Unlocking Language
The Classic Readability Studies

Volume 1
1921—E. L. Thorndike and the Familiar Word

William H. DuBay
Series Editor

Copyright

The following journal article is in the public domain:

Thorndike, Edward L. November, 1921. "Word Knowledge in the Elementary School." *The Teachers College Record*. Vol. 22, No. 5, pp.334-370.
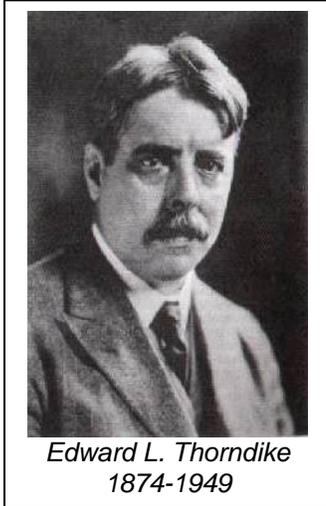
Readers' Comments

Please send all comments and suggestions regarding this document to:

William DuBay
Impact Information
126 E. 18th Street, #C204
Costa Mesa, CA 92627
Phone: (949) 631-3309
Email: bdubay@impact-information.com
Website: http://www.impact-information.com

1921—E. L. Thorndike and the Familiar Word

Introduction

Edward L. Thorndike
1874-1949

E DWARD L. THORNDIKE is recognized as one of the most important psychologists of the 20th century. A student of psychologist and philosopher William James in Harvard, he brought the new approach of the Applied Psychology to the classroom. Along with American philosopher and educator John Dewey, he was to dominate educational theory and practice around the world for 50 years.

During the 1920s, two major trends stimulated a new interest in readability:

1. A changing school population, especially an increase in "first generation" secondary school students, the children of immigrants. Teachers reported that these students found textbooks too difficult.

2. The growing use of scientific tools for studying and objectively measuring educational problems.

One such tool, Thorndike's *The Teacher's Word Book*, which came out in 1921, was the first extensive listing of words in English by frequency.

It is hard to overestimate the effects of this book on education and communications. It provided teachers with an objective means for measuring the difficulty of words and texts. It laid the foundation for almost all the research on readability that would follow, including the first readability formulas.

Thorndike, while doing research for the Teachers College in Columbia University, noticed that teachers of languages in Germany and Russia were using word counts to match texts with students. The more frequent a word is used, they found, the more familiar it is and the easier to use.

Around 1911, Thorndike began to count the frequency of words in English texts. In 1921, he published *The Teacher's Word Book,* which listed 10,000 words by frequency of use. Each word was given a score or "credit", depending on how frequently it is used. A word like *the* or *to* has a credit of 208, while *advisable* has a credit of 3. In 1932, he followed up with *A Teacher's Word Book of 20,000 Words*, and, in 1944 with Irving Lorge, *A Teacher's Word Book of 30,000 Words*.

A vocabulary test on the meaning of words is the strongest predictor of verbal and abstract intellectual development. The

knowledge of words has always been a strong measure of a reader's development, reading comprehension, and verbal intelligence. Chall and Dale wrote in 1995, "It is no accident that vocabulary is also a strong predictor of text difficulty."

It happens that the first words we learn are the simplest and shortest. These first, easy words are also the words we use most frequently. Most people do not realize the extent of this frequency. Twenty-five percent of the 67,200 words used in the 24 life stories written by university freshmen consisted of these ten words: *the, I, and, to, was, my, in, of, a*, and *it*. The first 100 most frequent words make up almost half of all written material. The first 300 words make up about 65 percent of it.

Educators, publishers, and teachers still use Thorndike's word-frequency lists to evaluate reading materials for schools. After Thorndike, there was extensive research on vocabulary.

In 1968, psychologist George Klare wrote about the central role vocabulary plays in reading skill, "Not only do humans tend to used some words much more often than others, they recognize more frequent words more rapidly than less frequent, prefer them, and understand and learn them more readily. It is not surprising, therefore, that this variable has such a central role in the measurement of readability."

The following article, reprinted here in full, accompanied the publication of the first *The Teachers Word Book* in 1921. In the introduction to the book, Thorndike refers readers to the article for the background of the list and how to use it.:

> A full account of the methods by which this list was selected, of the reliability of the credits attached to the words, and of the uses to which the list may be put, will be found in an article of thirty-seven pages, entitled "Word Knowledge in the Elementary School," by E. L. Thorndike, published in the Teachers College Record for September, 1921.

The article also shows Thorndike's beliefs about teaching and how students learn to read. He shows how to use the list to teach high school students and those learning English as a second language.

His comments on why youngsters prefer to read "trash" rather than current affairs goes to the lack of "suitable reading materials" for readers of all ages and backgrounds. He wrote:

> It is commonly assumed that children and adults prefer trashy stories in large measure because they are more exciting and more stimulating in respect to sex. There is, however, reason to believe that greater

ease of reading in respect to vocabulary, construction, and facts, is a very important cause of preference. A count of the vocabulary of "best sellers" and a summary of it in terms of our list would thus be very instructive.

In the mid-1930s, Edgar Dale and Ralph Tyler claimed that Thorndike's lists did not take into account the different meanings that some words have.

Edgar Dale began working on his own lists. In 1944, Irving Lorge published a formula using Dale's "short list." In 1948, teaming up with Jeanne Chall of the Harvard Reading Laboratory, Dale published a readability formula using his "long list" of 3,000 easy words.

In 1981, Dale and Joseph O'Rourke published their 40,000-word *Living Word Vocabulary*, a masterful tribute to Thorndike's original and prestigious work.

<div align="center">—WHD</div>

Word Knowledge in the Elementary School

By EDWARD L. THORNDIKE
Professor of Education, Teachers College

CONSIDER these simple questions: How many English words should the ordinary boy or girl know the meanings of at the end of Grade 8? Which words should all or nearly all pupils know at that stage? In what grades and in what connections should they be learned? If gifted and experienced teachers, supervisors, and authors of courses of study were to give answers to these questions, the answers would vary enormously. Nobody, in fact, knows the answers with even roughly approximate correctness. Nor is our condition better if we free the questions from the ambiguity of "ordinary boy or girl" and specify any particular child or type of child. The answers would still vary widely, and all of them might well be wrong. Nor is our condition better if we describe fully what else the pupil is to know and assign, say, 8.375 per cent of his time and energy to this particular feature of his education. We still cannot answer with any surety.

These questions, though concerned with details, and less inspiring than broad questions about health, morality, or citizenship, are important, as indeed all competent workers in the science of education will now admit.

It appears that one notable cause of our inability to answer them correctly is our lack of knowledge of the frequency of occurrence of words in the talk our pupil and graduate will or should hear, and the books, articles, letters, and the like, which he will or should read. Just as word counts of such material as the pupil may need to write are instructive in the pedagogy of spelling, so word counts of such material as the pupil may need to understand will be instructive in the pedagogy of reading, and indeed of all the school subjects which are presented with the aid of language.

So for about ten years I have made such counts as I could. They are as follows:

THE WORD COUNTS AND CREDITS

CHILDREN'S READING

1. Every word in Chapters 10 to .19 (and a few pages more) of *Black Beauty,* one of the most popular books if not the most popular book for children about nine years of age. About 11,500 words in all. Credits given as follows: 1 to 4 occurrences, 1, 5 to 9, 2; 10 to 19, 3; 20 to 29, 4; 30 to 49, 5; 50 or over, 6. That is, if a word occurred once, twice, three times, or four times in these chapters of *Black Beauty,* it was given a credit of 1; if it occurred from 5 to 9 times, it was given a credit of 2, etc.[1]

---

[1] The reader is asked to accept arbitrarily these credits since an explanation of the method by which

2. Every word in Chapters 1, 2 and 3 (and 14 lines of Chapter 4) of *Little Women..* About 13,000 words in all. Credits: As in No. 1.

3. Every word in Chapters 1 to 5 and part of 6 of *Treasure Island.* About 13,000 words in all. Credits: As in No. 1.

4. Every word in *Scrooge's Christmas,* a selection from *The Christmas Carol* as reprinted in a school reader. About 8,000 words in all. Credits: As in No. 1.

5. Every word in Irving's *Sleepy Hollow.* About 13,000 words in all. Credits: As in No. 1.

6. Every word in one issue of the *Youth's Companion,* omitting advertisements and fine print. About 25,000 words in all. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 19, 3; 20 to 29, 4; 30 or over, 5.

7. Every word in the fifty-six selections found by Hosic to be the commonest features of school readers.[2] About 27,000 words in all. These were divided into two halves, referred to hereafter as Hosic A and Hosic B. Credits for Hosic A were; 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 19, 3;

---

they were obtained is too involved to be given here.

[2] It would probably be more scientific to enter these along with items 22 to 30 as Standard Literature; and, in another connection, this is done. The list is as follows:

Allingham, The Fairies.	Kingsley, The Lost Doll.
Anderson, The Steadfast Tin Soldier.	Kipling, Toomai of the Elephants.
Anderson, The Ugly Duckling.	Lincoln, Address at Gettysburg.
Bjornson, The Tree.	Longfellow, The Arrow and the Song.
Browning, Home Thoughts from Abroad	Longfellow, The Village Blacksmith.
Browning, How They Brought the Good News from Ghent to Aix	Macaulay, Horatius
Browning, Pied Piper of Hamen	Miller, Columbus.
	Moore, A Visit From St. Nicholas.
Bryant, Planting of the Apple Tree.	Saxe, The Blind Man and the Elephant.
Bryant, Robert of Lincoln.	Scott, Loshinvar
Bryant, Song of Marion's Men	Scott, Love of Country (Lives there a man)
Bryant, To a Waterfowl.	
Burns, A Man's a Man.	Shakespeare, Orpheus with His Lute.
Carlyle, Today.	Shakespeare, Under the Greenwood Tree.
Drake, The American Flag	Sherman, Daisies.
Emerson, Concord Hymn	Smith, America.
Emerson, The Mountain and the Squirrel	Stevensen, My Shadow.
Franklin, Proverbs.	Stevensen, The Wind.
Franklin, The Whistle.	Tennyson, The Brook.
Gray, Elegy.	Tennyson, The Bugle Song.
Hemans, Landing of the Pilgrims.	Tennyson, The Charge of the Light Brigade
Holmes, The Chambered Nautilus.	
Holmes, The Deacon's Masterpiece.	Tennyson, Sir Galahad.
Holmes, Old Ironsides.	Thaxter, The Sandpiper.
Hunt, Abu ben Adhem.	Whitman, O Captain, My Captain.
Ingelow, Seven Times One.	Whittier, The Barefoot Boy.
Irving, Rip Van Winkle.	Wolfe, Burial of Sir John Moore.
Key, The Star Spangled Banner.	Wordsworth, Daffodils (I wandered lonely).

20 to 29, 4; 30 or over, 5. Credits for Hosic B were the same.

8. Every word in ten primers or first readers (for this material in its original form, I am indebted to Mr. C. N. Smith and the teachers who aided him). As a very rough estimate, we may take 80,000 words. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 19, 3; 20 to 29, 4; 30 to 39, 5; 40 to 49, 6; 50 or over, or fewer if present in all 10 books, 7.

9. Every word in ten second readers (for this material in its original form I am indebted to the University of Iowa and E. T, Hooch). About 150,000 words in all. Credits: Same as for No. 8.

10. Every word in ten third readers (for the material in its original form I am indebted to the University of Iowa and W. S. Miller). 283,000 words in all. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 19, 3; 20 to 29, 4; 30 to 39, 5; 40 to 74, 6; 75 or over, or fewer if present in all 10 books, 7.

11. Every word in Book One of the *Thorndike Arithmetics,* edition of 1917. About 32,000 words in all. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 or over, 4.

12. Every word in Book One of the original edition of the *Young and Jackson Arithmetics.* About 35,000 words in all. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 19, 3; 20 to 29, 4; 30 or over, 5.

13. Every word in Brigham and McFarlane, *Essentials of Geography, First Book,* first edition, pages 26 to 256. About 83,000 words in all. Credits: As for No. 12.

14. Every word in Straubenmuller, *A Home Geography of New York City.* About 37,000 words in all. Credits: As for No. 12

15 Every word in Thwaites and Kendal, *History of the United States,* edition of 1914, pages 26-100 inclusive. About 25,000 words in all. Credits: As for No. 12.

16. Every word in Forman, *History of the United States,* pages 101,102,121, 122, etc., to the end. About 17,000 words in all. , Credits: As for No. 12.

17. Every word in the first 25 pages of each of five standard First Books in Geography. About 40,000 words in all. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 29, 3; 30 to 49, 4; 50 or over, 5.

18. Every word in the first 25 pages of each of three text-books in United States History. About 20,000 words in all. Credits: As in No. 12.

19. The vocabularies of ten books on Elementary French. Credits: A credit of 1 for a word found in 1 or 2 of the ten; 2 for a word found in 3 or 4 of the ten; 3 for a word found in 5 or 6 of the ten; 4 for a word found in 7 or 8 of the ten; 5 for a word found in 9 or 10 of the ten.

20. The vocabularies of ten books on Elementary German. Credits: As in No. 19.

21. The vocabularies of five books on Elementary Spanish. Credits: o for a word found in only one of the five; 2 for a word found in 2 of the 5; 3 for a word found in 3 of the 5; 4 for a word found in 4 of the 5; and 5 for a

word found in all.[3]

22. Strong, *Exhaustive Concordance of the Bible.* About 900,000 words in all. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 occurrences up to a column-full,[4] 3; a column-full but not two columns-full, 4; two columns or over, 5.

23. John Bartlett, *Concordance to the Dramatic Works of Shakespeare.* (The supplementary concordance to the poems was not used.) About 925,000 words in all. Credits: 3 to 9 occurrences, i; 10 occurrences up to a column-full, 2; a column-full but not two columns-full, 3; two columns or more, 4.

24. Lane Cooper, *Concordance to Wordsworth.* About 400,000 words. Credits: As for No. 23.

25. Baker, *Concordance to Tennyson's Poetical and Dramatic Works,* Part II, on the Dramatic Works. About 120,000 words. Credits: As for No. 23.

26. Baker, *Concordance to Tennyson's Poetical and Dramatic Works,* Part I, on the Poetical Works. About 200,000 words. Credits: As for No. 23.

27. Neve, *Concordance to the Poetical Works of William Cowper.* About 200,000 words. Credits: As for No. 23.

28. Abbott, *Concordance to Pope* (this covers only a part of Pope's poetical works). About 90,000 words. Credits: As for No. 23.

29. Bradshaw, *Concordance to the Poetical Works of John Milton.* About 130,000 words. Credits: As for No. 23.

30. Bartlett, *Familiar Quotations,* 10th edition, pages 42-62, 201-220, 301-320, 401-420, 501-520, 601-620, 701-720, 801-820, 901-920 and 1001-1020; footnotes and all save the quotations themselves being omitted. About 32,000 words in all. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 19, 3; 20 to 29, 4; 30 or over, 5.

COMMON FACTS AND TRADES

31. The United States Constitution and the Declaration of Independence. About 8,000 words in all. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 19, 3; 20 or more, 4. The words of Common List A[5] were not counted but

---

[3] 1 am indebted.to Professor Bagster-Collins for the Spanish data and to Mr. Ben Wood for their translation.

[4] A column means about 116 occurrences in the Bible; about 90 in Shakespeare; about 92 in Wordsworth; about 72 in Tennyson's plays, or in his poems; about 80 in Milton; 82 in Pope; and about 55 in Cowper.

[5] The words of Common List A were as follows:

a	can	good	know	night	since	very
about	come	got		no	so	
after	could	great		not	some	
again			last	now	soon	was
against			let		such	way
ago	day	had	like			we
also	did	has	little	of		well
all	do	have	long	on	than	were
an	does	he		only	that	what

each was given a credit of 4. This same procedure was followed in Nos. 32 to 36 inclusive. Credits for *today, too, two, until* and *us* were also assigned by estimate in Nos. 31 to 36, as: 1, 2, 3, 2 and 1 respectively.

32. Farmer, *A New Book of Cookery,* pages 1, 11, 21, 31, etc. About 4,000 words in all. Credits: As in No. 31.

33. Allington, *Practical Seiving and Dressmaking,* pages 1, 11, 21, 31, etc. About 6,000 words in all. Credits: As in No. 31.

34. *Garden and Farm Almanac for 1914,* pages 7, 9, 11, 18, 19 to 32; 88 to 120; 132, 145 to 151 and 156 (but with a few omissions of notes, statistics, etc.). About 17,000 words in all. Credits: As in No. 31.

35. Five pages containing the United States postal regulations in popular form. About 1,700 words in all. Credits: As in No. 31.

36. The first ten questions and answers in each of thirty-one trade tests chosen from those published by the Committee on Classification of Personnel in the Army. About 5,000 words in all. Credits: As in No. 31.

37. The first word of each entry in the indexes of three large mail-order catalogues. The fullest index was taken as a basis. Credits: A credit of I was given if the word was in this fullest index only. If it was also in one other, a credit of 2 was given. If it was in all three, a credit of 3 was given.

NEWSPAPER READING

38. The 44,000 words of the Eldridge count from a Buffalo paper. Eldridge did not include proper names or numerals. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 19, 3; 20 to 29,4; 30 to 49,5; 50 to 99,6; 100^999,7; 1000 or over, 8. For numerals and proper names, credit same as in No. 39.

39. Selections from the newspaper of an eastern city, *The Examiner-New Era* of Lancaster, Pa., taken from six issues spread over the year 1920 and taken at random from page 1, col. 1; page 1, col. 2; etc. About 40,000 words in all. Credits: 1 to 4 occurrences, 1; 5 to 9, 2; 10 to 19, 3; 20 to 39, 41 30 or over, 5. The words of Common List A were not counted but given each a credit of 5. *Today, too, two, until,* and *us* were assigned credits of 1, 2, 3, 2, and 3, respectively.

and	down	her		one	the	when
any		here	made	or	thjeir	where
are		him	make	other	them	which
as		his	man	our	then	while
at	each	how	man	out	there	who
	every		matter	over	these	why
			may		they	will
be		if	me		through	with
been	few	in	might	place	thing	woman
before	first	into	more		this	work
best	for	is	most		time	would
better	from	it	much	same	to	
but		its	my	say		
				see		you
	get			she	up	your
	go	just	new	should		

8

40. The Cook-O'Shea list of frequencies derived from a count of 200,000 words of private correspondence was used. Credits: 2 to 9 occurrences, 1; 10 to 19, 2; 20 to 39, 3; 40 or over in their List III, 4; 40 to 49 in their Lists I and II, 4; 50 to 99 in Lists I and II, 5; 100 and over in List II, l00 to 299 in List I, 6; 300 to 999 in List I, 7; 1000 to 1999 in List I, 8; 2000 and over in List I, 9.

41. The Anderson list[6] of frequencies derived from students' counts of over 360,000 words in business and private correspondence was used. He reports the results for only 3087 words, occurring with a total frequency of 5 or more, and occurring in at least three of the six groups into which he divided his material (Professional, Business, Domestic, Miscellaneous, Personal and Farmers'). Credits: 5 to 19, 2; 20 to 49, 3; 50 to 99, 4; l00 to 499, 5; 500 to 999, 6; 1000 or over, 7.

In connection with the counts listed above, certain corrections and additions have been made where the need for them was obvious.

For example: the concordances omit altogether certain very common words like *the, and, of, it.* The credit for these was estimated, and usually with no risk of error, it being well above the maximum credits as stated on page 339. Sometimes the concordances give samples of a word's usage, but not complete inventories, and it is to be feared that they sometimes do this without informing the student of the fact. These cases have been treated as well as might be without elaborate study. There are doubtless errors on the part of myself and of my assistants in the counting and crediting and tabulating; but such probably act for the most part as variable errors. Where they do not, the most probable resulting error is an insufficient importance for abbreviations, and, to a less extent, for proper names. In the case of the counts from the concordances, to avoid tabulating data of no probable use in the final report, the custom was followed of not entering rare proper names. The material of the two spelling lists, and of the vocabularies of French, German, and Spanish First-Year Books, also does not include all the proper names used. One of the spelling lists includes all except those of "towns of less than 10,000 inhabitants; all strictly local publications; organizations and streets; and all family names, except those of historical characters or of men in the public eye"; the other includes only the names of the days of the week, months, and nationalities. So it may well be that on grounds of frequency alone some of the words of our last thousand or so should be displaced by names of persons and places. However, on grounds of real importance, this "error" may well have brought us nearer the truth.

The concordances, the Eldridge count, the Cook-O'Shea count, and the vocabularies rarely or never include abbreviations. Although Dr. Anderson instructed his helpers to count all abbreviations, it seems almost certain from the actual results that they did not do so. *Etc., Mr.,* and *Mrs.* receive credit

---

[6] W. N. Anderson, *The Determination of a Spelling Vocabulary Based upon Written Correspondence,* Aug. 1917. Manuscript in the library of Iowa State University.

from him; but *doz., ft., kr., in., pt., pk., St.,* and the like, do not. I fear, therefore, that his instructions were often disregarded. The sums of credits in the case of abbreviations in my list are thus unreliable and in general too low.

The Cook-O'Shea list does not make separate entry for words like *am, are, ate, been, began, begun, is, was*, and the like. I have not corrected for this because it seems undesirable to tamper with lists selected by others, even if their procedure seems indefensible. The vocabularies of the first-year foreign language books also are taken as they stand, although they omit many derived words which really are used in the body of the text. The general effect of the omissions in them and in the Cook-O'Shea list is to reduce the credit of derived forms. This is perhaps desirable.

I shall not defend the above as an especially good selection of material, though it is by no means a bad one. Some of the items were chosen partly for other reasons than the rating of words for frequency and range of occurrence. Some were chosen, notably the concordances, because of the amount of information gained per hour or dollar spent. Nor shall I defend the system of credits used above as an especially wise system of weighting frequency and range of occurrence, on the whole or within one sort of matter, such as children's literature. Indeed I am sure that I could now improve it. The general principle of weighting range as well as mere number of occurrences is sound, and the final result from the cleverest weighting would probably not be very much better than that secured here. Let us postpone further critical study of the counts until we have inspected some of the results.

GENERAL RESULTS

First, it should be noted that a plural formed by adding s was not counted separately, but entered under the singular form. The same procedure was followed, except in certain cases for special reasons, with plurals where *y* is replaced by *ies*, adverbs formed from adjectives by adding *ly*, comparatives formed by adding er or r, superlatives formed by adding *est* or *st*, verb forms derived by adding *s*, *ed* or *d*,[7] *n*, and *ing*, in cases where the derived form would probably be easily read and understood by the pupil when he experienced it, if he knew the primary word. Adjectives formed from proper nouns by adding *n* are also, as a rule, counted with the noun. For example:

days	nights	cherries	counted under	day	night	cherry
gladly	proudly	counted under	glad	proud		
stronger	weakest	counted under	strong	weak		
shows	showed	shown	showing	counted under	show	
Russian	Bolivian	counted under	Russia	Bolivia		

This greatly reduces the number of entries and seems desirable for our purpose. If a count is to be used as a guide to instruction in spelling, on the contrary, we need estimates of each such derived form.

It was permissible to omit from entry rare names of persons and places. Apart from this, every word or abbreviation was to be counted and was counted in most of my work. In some that was done without my direct supervision abbreviations were, I fear, somewhat neglected. In the concordances and vocabularies of text-books in French, etc., and in the counts

---

7 Including changes of y to *ies, ier, iest* and ied.

by Eldridge, Cook and O'Shea, Miller, Housh and Anderson, abbreviations have been very largely neglected.

In all, over 20,000 words or abbreviations received a credit of 1or more.

It should be made clear at this point that the credit assigned to a given word in each of the several counts depended upon the number of times the given word occurred in the particular book or article. It follows, therefore, that a given word might have a credit of 4 in the count of one book and a credit of 5 or more in the count of a second book. As an illustration in the count of *Black Beauty* the word *angel* has a credit of 1, in the Bible this same word has a credit of 5, while in Tennyson's plays it is credited as 2. It is evident that the credit received by the word *angel* in *Black Beauty* would not be a sufficient indication of its importance or occurrence in all the books counted.

TABLE I

	and	*angel*	*anger*	*angle*	*anguish*	*animal*
Bible	5	5	5	1	3	
Wordsworth	4	3	2		2	2
Tennyson, plays	4	2	2			
Cowper	4	2				I
Familiar Quotations	5	4	1	1		1
Hosic A	5	1				1
Thwaites-Kendal History, 75 pp.	5		1			2
Brigham-McFarlane Geography, 230 pp.	5			1		5
Thorndike Arithmetics	5			2		
Geography, beginning, 125 pp.	5					4
First Readers or Primers, 10	7					1
Second Readers, 10	7		3			6
Third Readers, 10	7	4	2			7
Shakespeare	4	2	2	2	1	1
Milton	4	3	2		2	1
Tennyson, poems	4	2	2	1	1	
Pope	4	2	1			
Youth's Companion	5		1			3
Hosic, B.	5	1				
Forman History, 70 pp.	5					1
Straubenmuller Geography, 218 pp.	5	1	1	1		3
Young-Jackson Arithmetic	5					
History.beginning, 75pp.	5	1				1
French 1st-year books, 10	5	2	2		1	4
German 1st-year books, 10	5		3			5
Spanish 1st-year books, 5	5		2			4
Cook-O'Shea Correspondence	7	1				2
Anderson Correspondence	7					3
Cook book	4	2				1
Sewing book	4					
Trade tests	4			1		
Farm almanac	4			2		4
Mail-order catalogues	10			2		2
Postal regulations	4					
Constitution and Declaration	4					
Eldridge newspaper	8		1			1
Examiner newspaper	5		1	1		2
Black Beauty	6	1	1			1
Little Women	6	1			1	
Scrooge	6					
Treasure Island	6		1			
Sleepy Hollow	6					1
SUM OF CREDITS	210	40	36	15	11	70

It became necessary therefore to devise some means of designating the importance of a word by adding together the credits received by that word in all the counts. The result thus obtained for each word is called the "sum of credits" or "credit sum."

We thus have for each of the 20,000 words which received a credit of 1 or a record like that shown in Table I for the words *and, angel, anger, angle, anguish,* and *animal.* This record is summarized in a number, the "sum of credits," found at the bottom of each column. Thus the word *and* has a "credit sum" of 210; the word *angel*, 40; *anger,* 36; *angle,* 15; *anguish,* 11; *animal,* 70.

TABLE II

DISTRIBUTION OF WORDS IN RESPECT TO FREQUENCY AND RANGE OF OCCURRENCE

Sum of Credit of:	
0	50,000 (approx)
1 and 2	10,000 (approx)
3 and 4	2,878
5 and 6	1,569
7 and 8	1,074
9 and 10	759
11 and 12	565
13 and 14	478
15 and 16	413
17 and 18	291
19 and 20	248
21 and 22	256
23 and 24	189
25 and 26	198
27 and 28	181
29 and 30	152
31 and 32	100
33 and 34	128
35 and 36	103
37 and 38	96
39 and 40	96
41 and 42	74
43 and 44	69
45 and 46	65
47 and 48	66
49 and 50	60
51 and 52	49
53 and 54	40
55 and 56	50
57 and 58	59
59 and 60	47

with 715 cases spread from 61 to 211 with decreasing frequency.

The highest credit sum found among the 20,000 words, which belongs to the word *in,* is 211; the word which is 500th in rank has a credit sum of 75; the l000th word has 49 as a credit sum; the 1500th word has 36 as a credit sum; the 2000th word has 28 as a credit sum; the 2500th has 23 as a credit sum. The credit sums of the next 2500 words (to the 5000th) range between 23 and 10; those of the next 5000 range between 9 and 3; and there are about

10,000 words with credit sums of only 1 or 2. That is, the distribution of words in respect to frequency and range of occurrence is as shown in Table II.

The meaning of these credit sums can be realized by a rapid inspection of the list below which gives four or five words taken at random from those receiving the following credit sums: 200, 180, 160,140,120,100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 8, 6, and 4.

Sum of Credits	Samples
200	for it on one
180	an no or some were
160	go great her new well
140	did just live say
120	eye light right white woman
100	mean nothing order sea seem
90	almost ground mark rich table
80	buy dead hat known lead
70	able act animal appear cannot
60	class cup distance doubt equal
50	bit breakfast clothing (e, ed) crowd crown
40	advantage ah angel aunt bend
30	advice avenue ax (axe) beach bean
20	accordingly acquire ahead ample anyone
10	abate abolish abound accommodate acorn
8	absolve absorb absurd accommodation accordance
6	abstain adjacent adjective admirable admirer
4	abashed abridge accessory Adams adjudge

In the above list the exact credit sums of some of the words are not multiples of ten. These are: *for,* 201; *it* and *one,* 199; *an, no, or, and some,* 181; *great,* 159; *her,*161; *live and water,* 139.

The general nature of the list may be realized from another point of view by an inspection of the following words which mark the ends of the first, second, third, fourth, and fifth thousands:

End of 1st 1000.   Sum of credits, 49

chain circle condition date discover double escape fancy fence fool grave health history jump mail merry mighty noise

End of 2nd 1000.   Sum of credits, 28

addition Africa Asia authority bosom brass bride broke brow button complain congress cottage create crime crush curl deceive defend devil Dick doll eagle echo eleven farewell

End of 3rd 1000.   Sum of credits, 19

Alice amaze argument Arthur banish barren bat belief beloved bloody bower breadth Britain Caesar capable cedar cloudy combination commend confusion conquest consequence construct crept dame

End of 4th 1000,   Sum of credits, 13

accursed actor Adam affair afflict airy Alexander almighty alms ambassador amid applause arrangement attorney auto awoke bait balm beguile Benjamin Bible boundless bounty brace Carl

End of 5th 1000.   Sum of credits, 10

abate abolish abound accommodate acorn admission adoption adversary affectionate . agency aisle ale allegiance aloft anniversary annoy anoint antiquity anvil apparel aspect asunder Athens attic attitude

THE RESULTS IN DETAIL

The most frequently occurring 10,000 words of our list are printed in alphabetical order with a rating of the importance of each as indicated by its sum of credits, and in a form convenient for use by teachers and others in The Teacher's Word *Book*[8]. When the term "list" is used hereafter in this article, it will mean this list of 10,000 words.

The inclusion of all the words obtaining 3 as a sum of credits would extend the list to nearly 11,000. So I have eliminated those "3" words which seemed in the combined opinion of five judges to be the least important.[9] I have also eliminated some words[10] receiving credits so exclusively from one source that it seemed absolutely certain that a more extensive count would not include them. With this exception the list represents the unmodified results of the counts.

THE ADEQUACY OF THE COUNTS

Before drawing conclusions from the facts of the list and planning uses for it, we must deal with certain questions about the quantitative and qualitative adequacy of the counts.

The question of quantitative adequacy may best be stated thus. What are the probable divergences of the present list from one that would be made up from counts of the same kind of selections in the same proportions, but hundreds of times as extensive? The full answer is given in Table III. The gist of it is that the present counts are adequate to determine the first one thousand words with a small probable error, and the next four thousand well enough for many educational purposes, and the last five thousand to an extent that is useful, though far from accurate.

---

[8] This is a book of 132 pages so arranged that relevant facts about any of these words can be entered. In the case of the 5000 most important words, the credit sum is followed by a number and letter stating in which thousand and in which half thereof the word belongs. Thus "43 2a" means that the word has a credit sum of 43 and is in the first half of the second thousand, "21 3b" means that the word has an importance of 21 and is in the second half of the third thousand. Within the first five hundred there is a further distinction into hundreds, lai meaning that the word is one of the first hundred; ia2 meaning that it is in the second hundred; 133 meaning that it is in the third hundred, and so on. The book is obtainable from the Bureau of Publications, Teachers College, 525 West 120 Street, New York City.

[9] As a consequence the "3" words remaining in the list are probably on the whole nearly or quite as important as the "4" words.

[10] Such were, for example: Agrippa, Albion, Arabella, Castoria, chee, chiffonette, contemn, Coronado, Cyprus, Dagon, drave, Eli, Enoch, Ephriam, Galatians, Gardinar, Hepsy. Jephthah, Jip, linnet, ope, pate, pied, Titus, trow, Tubal, tweet.

TABLE III

THE EXTENT TO WHICH THE RESULTS OF AN INFINITELY EXTENSIVE
COUNT OF THE SAME SORT MAY BE EXPECTED TO DIVERGE FROM
THE RESULTS OBTAINED FROM THE PRESENT COUNT.

Sum of Credits	P.E. of sum of Credits	Median probable displacement from true position in terms of number of words	Sum of Credits	P.E. of Sum of Credits	Median probably displacement	Sum of Credits	P.E. of Sum of Credits	Median probably displacement
4	1.50	1900	54	3.70	90	100	4.20	25
6	1.60	1150	6	3.74	86	105	4.20	23
8	1.75	850	8	3.78	83	110	4.20	20
10	1.87	675	60	3.81	80	115	4.20	17
2	2.00	550	2	3.84	76	120	4.20	15
4	2.11	465	4	3.86	72	125	4.20	13
6	2.22	405	6	3.88	69	130	4.20	12
8	2.33	360	8	3.90	65	135	4.20	8
20	2.43	325	70	3.92	62	140	4.20	10
2	2.52	300	2	3.94	58	145	4.20	10
4	2.60	275	4	3.96	55	150	4.20	9
6	2.68	250	6	3.98	51	155	4.20	9
8	2.76	225	8	4.00	48	160	4.20	8
30	2.84	202	80	4.02	45	165	4.20	8
2	2.92	185	2	4.04	42	170	4.20	7
4	3.00	170	4	4.06	40	175	4.20	7
6	3.08	157	6	4.08	38	180	4.20	6
8	3.16	145	8	4.10	36	185	4.20	6
40	3.24	135	90	4.12	34	190	4.20	6
2	3.32	126	2	4.14	34	195	4.20	5
4	3.40	118	4	4.16	30	200	4.20	5
6	3.47	111	6	4.18	29	205	4.20	5
8	3.54	105	8	4.20	27	210	4.20	5
50	3.60	99						
2	3.65	94						

Of the words put in the first 500 of our list, about 25 words would, by an infinitely extensive count, be put lower. Of the words in the second 500, about 31 would be put in the next higher 500, and about 57 in the next lower 500. Of the words in the top 5000, about 350 would be put lower. Of the words in the entire 10,000, about 1000 would be displaced downward, and replaced by others from the 60,000 not listed here, but almost all of them would come in the next few thousands, so that it is not a serious error to regard them as belonging within the 10,000.

There are 957 chances in 1000 that a word of credit above 100 would not by the infinite count receive a credit 13 higher or lower than it does; that a word of credit 50 would not receive a credit it higher or lower than it does; that a word of credit 25 would not receive a credit 8 higher or lower than it does; that a word of credit 20 would not receive a credit 7½ more or less than it does; and that a word of credit 15 would not receive a credit 6½ more or less than it does.

It is impossible to measure the qualitative adequacy of the counts and

credits except by further very extensive counts. For example, only further counts of children's reading can decide whether our selections are an adequate sampling of the matter children do or should read. Only elaborate counts of correspondence will decide whether the Cook-O'Shea and Anderson counts are fairly representative of correspondence in general as to its quality. Moreover, there will always be some room for diversity of opinion as to the balance between what is read and what we should consider important to be read. As to the relative weight to attach to children's reading, correspondence, literary classics, the Bible, the newspaper, and so on, the diversity of opinion exists and will long remain. I hope some time to be able to publish all the original counts so that any competent person can use them with any weights that he thinks desirable.

In various details notable changes might be made. The importance .of-*hath* and *doth,* for example, depends almost entirely on the weight given to the reading of the Bible. If a revised version of it used *has* and *does,* they would drop down or out of our list. The importance of *acid, ampère, atom,* and the like, depends on the weight attached to reading on scientific and industrial topics as contrasted with "literary" reading.

The scientific student of the list can use the facts given about the counts and credits to amend it where he thinks wise. So it seems undesirable here to deal at any length with the principles of selection and weighting. I shall therefore simply note four principles, and illustrate in representative cases how the selection and weighting work out.

Pains were taken to preserve some fair balance between importance for the boy and girl during the elementary school course, and importance for them after graduation. Rather large weight was given to appearance in a variety of sorts of reading, and relatively small weight to a large number of occurrences in any one sort alone. A balance was sought between *reliability* (attainable by a large count, as in the concordances) and *significance*. A word appearing in a small count is, other things being equal, more important, of course, than one appearing in a large count.

TABLE IV

CREDITS RECEIVED FROM DIFFERENT SOURCES BY EACH OF FIFTY
REPRESENTATIVE WORDS.

	Children's Literature	Bible and English Classics	Textbooks and Vocabularies	Common Facts and Trades	Newspapers	Correspondence	Total Sum of Credits
mean	22	25	31	4	7	11	100
nothing	28	29	24	2	7	10	100
sea	24	34	32	0	6	4	100
seem	28	35	19	I	6	ii	100
spring	22	26	30	8	5	9	100
almost	21	22	28	3	6	10	90
ground	22	25	27	6	7	3	90
mark	16	25	29	11	4	5	90
rich	18	29	30	4	4	5	90
table	23	18	29	10	4	6	90
buy	20	15	27	1	6	11	So
dead	18	34	16	1	5	6	80
hot	17	16	22	13	3	9	80
lead	13	25	24	7	7	4	80
picture	15	14	33	3	5	10	80
able	19	12	25	1	3	10	70
act	14	20	19	*5*	5	7	70
animal	18	8	29	7	3	5	70
appear	13	26	17	2	7	5	70
cannot	21	22	12	I	2	12	70
class	5	2	31	6	6	10	60
cup	14	18	13	7	4	4	60
distance	10	14	27	4	3	2	60
doubt	7	24	17	0	5	7	60
equal	3	19	25	7	2	4	60
bit	16	12	11	4	3	4	50
breakfast	16	6	15	4	1	8	50
demand	7	17	14	3	5	4	50
crowd	7	17	14	1	6	5	50
crown	10	25	7	4	3	1	50
advantage	1	10	17	1	4	7	40
ah	14	21	3	0	2	0	40
angel	6	27	4	2	0	1	40
aunt	13	4	11	0	2	10	40
brick	11	5	13	6	2	3	40
advice	3	8	12	I	2	4	30
avenue	4	3	13	0	5	5	30
ax (axe)	9	II	7	3	0	o	30
beach	8	9	7	1	3	2	30
bean	8	2	7	6	2	5	30
accordingly	2	6	7	1	I	3	20
acquire	0	3	12	1	I	3	2O
ahead	6	0	7	0	I	6	2O
ample	I	12	2	1	2	2	2O
anyone	7	0	7	0	I	5	20
abate	0	8	2	0	0	0	10
abolish	0	4	3	I	2	0	10
abound	0	8	I	0	I	0	10
accommodate	1	1	4	I	I	2	10
acorn	5	2	3	0	0	0	10

TABLE V

RANK ORDER OF THE FIFTY WORDS OF TABLE IV, ACCORDING TO THE
CREDIT RECEIVED FROM EACH SOURCE

	Children's Literature	Bible and English Classics	Textbooks and Vocabularies	Common Facts and Trades	Newspapers	Correspondence	Total Sum of Credits
mean	6	11	3½	16½	3	3	3
nothing	1½	4½	l6½	25	3	7½	3
sea	3	2½	2	45½	8½	31	3
seem	1½	1	19½	33½	8½	3	3
spring	6	7½	5½	4	15	11½	3
almost	8½	15½	10	21½	8½	7½	8
ground	6	11	12	10½	3	36½	8
mark	17	11	8	2	21	23½	8
rich	13	4½	5½	16½	21	23½	8
table	4	20½	8	3	21	18	8
buy	10	25	12	33½	8½	3	13
dead	13	2½	24	33½	15	18	13
hot	15	24	18	1	27	11½	13
lead	24	11	l6½	6½	3	31	13
picture	19	26½	1	21½	15	7½	13
able	11	29	14½	33½	27	7½	18
act	21	18	19½	13	15	15	18
animal	13	35½	8	6½	27	23½	18
appear	24	7½	22	25	3	23½	18
cannot	8½	I5½	32	33½	35	1	18
class	38½	46	3½	10½	8½	7½	23
cup	21	20½	29	6½	21	31	23
distance	27½	26½	12	16½	27	40½	23
doubt	33½	14	22	45½	15	15	23
equal	41½	19	14½	6½	35	31	23
bit	17	29	34½	16½	27	31	28
breakfast	17	38½	25	16½	43	13	28
demand	33½	22½	26½	21½	15	31	28
crowd	33½	22%	26^	33½	8½	23½	28
crown	27½	11	39	16½	27	43½	28
advantage	45	32	22	33½	21	15	33
ah	21	17	46	45½	35	47½	33
angel	36½	6	43½	25	48½	43½	33
aunt	24	41½	34½	45½	35	7½	33
brick	26	40	29	10½	35	36½	33
advice	41½	35½	32	33½	35	31	38
avenue	40	43½	29	45½	15	23½	38
ax (axe)	29	31	39	21½	48½	47½	38
beach	30½	33	39	33½	27	40½	38
bean	30½	46	39	10½	35	23½	38
accordingly	43	38½	39	33½	43	36½	43
acquire	48½	43½	32	33½	43	36½	43
ahead	36½	49½	39	45½	43	18	43
ample	45	29	48½	33½	35	40½	43
anyone	33½	49½	39	45½	43	23½	43
abate	48½	35½	48½	45½	48½	47½	48
abolish	48½	41½	46	33½	35	47½	48
abound	48½	35½	50	45½	43	47½	48
accommodate	45	48	43½	33½	43	40½	48
acorn	38½	46	46	45½	48½	47½	48

The practical principle of obtaining the best result per hour of time spent in the counts was considered throughout, and explains the use of counts of vocabularies in foreign-language books, and the use of some of the concordances. It is not assumed, for example, that many elementary-school

graduates will read Wordsworth, or Milton, or Pope, or Cowper to any considerable extent. The concordance counts, however, are made rapidly and with a moderate weight probably benefit the total result more than would the same amount of time spent on textual counts.

To turn to some illustrative cases, there are given in Table IV the sources of the credits for five words earning 100, 90, 80, 70, 60, and 50, 40, 30, 20, and 10 respectively. The words of credit above 100 would be at or near the top by any sort of count whatever.

In the next table (Table V) are given the rank order for these fifty words by the total sum of credits, and by children's reading alone, elementary text-books alone, standard literature alone, and so on.

It should be noted that I have included neither the counts from Hosic's 57 commonest selections in school readers nor *Sleepy Hollow* under Children's Literature here, but under English Classics.

Tables IV and V show that there is a correspondence between the sums of credits for the same word from different sorts of material, but that it is far from perfect. The coefficients of correlation (by the Spearman foot-rule) of Total with Children's Literature, English Classics, etc., are, in order, about .9, 8, .9, .5,.8, and .7.[11] The high values of the first three (.9, .8, .9) are in part due to the fact that they predominate in determining the Total. Common Facts and Trades has the lowest correlation partly because it has the least share in determining the total.

After this effect of the composition of the Total is allowed for, it still remains true that the vocabularies of Common Facts and Correspondence seem notably specialized in the table. That this is true in general seems certain to one who has made the counts. Indeed the Common Facts and Trades counts were selected to supply an obvious lack.

The greatest displacement for the Children's Literature is of *equal,*18½ ranks too low (41½ – 23), for the Bible and Classics, it is *angel,* 27 ranks too high; for the Vocabularies and Text-Books it is *class,* 19½ ranks too high; for Common Facts and Trades it is *sea,* 42½ ranks too low; for Newspapers it is *crowd,* 19½ too high; and for Correspondence it is *ah,* 28½ too low. These displacements are all such as might be expected even in an infinitely extensive count. To obtain a measure of general importance for the elementary school p3upil and graduate, we have to assign weights. If we weight Common Facts more heavily, *angels* will go down and *bricks* will go up!

The sums of credits at the basis of our list down to 10, the end of the fifth thousand, are determined to over three fourths of their amount by the counts from children's reading, classics, text-books and vocabularies. The result of this weighting corresponds probably rather closely with importance as measured by the prevailing ideals of what an elementary school pupil and graduate *should* read. These ideals are, however, themselves perhaps somewhat inappreciative of science, technology, busi-ness, and politics in comparison with literature and morals. They also perhaps undervalue the present and future in comparison with the past. There is further an increasing trend toward considering what the pupils *will* read as well as what they

---

[11] The absolute magnitudes of these coefficients lack their ordinary meanings, since the words are taken at intervals of 10 in respect to the credit sum. The relative magnitudes are all that I mean to measure by them.

should read. The present list may therefore be criticized as too weak in credit to children's literature and newspapers, and much too weak in credit to words relating to common life and trades.

It will, however, serve reasonably well until more extensive and specialized counts are made.

The following would be specially worth making:

A count of 250,000 to 1,000,000 words taken at random from a standard encyclopedia.

A count of 250,000 to 1,000,000 words taken at random from newspapers and weekly magazines.

A count of 250,000 to 500,000 words taken at random from boys' magazines.

A count of 250,000 to 500,000 words taken at random from girls' -magazines.

A count of 250,000 to 500,000 words taken at random from reading of a definitely utilitarian nature, such as printed directions for the use of machines, tools, medicines, etc.; laws and ordinances; proclamations; hand books; railroad guides; civil service examinations for policemen, firemen and the like.

USES OF THE LIST IN THE TEACHER'S WORD BOOK

THE TEACHING OF WORDS

Conscientious teachers now spend much time and thought in deciding what pedagogical treatment to use in the case of words which offer difficulty to pupils. In the third readers which they use they find, according to Miller, over nine thousand different words.[12] Some of these probably should not be taught at all in that grade; some should be explained at the time to serve the purpose of the story or poem, but then left to their fate; some should be thoroughly taught and reviewed. The *Teacher's Word Book* helps the teacher to decide quickly which treatment is appropriate.

The same service is performed, of course, in each of the school grades. Consider, for example, these words taken from Shelley's "Ode to a Skylark," a poem often found in school readers for grade seven or eight.

blithe	harmonious	rainbow	unbidden
bright'ning	hidden	rapture	unbodied
chaunt	hymeneal	shrill	unpremeditated
chorus	hymns	soar	vaunt
divine	ignorance	spirit	wert
dost	madness	sprite	wherein
forth	matched	strains	wrought
fountain	melody	sunken	
gladness	panted	thine	
heeded	profuse	triumphal	

If the reader will decide for himself in which thousand each of these belongs, keep account of the time spent to reach a decision, and then compare his ratings with those on page 357 [the next page] derived from the list, he will have a sample of the gain in time and correctness of judgment due to using the list.

Even expert teachers have very inadequate and inaccurate notions of the relative frequency and importance of words. For example, thirteen expert

---

[12] Many of these, however, are derived forms.

teachers were asked to rank certain words as 10, if in the first thousand for importance; 9, if in the second thousand; 8, if in the third thousand; and so on, using 0 for words below the tenth thousand. They differed enormously from one another in their ratings for the same words; and any one of them gave widely different ratings to words which are of closely equal importance by our count or by the average voice of the thirteen. For example, their ratings for ten words all having 10 as a sum of credits, and thus being all in or near the lower half of the fifth thousand, were as shown below:

RATINGS OF 10 WORDS BY 13 TEACHERS (I, II, III, ETC.)

(The + sign means "more" or "over.")

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	Range for the same word
abate	1	3	6	4	7	7	0	3	6	8	9	3	6	9000 or +
abolish	1	5	8	3	8	9	9	6	8	10	10	7	7	9000
abound	6	6	8	6	6	8	4	6	8	10	6	3	7	7000
accommodate	8	10	7	4	8	6	10	8	8	10	10	8	7	6000
acorn	1	1	7	10	7	7	9	7	9	5	10	8	8	9000
admission	7	6	9	4	9	8	10	8	8	10	8	10	10	6000
adversary	2	4	7	3	6	5	5	0	7	7	5	2	5	7000 or +
agency	4	6	5	8	8	9	9	7	8	10	10	6	6	6000
aisle	5	5	4	6	10	6	10	7	7	10	8	3	8	7000
allegiance	3	2	4	..	9	7	9	10	8	10	8	6	10	5000 or +
Range for the same teacher in 1000's	6	9	5	7	4	4	10 or +	7	3	3	5	8	5	

RATINGS OF THE WORDS FROM THE "ODE TO A SKYLARK" BY THE LIST, THE NUMBERS REFER TO SUCCESSIVE THOUSANDS. "NOT" MEANS THAT THE WORD IS NOT IN THE 10,000.

blithe	7	harmonious	7	rainbow	3	unbidden	not
bright'ning	not	hidden	4	rapture	4	unbodied	not
chaunt	10	hymeneal	not	shrill	4	unpremeditated	not
chorus	7	hymns	4	soar	5	vaunt	7
divine	2	ignorance	4	spirit	1	wert	4
dost	3	madness	4	sprite	6	wherein	4
forth	1	matched	2	strains	3	wrought	3
fountain	2	melody	4	sunken	10		
gladness	5	panted	3	thine	3		
heeded	3	profuse	7	triumphal	7		

In teaching arithmetic, history, geography, civics, or elementary science, there will be found in the book lessons many words which some of the pupils will not understand. Which are these and in which cases should the occasion be used to master a word for future use? Decision obviously depends in part upon how important the word is. For example, the first twenty-five pages of a standard geography for elementary schools contain these words:

accumulate	cable	consequently
alluvial	canal	continual
ant	capital	continue
area	capitol	definition
barrier	camel	delta
beaver	cascade	department
boulder	churn	deserted
bounce	clay	ditch
Buffalo	cliff	drain
burrow	climber	dweller

What the teacher should do with each of these words depends partly on its special importance for geography, and partly on its general importance then and later for pupils of the grade in question. It will be found, with respect to the latter, that some of these words rank as high as the first thousand, while some of them do not appear in our list at all, and probably would not appear even in a list fifty per cent larger.

The *Teacher's Word Book* does not, of course, rate correctly for any one community, words which are very important locally (as, for New York City, *subway, elevated, Brooklyn).* By its very existence, however, it directs attention to this issue, and stimulates the educational authorities to extend and amend it in respect to words of special local importance. In the case of spelling, the publication of general lists has been notably effective in producing the reaction of attention to special local lists; and we may expect the same effect from this reading- or meaning-list.

THE ESTABLISHMENT AND CLARIFICATION OF STANDARDS

The list makes it much easier than it has been in the past to put standards for word knowledge, by grades, by ages, or by mental ages, into clear, definite comprehensible form. For example, we may say that at a certain mental age or grade the minimum standard should be knowledge of the meanings of 95 per cent of the first 2500 worlds, 80 per cent of the next 1000, 60 per cent of the next 1500, and 20 per cent of the next 5000. If it seems desirable, we can specify still more narrowly, for example, 100 per cent of the first 1000, 95 per cent of the next 1000, 90 per cent of the next 1000, 80 per cent of the next 2000, with or without stipulation of a knowledge of the second 5000.

The actual learning of meanings is probably best accomplished by a large amount of relatively easy reading, plus a much smaller amount of harder reading with recourse to the dictionary, plus a still smaller amount of specific teaching of meanings as such. This actual learning of meanings also may be accomplished by means of varying and unsystematic stimulation of individual pupils. But the testing and keeping account of the knowledge gained does need to be in terms of specific word knowledge, and the list is a great aid in defining and testing such word knowledge.

THE EVALUATION OF TEXT-BOOKS

Within very recent years there has been quietly developing a demand for objective, scientific evaluation of text-books and other instruments of instruction. A first book in reading, for example, is being judged by a system

of credit points for type, spacing, number of words used, quality of the English, interest of the selection to little children, and the like. One element in such an evaluation of almost all text-books is the suitability of their vocabulary to the grade for which they are intended. This can be measured with absolute impartiality with the aid of the list. One has only to make a word count of a sufficient sampling of pages from the book in question, and look up the ratings of the words on the list. For example, when it is found that of two contemporaneous beginners' books in arithmetic to be read by pupils in the first half of grade 3, one has in the first fifty pages eleven words that are not in our 10,000 at all, and twenty-five more that are not in the first 5000, whereas the other has five and fifteen as the corresponding numbers, it is obvious where credit belongs for wisdom and care in the choice of words.

<center>SELECTION AND GRADATION IN READERS</center>

The importance of the words in readers and other reading material and the gradation of this material should, of course, be one element in its evaluation. Without a word list such as this, however, the judgments have necessarily been subjective and rather vague. With this list, they can be absolutely impartial and precise to any desired degree. The results of the studies which I hope this list will stimulate, are likely to be far-reaching in their exposure of imperfect selection and gradation of material in even our best instruments of instruction. Until such a list was available, indeed, the labor of inspecting material for details of vocabulary was too great perhaps to be expected from authors. As an illustration, I have taken at random the last pages of Book II, the first pages of Book III, and the last pages of Book III, counting approximately 4500 words in the case of each, from one of the best of present series of readers.

<center>TABLE VI</center>

<center>WORDS IN THE FIRST 10,000 BECAUSE OF THEIR PRESENCE IN
FIRST, SECOND AND THIRD READERS[13]</center>

bagpipe	Blitzen	bushtail	cluck
banter	bluebell	buzzard	cob
begone	Bobby		codfish
belfry	bonny		comical
bin	bo-peep		coon
blink	bowlful	cackle	coverlet
blur	Boxer	calyx	Cox
bonfire	boxwood	cipher	cozily
bun	brawn	crier	cozy
baa	brawny	cruise	crape
badger	Bremen	caller	creak
ball-room	brig	caramel	cress
bantam	brimful	catnip	crock
barnyard	Brom	can	crocus
bask	broth	changeful	cub
beanstock	Brownie	chanticleer	cupola
beck	browse	charley	curd

---

[13] Some of these are not in the 10,000 as printed since there were over 1700 words of credit 3, which carried the list to about 10,800. So the less important of the "3" words were left out of the final list.

<center>23</center>

bedside	bruce	chee	curt
beefsteak	brunette	cheerless	czar
belated	Bruin	chick	czardom
bespoke	bumble	Chinaman	
betook	bumble-bee	chirk	
Betsy	bummel	Cinderella	
birdie	Bunny	clapper	
blare	burrow	clement	

The vocabulary of the beginning of Book III is actually wider than that of the end of the book: 842 words to 736! From the end of Book II to the beginning of Book III there is a jump of over 20 per cent, from 681 to 842.

Attention to the importance and difficulty of words in selection and gradation is only one of many possible desiderata to be considered in a series of readers. It may well be sacrificed from time to time for the sake of literary excellence, or interest, or informational value, or other worthy qualities. But it should not be sacrificed to no purpose; and it has just claims for much consideration not only in grades 1 to 3, but to the very end of the elementary course.

The list, when used in connection with a word count of any instrument of instruction, will probably often lead to constructive recommendations of some importance. Two such may be noted here. The first is that primers and first readers should try to secure interest and adaptation to childish ideas and activities, without recourse to rare and even fictitious words. The second is that they should try to provide for phonic experience and practice without recourse to such rare or fictitious words. There are words in primers and first readers which do not even rank in our 10,000, or would not rank there except for the credits they get by virtue of their use in primers and first readers. Children are taught to read words in the first year of school which they may actually not see again for years.

TABLE VII

WORDS IN THE FIRST, SECOND AND THIRD READERS WHICH ARE
NOT IN THE 10,000

backwoods	bluejay	calmy	cleak
benighted	bluebird	Campbell	climax
baff	blumb	cantaloupe	climber
bagful	Blynkin	Canute	clop
baggy	bobtail	carman	clump
bab	boo	Caroline	cockleshell
bannock	Borneo	Carrara	columbine
barbecue	Bose	Caspar	commoner
barrelful	bossy	Charlotte	Conrad
basketful	Bowden	chatterer	Cora
beanshooter	bricklayer	cheep	corncob
bearskin	Brinker	chipmunk	Corrola
beehive	brooklet	chirrup	counterpane
Bert	broomstick	chore	craggy
Biddy	burdock	chubby	creepy
biff		chug	crispy
bight		chump	croon
birdling		churnful	crotch
blithesome		citron	crumple

A list of words which would not be in the 10,000 except for their presence in the first, second or third readers is instructive from many points of view. This list for words beginning in *b* and *c* is given on page 23 [Table VI]. Above [Table VII] is a list of words which are in the first, second, and third readers, but are not even in the 10,000.

MATERIAL FOR PHONIC DRILLS

It is interesting to note those words which are suitable to develop phonic insights and habits and are among the thousand most important words according to our count. I have, therefore, taken about seventy of the most useful phonograms, and entered after each phonogram the words from the first thousand of our list which present it clearly. Some of these words are not equal in interest to the words now used in beginning reading for the purpose, but on the whole, they will form a very serviceable basis for phonic drills; and every one of them is well worth learning for its own sake.

ace—face place race space
ack—back black
ade—made shade trade
ail—sail
ake—cake lake shake take
all—all ball call fall hall small tall wall
ame—came game name same
an—an can cannot man manner plan ran than
and—and band command demand hand land sand stand
at—at fat hat matter sat that
ate—gate late state
ay—away bay day gray lay may pay play say stay today way
bl—black bless blind blood blow blue
br—branch brave bread break breakfast bridge bright bring
    broad broken brook brother brought brown
ch—chair chance change charge chief child children choose church
cl—class clean clear clock close cloth clothing clothes
cr—cried cross crowd crown cry
dr—draw dream dress drink drive drop dry
eam—dream
eep—deep keep sheep sleep
eet—meet street sweet
ell—bell fell fellow tell well
en—men pen ten then when
ent—cent center different enter entire sent went
est—best nest rest yesterday
et—get let letter met set settle yet
fl—floor flow flower fly
fr—free French fresh friend from front fruit
gl—glad glass
gr—Grace grain grant grass gray great green grew ground grow
ice—nice price
ick—prick quick sick stick thick
ide—beside decide divide guide hide ride side wide
ight—bright delight fight light might night right sight
ill—bill fill ill kill mill still till will
in—begin in inch increase indeed Indian instead interest into
    skin thin win
ine—line mine nine shine
ing—being bring coming during evening going king morning
    ring sing and many others
ip—lip ship trip
it—fit it its sit
ite—quite white write
oat—boat coat
ock—clock lock rock stock
old—cold gold hold old sold told
ong—along belong long song strong wrong
ook—book brook cook look took
oon—noon soon
op—shop stop top
ot—hot not
ound—around found ground pound round sound
out—about out outside shout
own—brown crown down town
ox—box
pi—place plain plan plant play pleasant pleasure
pr—practice press pretty price prince promise proper prove proud
qu—quarter queen question quick quiet quite require
sh— shade shake shall shape she sheep shine ship shoe
shop shore short should shoulder shout show shut
sk—skin sky
sl—sleep slow
sm—small smoker
sp—space speak spend spirit spoke spot spread spring

st—stand star start state station stay step stick still stock
    stone stood stop store storm story study
str—straight strange stream street strength strike strong
tr—trade train travel tree trip
wh—what wheat wheel when where whether which while
    white who whole whom whose why

A STANDARD VOCABULARY FOR THE TEACHING OF FOREIGNERS

By the elimination of certain specially childish or "literary" words from the first 500 of our list and the addition of certain words of special importance to the newcomer to America, such as *danger, poison, cent, dollar, entrance, exit,* we shall have a basic list of great value in teaching foreign adults to read English. A tentative first 500 for foreigners may be formed simply by omitting *ball, being, pretty,* and *soldier* from the list on page 365, and adding *danger, poison, cent* and *dollar.* A second 500 can be formed in much the same way. The use of two such lists, with any additions necessary to arouse interest and meet local needs, may be expected to improve and facilitate the teaching of English to foreigners above the age of twelve.

THE VOCABULARIES IN ELEMENTARY BOOKS FOR TEACHING A FOREIGN LANGUAGE

Investigations by Bagster-Collins have shown that first-year books in French, German, and Spanish differ enormously in the words used, and that many of the words in any one of them are such as should not be learned by the beginner. It is also probably the case that in any one of them some very important words will be given little or even no attention. Until counts for the foreign language itself, comparable to our count for English, are available, it will be worth while to check the vocabulary of a foreign language text-book against our list. Words of little importance according to our list should receive little emphasis in teaching, unless they are of clear service to the student. Words of much importance according to our list which are given little or no practice by the text-book should be provided for by additional exercises unless they are words like *acre, baseball,* or *inch* where use is restricted chiefly to English speakers and writers.

THE FIRST 5OO WORDS

A   about   above   across   add   after   again   against   air   all   almost
alone   along   also   always   am   among   an   and   another   answer   any
apple   are   arm   around   as   ask   at   away   back   bad   ball   bank   be
bear   beautiful   because   become   bed

been   before   begin   behind   being   believe   best   better   between   big   bird
black   blow   blue   body   book   both   box   boy   bread   bring   brother
brought   build   burn   but   buy   by   call   came   can   care   carry   case   cause
certain   change   child   children   church

city   clear   close   cold   color   come   company   corn   could   country   course
cover   cross   cut   dark   day   dead   dear   death   deep   did   die   do   does
done   door   down   draw   dress   drink   drive   drop   during   each   ear   early
earth   east   eat   egg

end   enough   even   ever   every   eye   face   fair   fall   family   far   fast
father   fear   feel   feet   few   field   fill   find   fine   fire   first   five   floor
flower   fly   follow   food   foot   for   form   found   four   free   fresh
friend   from   front   full

garden   gave   general   get   girl   give   given   glad   go   God   gold   good   got   great   green   ground   grow   had   hair   half   hand   happy   hard   has   have   he   head   hear   heart   heavy   help   her   here   high   hill   him   himself   his   hold   home

hope   horse   hot   hour   house   how   hundred   I   if   in   into   is   it   its   just   keep   kill   kind   king   know   known   land   large   last   late   laugh   law   lay   lead   learn   leave   left   length   less   let   letter   lie   life   light   like

line   little   live   long   look   lost   love   low   made   make   man   many   mark   matter   may   me   mean   measure   meet   men   might   mile   milk   mind   mine   miss   money   month   more   morning   most   mother   mountain   move   much   must   my   name   near   need

never   new   next   night   no   north   not   nothing   now   number   of   off   often   old   on   once   one   only   open   or   order   other   our   out   over   own   paper   part   pass   pay   people   person   picture   piece   place   plain   plant   play   please   point

poor   power   present   pretty   put   quick   rain   raise   reach   read   ready   reason   receive   red   remain   remember   rest   rich   ride   right   river   road   rock   roll   room   round   run   said   sail   same   save   saw   say   school   sea   second   see   seem   seen   send

sent   serve   set   several   shall   she   ship   short   should   show   side   sight   silver   since   sing   sister   sit   six   sleep   small   so   soft   soldier   some   something   sometime   son   soon   sound   south   speak   spring   stand   start   state   stay   step   still   stone   stop

story   street   strong   such   summer   sun   sure   sweet   table   take   talk   tell   ten   than   thank   that   the   their   them   then   there   these   they   thing   think   third   this   those   though   thought   thousand   three   through   till   time   to   today   together   too   took   top

town   train   tree   true   try   turn   two   under   until   up   upon   us   use   very   visit   voice   wait   walk   wall   want   war   warm   was   watch   water   way   we   week   well   went   were   what   when   where   which   while   white   who   whole   why

wide   will   wind   window   winter   wish   with   without   woman   wood   word   work   world   would   write   year   yet   you   young   your

TEXT-BOOKS FOR HIGH SCHOOL PUPILS

The instruments of instruction in mathematics, science, history, civics, and even literature, used for pupils in high schools, especially in Grade 9, may well be scrutinized from the point of view of the *Teacher's Word Book*. Other things being equal, it is better not to burden a subject like algebra or chemistry with unnecessary linguistic difficulties. The slight gain from a widened vocabulary is more than balanced by the loss in ease of comprehension of the principle to be taught.

As an illustration, note some of the words used in the first fifty pages of two well-known text-books in algebra which are not in the list at all. Some of these should probably not have been used. In the case of others the preparation for the lessons in question should include special attention to the meanings of the words.

WORDS IN THE FIRST FIFTY PAGES OF A STANDARD ALGEBRA
WHICH ARE NOT IN THE 10,000

abbreviate	Diophantus	literal	prefix
Ahmes	dissimilar	mathematical	Pythagoras
algebraic	distinctive	mathematician	quadrilateral
algebraist	elementary	minuend	rearrange
arithmetical	et	monomial	redwood
assets	exponent	multiplicand	reintroduce
binomial	facilitate	multiplier	rewrite
bricklayer	formula	ninth	scalepan
casting	formulae	nitrogen	simplify
coefficient	Harriot	numerical	Stifel
complementary	haw	Oughtred	subtrahend
computation	Herigone	parallelogram	supplement
debit	Hindu	parenthesis	supplementary
Demosthenes	ingenuity	Pell	trinomial
denominate	integer	perimeter	Vieta
Descartes	Leibnitz	polynomial	Widmann
digit	likelihood	potentia	workmanlike

THE AMENDMENT AND EXTENSION OF SPELLING LISTS

Ours is not a spelling list, and the order of importance of words for spelling will often diverge widely from that of this list. Words like *ache, cough, hoarse, doctor, medicine, coal, shoes, waist, dear madam, yours truly, Mass., Conn., Ill., Neb.,* will figure relatively much more frequently in writing than in reading. In a spelling list also the derivatives may best be counted separately.

This list will, however, be useful to correct certain notable omissions from the spelling lists, such as parts of irregular verbs from the Cook-O'Shea list, or names of the days, months, states, large cities, and the like from the Ayres list. It will be still more useful to extend spelling lists beyond 2500 or 3000, as seems necessary with pupils in high schools and pupils who expect to become stenographers; and this may also be desirable for others. The work being done at the University of Iowa by Horn and Ashbaugh will soon, I trust, provide us with a list even more adequate for spelling than this list is for reading.

THE IMPROVEMENT OF READING SCALES AND TESTS

This list makes it possible to devise tests and scales for word knowledge which will be very much superior to any that we now have. Existing tests and scales are made better at once because we can give to each element of the test a rough measure of importance.

It also becomes much easier than it would otherwise be to extend tests of word knowledge by alternative forms; and to assign a provisional gradation for difficulty within each form. Finally, certain very promising new methods of testing word knowledge become practicable, when we have ten thousand words graded fairly well as to importance.[14]

---

[14] It may be noted here that the List will be of value in the arrangement of psychological tests in general, by enabling their authors to be sure that the words in the instructions for a test are

OTHER USES

It is hoped that the *Teacher's Word Book* will be of service to students of education in many other ways. It seems, for example, to offer an excellent chance to measure the relative importance of words of Latin derivation, and of the extent to which a given knowledge of Latin may be expected to help a pupil of a given degree of ability to understand the present meanings of these words.

The frequency of the use of words in the reading matter of any given time for any given group is of some value as an index of the knowledge and interests of that group at that time. The words which are not in concordances of Shakspeare and Milton but are very common to-day would, I think, make an instructive list. The words which are not in Shakspeare, Milton, Pope, Cowper, Wordsworth, or Tennyson, but are common today, are perhaps more instructive. The words which figure largely in the two newspaper counts but only slightly elsewhere in our material form an interesting group. Among them are, for example, *hosiery, millionaire,* and *nomination!*

There is, of course, a notable lag of a list like this behind the actual reading material of 1920. It may well be that *automobile, auto,* and *Ford,* are now read oftener than *horse,* though the credit sum on our list is 108 for horse to 33, 13, and less than 10^{15} for the other three, respectively. The lag in school readers is perhaps even greater than that in our list.

Children read about giants and fairies, knights and castles, kings and queens, forms of work and fighting, and ways of thought and belief, which are really a sort of paleontology of civilization. This does not, of course, do much harm, since the misleadings about the facts of human nature and the world are probably easily curable, and the literature of folk-lore, feudalism, and militaristic societies is reputed to have great merits of simplicity, interest, and literary quality. In general, material finds entrance to our readers a generation or more after it is written. Possibly this is wise. Literary critics as a rule agree that you must wait for the long test of time to decide what is really great and fine in literature. They would presumably not make such a confession of their own lack of acuity if it were not true. The contrast with science, where the elementary student may learn about, say, the electrons, a year or two after their discovery, should, however, receive our attention. Since school histories also usually spend 95 per cent of their energies on the world minus the decades since the pupil and his older brothers and sisters were born, the pupil's academic reading acquaintance with human affairs is almost entirely out of date. It may be well for it to be so. It surely is well that teachers should understand that it is so. The *Teacher's Word Book* may also help us to understand why so many children and so many adults read what the cultivated man condemns as trash. Dr. Jordan[16] has shown that boys will wait for an hour at public libraries to get books by certain present-day writers whose names the cultivated man has never even heard. It is commonly assumed that children and adults prefer trashy stories in large measure because they are more exciting and more stimulating in respect to sex. There is, however, reason to believe that greater ease of reading in respect to vocabulary, construction, and facts, is a very important cause of preference. A count of the vocabulary of "best sellers" and a summary of it in terms of our

sufficiently easy to understand.

[15] 18 for ford and Ford together. At least 8 are for ford.

[16] Jordan, A. W.. Children's Interests in Reading. Teachers College, Contributions to Education, No. 107.

list would thus be very instructive.

PREFIXES AND SUFFIXES

It was formerly customary to teach the meanings of prefixes and suffixes. Such teaching became discredited, partly because it extended to rarities like *agogue* and *ambulist,* partly because the prefixes and suffixes were divorced from their connections in real words, and partly because of the ambiguities of many of them. The word counts reveal the very great importance of certain prefixes and suffixes, and suggest that they are worth teaching, in proper connections, even though the pupil will be occasionally misled. For example, there are in our list 170 words beginning with *un.* Of these, 15 begin with *under,* leaving 155. The remainder of the word is, in 140, or 90 per cent, of these 155 cases in the list as a word by itself. In only about four per cent will the simple negative or oppositional meaning of *un* mislead pupils (save perhaps a very few very dull pupils). If pupils were taught[17] *un—*after learning, say *unhappy, unkind,* and *unwilling—* it would probably represent a considerable economy over leaving them to discover its meaning haphazard in the course of further reading. The advisability of teaching any prefix or suffix should be considered in the light of similar data about it now readily available in the list.

This report has already overrun its allotment of space, but I must at least mention the fact that one chief service of the *Teacher's Word Book* will be to aid in the production of some much better list, from wider counts, to replace it. The entire procedure in counting, entering and crediting words is made very much easier, once we have an approximately correct list for use in recording the counts economically.

---

[17] The cases where such misleading might be considerable are: unanimous, uniform, unit, unless, unto, and untoward.