

WILEY



Using a Computer to Calculate the Dale-Chall Formula

Author(s): Jeanne Gardner Barry and Timothy E. Stevenson

Source: *Journal of Reading*, Vol. 19, No. 3 (Dec., 1975), pp. 218-222

Published by: [Wiley](#) on behalf of the [International Reading Association](#)

Stable URL: <http://www.jstor.org/stable/40011418>

Accessed: 18/06/2014 18:12

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Wiley and *International Reading Association* are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Reading*.

<http://www.jstor.org>

Using a Computer to Calculate the Dale-Chall Formula

JEANNE GARDNER BARRY
TIMOTHY E. STEVENSON

Barry is an associate professor of mathematics at California State University, Hayward. Stevenson is a doctoral candidate in computer science at the University of California, Davis, Lawrence Livermore Laboratory.

■ Determining readability levels of textbooks is a matter of concern to teachers. Readability formulas such as the Dale-Chall formula provide an accurate measure of readability, but their use can be cumbersome and time-consuming. Since the application of the Dale-Chall formula is mechanical in nature and since computers can handle character data as well as number data, a solution to the problem seems possible on a computer.

The following program is designed to follow exactly the computation of the Dale-Chall Readability Formula (Dale and Chall, 1948). The program is written in Fortran, the most widely used computer language. Two types of input are possible—typewriter terminal input and punched cards. Both methods are easy to use and require no computer expertise. The results of the analysis of the example in the Dale-Chall article agree exactly with the published results.

We have analyzed a wide spectrum of texts from grades four through fifteen. The computation of the reading level for a 500-page text on the CDC

3150 (Control Data Corporation computer number 3150) takes less than three minutes, compared with a hand calculation of many days. We feel that this research should prove very useful to a wide range of computer users including teachers, administrators, textbook editors and writers, as well as elementary, secondary and college textbook selection committees.

The Dale-Chall formula is based on two factors: 1) vocabulary load (relative number of words outside the Dale list of 3000 words) and 2) a factor of sentence structure based on average sentence length. Since it applies only to continuous textual material, textbooks containing many formulas, graphs, or charts must be excluded.

To use this formula for books, samples of 100 words are selected from about every tenth page. For articles, about four evenly spaced 100-word samples per 2000 words are selected.

A count of the unfamiliar words is made from the samples. A familiar word is either a word appearing in the Dale basic vocabulary of 3000 words or 1) a plural or possessive of a basic word in the list; 2) a verb from the list formed by adding *s*, *ies* (from *y*), *ing*, *n*, *ed*, or *ied* (from *y*) or formed by doubling the consonant before adding *ing* or *ied*; 3) a comparative or superlative of an adjective in the list; 4) an adverb formed by adding *ly* to a word in the list.

In addition, 5) names of persons and places are considered familiar even though they do not appear on the list; 6) names of organizations, laws, documents, titles of books or movies, when used several times in a sample of 100 words, are counted only twice; 7) abbreviations are counted as one word; and 8) hyphenated words are counted as familiar

only if both words in the compound appear on the word list.

Once this count is made, the remaining steps are easily followed. 1) The average sentence length, X_1 , is computed by dividing the number of words in the sample by the number of sentences in the sample. 2) The percentage of words outside the Dale list, X_2 , is obtained. 3) $X_3 = .1579X_2 + .0496X_1 + 3.6365$ where X_3 is the formula raw score for one sample. 4) For more than one sample, the formula raw scores per sample are averaged. 5) The average formula raw score is converted to a corrected grade reading level.

The Computer Program

Basic to the entire analysis is the representation of the English language within the computer. This is machine-oriented and too technical to be discussed here. The Dale-Chall 3000-word vocabulary is set up as a dictionary in arrays based on the length of the words. The words in the samples of the texts are also sorted into arrays based on the number of characters in each word and stored in the format of the dictionary. Each array of the sample is checked against the dictionary. If a word is not found in the dictionary, various suffixes and derivatives of the word must be checked. A special routine will determine if the word has a suffix of *ied*, *ies*, *ing*, *est*, *s*, *es*, *ed*, *er*, *ly*, *n*, or *s*. If the suffix is found, it is removed and the word checked again in the dictionary. If the word is not found and if a double consonant is detected at the end of the word, the consonant is removed and the word checked again. A count is kept of all unfamiliar words except abbreviations, hyphenated words, proper names, and organiza-

tions. A word count and sentence count are maintained.

Samples of one hundred words from every tenth page for books, or four such samples per 2000 words for articles, are fed into the computer using either typewriter terminals or punched cards.

The following restrictions apply to input:

1. A sample must never begin or end in the middle of a sentence, since the number of sentences in a sample are counted.

2. Words are not hyphenated at the end of a line of input; otherwise they would be processed as hyphenated words.

3. Names of persons and places appear in quotes. This is because they must be singled out as familiar words even though they do not appear in the list. If quotes appear in the selection, they are not included in the input (they have no effect on the reading level). For example, *Joe said, "I did it"* will be *"Joe" said, I did it.*

4. Names of organizations, laws, documents, and titles of books or movies are enclosed by slashes. Again this is because of special rules applying to them, as outlined previously. For example, *Chicago Building Association* would be typed as */"Chicago" Building Association/*, with slashes for the organization and quotes because Chicago is a place.

5. Abbreviations are treated as one familiar word. Since a period terminates sentences for the computer program, the periods in abbreviations are changed to asterisks. *Y.M.C.A., Nov., Mr., U.S., or a.m.* are typed for input as *Y*M*C*A*, Nov*, Mr** and so on.

The program will print out these instructions, if desired, before accepting text material. The program then reads the title, author, publisher, date of publication, and the samples of textual material into the computer in that order.

The final output gives 1) the title, author, publisher and date; 2) the number of words in the sample; 3) the number of sentences in the sample; 4) the number of unfamiliar words; 5) av-

erage sentence length; 6) the Dale-Chall score; 7) the raw score; and 8) the corrected grade level.

Examples

The following example appears in the original Dale-Chall article (1948) and agrees exactly with Dale and Chall's results. Typewriter terminal input was used.

Input

Exfor

New or old—old

Old file name—Textanal

Would you like an explanation of this program and the restrictions and methods of input?

No

What is the title of the book or article?

Your Baby

What is the author's name?

Unknown

Who is the publisher?

Nat'l TB Assoc.

What is the date of publication?

1945

What method of input do you wish to use?

(File or terminal)

File

A happy, useful life, that's what you want for your baby, isn't it? And because a healthy mind and body are so necessary to happiness and long life, you must do all you can to get your baby off to a good start. There is much you can do. . . . Kissing is one way of spreading TB as well as other germs. Tuberculosis of the bones or joints or of certain organs of the body besides the lungs can come to the bottle-fed baby in milk which has not been pasteurized or boiled.

Output

Article/title: Your Baby

Author:

Publisher: Nat'l TB Assoc.

Date: 1945

1. Number of words in sample 374

2. Number of sentences in sample... 22
 3. Number of words not in Dale list... 43
 4. Average sentence length... 17
 5. Dale score... 11
 6. Raw score... 6.2166
 Corrected grade-level... 7-8th grade

Do you have any other text to analyze?
 No
 Stop 0
 Ready

The following examples omit the input and list only the final results.

A. This text was analyzed using our on-site CDC 3150 with punched card input. The last line gives the total computer time. For example, 00/02/54 indicates no hours, two minutes and fifty-four seconds for the analysis of *Living in the United States*, a fifth grade text of 425 pages.

Article/title: Living in the United States
 Author: Cutright and others
 Publisher: Macmillan
 Date: 1966
 -
 1. Number of words in sample... 2842
 2. Number of sentences in sample... 219
 3. Number of words not in Dale list... 263
 4. Average sentence length... 13
 5. Dale score... 9
 6. Raw score... 5.7024
 Corrected grade-level... 5-6th grade
 Job elapsed time 00/02/54

B. This example is a state adopted text for the sixth grade in California. The Dale-Chall reading level indicates seventh grade. The results were obtained using terminal input, so no computer times are available.

Article/title: Life in Latin America
 Author: Edmund Lindop
 Publisher: Ginn and Company
 Date: 1964

1. Number of words in sample... 3875
 2. Number of sentences in sample... 239
 3. Number of words not in Dale list... 418
 4. Average sentence length... 16
 5. Dale score... 11
 6. Raw score... 6.1670
 Corrected grade-level... 7-8th grade

C. This 511-page California state adopted text was analyzed in two minutes, forty-six seconds. The grade level is consistent with use.

Article/title: Projection in Literature
 Author: Pooley, Daniel, Farrel
 Publisher: Scott, Foresman and Company
 Date: 1967

1. Number of words in sample... 2318
 2. Number of sentences in sample... 110
 3. Number of words not in Dale list... 260
 4. Average sentence length... 21
 5. Dale score... 11
 6. Raw score... 6.4150
 Corrected grade-level... 7-8th grade
 Job elapsed time 00/02/46

D. The following text is in use in some ninth grades. The reading level indicates eleventh grade.

Article/title: The Ecumene Story of Humanity
 Author: William H. McNeill
 Publisher: Harper and Row
 Date: 1973

1. Number of words in sample... 3834

2. Number of sentences in sample . . .	185
3. Number of words not in Dale list	864
4. Average sentence length	21
5. Dale score	23
6. Raw score	8.3098
Corrected grade-level	11-12th grade
Job elapsed time 00/04/59	

E. Result using a high school sociology text with terminal input.

Article/title: High School Sociology
 Author: William E. Cole
 Publisher: Allyn and Bacon, Inc.
 Date: 1963

1. Number of words in sample	2560
2. Number of sentences in sample . . .	168
3. Number of words not in Dale list	664
4. Average sentence length	15
5. Dale score	26
6. Raw score	8.4859
Corrected grade-level	11-12th grade

F. This text is used in the freshman sociology course at California State University, Hayward. The reading level indicates a college level of thirteenth to fifteenth grade.

Article/title: Toward a New Sociology, A Critical View

Author: Charles H. Anderson
 Publisher: The Dorsey Press
 Date: 1971

1. Number of words in sample	3445
2. Number of sentences in sample . . .	131
3. Number of words not in Dale list	1102
4. Average sentence length	26
5. Dale score	32
6. Raw score	9.9789
Corrected grade-level	13-15th grade (college)

The almost unexpected accuracy of this program together with its economy of time spent in analysis gives rise to a series of questions. Should all texts be tested? Before or after publication? Should all students' reading ability be periodically tested to insure proper matching of student to text? Are grades given students using inappropriate texts valid? Should texts dealing with identical material be published on several readability levels?

The theory of this computer program is not limited to the Dale-Chall formula. Work is in progress on computerizing the Spache formula.

Reference

Dale, E. and J. S. Chall. "A Formula for Predicting Readability." *Educational Research Bulletin*, vol. 27 (1948), pp. 11-20.

Did You Know . . .

that the IRA state council in Indiana has its own journal? Inquiries and contributions can be sent to J. David Cooper, Indiana Reading Quarterly, 306 Teachers College, Ball State University, Muncie, Indiana 47306.