimental validity

Correlation between the reading speed and grade level Scores as measured by the formula : The results of this test are given in Table VI. The value of correlation coefficient was found to be 0.98, which

TABLE V
Correlation between grade level scores of the two groups of judges on identical passages

Grade level of passages analysed	Grade level score obtained		'r' value
	Group - I	Group - I	
6	6.21	6.3	0.98 **
7	6.77	6.88	
6	6.02	5.92	
5	5.03	5.24	
5	4.54	4.42	
6	6.23	6.31	

** Significant at 0.01 level of probability

TABLE VI
Correlation between the reading speed and grade level Scores as measured by the formula

Grade level (as measured by the formula)	Average Time taken to read(minutes)	Reading speed Words/minute	'r' value
1	5.90	5.80	0.98 **
5	1.28	78	
6	1.34	74	
7	1.40	71	
8	1.45	69	
9	1.47	68	
10	1.49	67	

** Significant at 0.01 level of probability

points out a highly significant relationship between the reading speed and the readability scores as measured by the formula. The data in Table reveals that reading speed decreased as the grade level of the passage increased. Hence, the reading ease formula developed is supposed to be a valid instrument.

It was found that there was a positive and significant relations between reading speed and the readability scores of passages as measured by the formula (Table VI). A positive relationship between reading speed and readability scores have been reported by Flesch (1948), Hackman and Kershner (1951), Brown (1952), Klare (1952), Klare *et al.* (1957) and Nanjappa (1992).

It is quiet natural that simple material is read faster and difficult material takes longer time to read. On this ground, it is expected that the time required to read a unit of 100 words of 4th standard passage should be less than the time required to read the same unit of words of any higher standard passage. This relationship was clearly established by the time taken to read the passages of 5th to 10th standards by the respondents in the present study, as the time taken increased along with the standard and reading speed decreased. Hence, this establishes the experimental validity of the readability formula developed.

Norms of distribution of readability variables

The norms of distribution of two readability variables from 5th to 10th standard had shown definite trends presented in Table VII. Mean of average

TABLE VII
Mean, standard deviation and range of the two readability variables with respect to 5th to 10th standard

Standard	Average sentence length			Percentage Non familiar words		
	Mean	S. D	Range	Mean	S. D	Range
5	8.57	0.96	6.81 - 9.79	27.34	6.80	18.31 - 36.29
6	8.18	0.92	6.39 - 9.42	24.92	3.96	19.04 - 32.05
7	8.76	1.13	7.14 -12.00	31.14	3.76	24.45 - 37.00
8	9.78	1.34	8.12 -14.25	40.56	5.27	26.78 - 51.39
9	9.80	1.51	7.00 -13.75	47.45	5.70	32.99 - 57.80
10	9.78	1.35	6.92 -13.70	51.23	5.51	42.07 -61.45

** Significant at 0.01 level of probability

sentence length varied from 8 to 10 words and ranged from 6-14 words. While, the mean of percent non-familiar words varied from 27 to 51 words from 5th to 10th standard, ranged from 18 to 61 words (Table VII and Fig. 2). This results revealed that there is a scope to vary the percentage non-familiar words and average sentence length in manipulating the level of readability. Smaller the number of different words, the easier the material, the results are in line with Lorge (1939), Dale and Chall (1948), Chall, (1974).

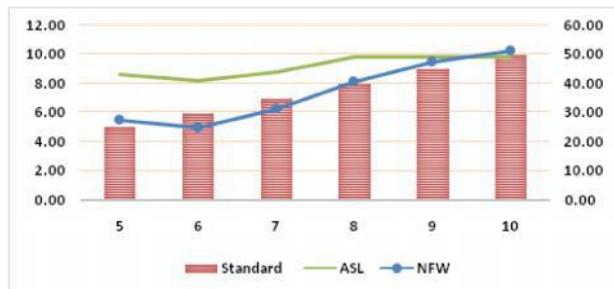


Fig. 2: Changes in ASL and PNFW over 5th to 10th standards

The average sentence length ranged from 6 to 14 words from 5th to 10th standard and found to be significantly related to grade level. Generally the longer the sentences were, the harder the text. The results are in conformity with Nanjappa (1992) Nomesh Kumar (2002). There are no studies of readability using Kannada word list.

Readability Chart

To find out the range of readability level of a given passage a readability chart will be more

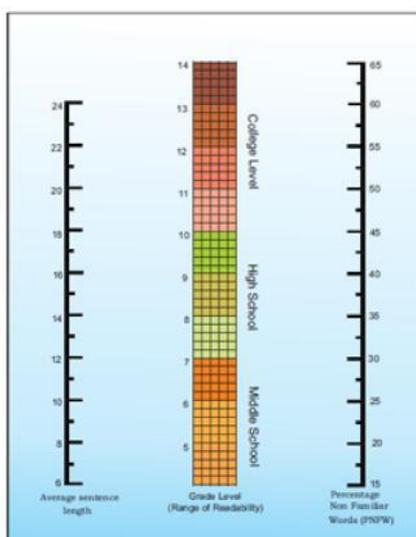


Fig.3 : Readability chart

convenient which is developed and presented in Fig. 3. This chart can be used by finding out the average words per sentence and non-familiar words per 100 words and connecting the respective columns outside of the readability column in the centre. The line passing through the middle column representing the grade level difficulty provides the range of readability of the passage.

The readability formula developed in the present study has direct application in the field of education, mass communication, journalism and agricultural extension. The developed formula can be used by the communicators, writers and authors to know the grade level of their writing. It is also possible to modify the level of difficulty to suit to the required literacy level of the readers. The formula will be particularly helpful for prior testing of material before publishing it for the benefit of readers in general and farmers in particular. By using the present readability formula, it is also possible to analyze the already published extension literature and other publications to determine their suitability to the intended readers. This work will provide the required feedback for communicators and writers to appropriately modify the written material for the benefit of readers.

REFERENCES

- ANONYMOUS, 1963, Simple reading materials for adults, manuals and adult and youth education, NO.3, UNESCO, Paris.
- BORMUTH, J. R., 1966, Readability : A new approach, *Reading Res. Quarterly*, **11** (3) : 79 - 131.
- BROWN, J. I., 1952, The flesh formula 'Through the Looking Glass'. *College English*, **13** : 393 - 394.
- CHALL, J., 1974, Readability : An appraisal of research and application. Essex: Bowker Publishing Company.
- DALE, E., AND TYLER, R. W., 1934, A study of the factors influencing the difficulty of reading materials for adults of limited. Reading ability, *Library quarterly*, **4** : 384 - 412.
- DALE, E. AND CHALL, J. S., 1948, 'The concept of readability' *Elementary English*, **26** : 19 - 26.

- DUBAY, W., 2004, The principles of readability. Impact information, pp: 1 - 76.
- FARR, J. N., JENKINS, J. J. AND PETERSON, D. G., 1951, Simplification of Flesch reading ease formula, *J. Appl. Psychol.*, **35** : 333 - 37.
- FLESCH, R. F., 1948. A new readability yardstick. *J. Appl. Psychol.*, **32** : 221 - 233.
- FLESCH, R. F., 1943, Marks of readable style: A Study in adult education, New York : Bureau of Publications, Teachers College, Columbia University.
- FORBES, F. W. AND COTTLE, W. C., 1953, *A New Method for determining Readability of Standard Tests* McGraw Hill, New York.
- HACKMAN, R. C. AND KERSHNER, A. M., 1951, The determination of criteria of readability (Technical report), University of Maryland, Baltimore city.
- JACOBSON, M. D., 1965, Readability difficulty of physics and chemistry textbooks. *Edu. Psychol. Measurement*, **25** (2) : 449 - 457.
- KAUR, D. K., 1978, Developing textual and format criteria for an effective Hindi farm magazine, *Ph.D. Thesis (Unpub.)*, Indian Agricultural Research Institute, New Delhi.
- KLARE, G. R., 1952, Measures of the readability of written communication: An evaluation," *J. Educ. Psychology*, **43** : 385 - 399.
- KLARE, G. R., SHUFORD, E. H. AND NICHOLOS, W. H., 1957, The relationship of style difficulty, practice and ability to efficiency of reading and to retention. *J. appl. Psychol.*, **41** : 222 - 226.
- LIVELY, B. A. AND PRESSEY, S. L., 1923, A method for measuring the vocabulary burden of textbooks. *Edu. Adm. Supervision*, **9** : 389 - 98.
- LORGE, I., 1939, Predicting reading difficulty of selections for children, *Elementary English Review*, **16** : 229 - 33.
- MCCLUSKY, H. Y., 1934, A quantitative analysis of the difficulty of reading materials", *J. Educ. Res.*, **28** : 276 - 282.
- NANJAPPA, D., 1992, Development of readability formula for kannada and its application. *Ph.D. (Agri.) Thesis (Unpub.)*, Univ. Agric. Sci., Bangalore.
- NOMESH KUMAR, N., 2002, Development of reading ease and human interest kannada readability formula for farm publications. *Ph.D.(Agri.) Thesis (unpub.)*, Univ. Agric. Sci., Bangalore.
- OLIVER, J., 1985, Treatment for effective message and development of readability measure, *Ph.D. Thesis (Unpub.)*, Tamil Nadu Agricultural University, Coimbatore.
- RAJAN, P. K., 1982, Effectiveness of communication through the farm news service of Kerala Agricultural University". *M.Sc. (Ag.) Thesis (Unpub.)*, Kerala Agricultural University, Mannuthy.
- SINHA, M., SHARMA, S., DASGUPTA, T. AND BASU, A., 2012, New readability measures for Bangla and Hindi texts. In Proc. of COLING, 2012; Posters, Mubai, India. pp : 1141-1150.
- WILSON, M. C. AND GALLUP, G., 1956, Extension teaching methods, new USDA Federal Extension Service Circular, No. 459, New York.

(Received : November, 2017 Accepted : February, 2018)