Introduction to Edgar Dale and Joseph O'Rourke's *The Living Word Vocabulary*

This 25-year study of vocabulary provides objective, familiarity scores on 44,000 word meanings. *The Living Word Vocabulary* may be used in many ways. With this list writers now can find the familiarity scores of the words they are using. If the score indicates that a word is unfamiliar, the writer can explain it either directly or indirectly, or remove it. In the past we have had extensive word lists by E. L. Thorndike, *The American Heritage Word Frequency Book* (Houghton Mifflin Company, Boston, MA 02107), and others solely on the basis of frequency in literature, textbooks and other writing. *This is the first time we have had an extensive list of word meanings and their familiarity scores.*

Persons preparing directions for filling out official applications and blanks of various kinds can now check the familiarity of words for intended readers. Public documents prepared for universal consumption can be written at the sixth to eighth grade level instead of at the tenth to twelfth, as is now common. Writers of textbooks preparing material for stated grade levels can turn to *The Living Word Vocabulary* for guidance. For example, writers of articles for *The World Book Encyclopedia* use this list to guide them in producing readable copy at varied age and grade levels. Test makers can be sure that the words in their tests are not too hard (or too easy) for prospective users. Vocabulary tests can be made to fit the persons being tested. In short, materials of instruction and information can be tailor-made.

In presenting this list of 44,000 words and their familiarity scores we can understand what Samuel Johnson meant in his “Preface to the Dictionary.” He notes the possible presence “of a few wild blunders, and risible absurdities, from which no work of such multiplicity was ever free, [and] may for a time furnish folly with laughter.” He points out that “no dictionary of a living tongue ever can be perfect, since while it is hastening to publication, some words are budding, and some falling away; that a whole life cannot be spent upon syntax and etymology, and that even a whole life would not be sufficient.” (*Johnson's Dictionary—A Modern Selection.* Pantheon Books, 1963, p. 28.)

This list provides an inventory of the written words known by children and young people in grades, 4, 6, 8, 10, 12, 13, and 16. It discloses what they know and don't know. It meets, in part, the definition of language given by Jerome S. Bruner, Jacqueline J. Goodnow and George A. Austin in *A Study of Thinking* (New York: John Wiley & Sons, 1956, p. 311): “For language is nothing less than inventory of all the ideas, interests, and occupations that take up the attention of the community. The study of semantics, in this extended sense, cannot be distinguished from the general study of culture.”

There were three stages in the development of *The Living Word Vocabulary*. Our first concern was to develop a list of common terms used widely in all writing: “A Comparison of Two Word Lists,” (*Educational Research Bulletin*, December 9, 1931, Vol. X, No. 18, pp. 484-87). These 769 words appeared:

1. in the international Kindergarten Word List of 2596 words (International Kindergarten Union, Child Study Committee, *A Study of the Vocabulary of Children before Entering the First Grade.* Washington, D.C.: Association for Childhood Education, 1928), and

Clarence R. Stone revised the Dale 769-word list by removing 173 words and adding 173 words. Here are some of the words he removed: bank, blind, blood, body, bone, born, burn, coal, cool, dead, doctor, finger, rose, salt, ship, silver, size. He also added such children’s words as bow-wow, chick, cock-a-doodle-doo, pet, moo, puppy, sled, doll, rooster, hen, rabbit. The 769-word list was used by Irving Lorge in “Predicting Readability,” (Teachers College Record, Vol. 45, No. 6, pp. 404-419, March 1944).

E. W. Dolch also revised the 769-word list, extending it to 1,000 words. (Problems in Reading, Garrard Press, 1948, pp. 108-129) This list was classified into categories dealing with The Child’s Person, Home Environment, School, Nature, War and Fighting, General Things, Qualities and Relationships, Miscellaneous Parts of Speech, Verbs. This classification makes the meanings much clearer and does not depend primarily upon a frequency score.

Twenty years later George Spache added 361 words to the revised Stone word list and removed 87 words, making a total of 1,041 words. (Good Reading for Poor Readers, revised 1974. Garrard Publishing Co., pp. 195-207.) Some of the words added by Spache were from an unpublished list developed by Edgar Dale and Emily Schuh titled, “A List of 1400 Words Known by 75% or More of First Grade Children in the Enrichment Program of the Columbus (Ohio) Public Schools,” March 1970.

Next we developed a list of 8,000 common words and used student judgment tests to check their familiarity. There were data to indicate a high correlation between student judgment and the familiarity of the word. We asked students in grades 4, 6, and 8 to check whether they knew the words which we presented to them in list form. The procedure worked well with easy words and with hard words, but not as well with the words in between. Further, it does not adequately take into account the large number of simple words which have multiple meanings, although we did present some of the words in short phrases. Students were reacting to words whose spelling sometimes confused them. Thus belie was stated as known because some thought it meant stomach. Eunuch was confused with unique.

A second approach to the use of pupil judgment involved having children in the fourth grade read Black Beauty and underline the unknown words. We got some valuable data to use in our 3,000-word list,* but the findings lacked needed accuracy. It did give a useful indication of familiarity.

A third approach involved placing The Bantam Concise Dictionary in the hands of college freshmen at The Ohio State University and Stephens College and asking them to read stated portions of the dictionary and check unknown words. This list proved valuable in discovering “hard” words but did not give us data on easy- and middle-level words. Here are samples of words checked as unknown by many of these college freshmen: acrimony, actinism, acumen, afflatus, agave, aggrandize, agnostic, albeit, aliquot, alluvial, alveolus, amaranth, ambergris, anachronism.

It became clear that we would have to turn to a more rigorous form of testing, namely a multiple-choice test. One of the writers had explored the nature and possibilities of such a test in his doctoral dissertation titled, “Factual Basis for Curriculum Revision in Arithmetic with Special Reference to Children’s Understanding of Business Terms,” (Edgar Dale, University of Chicago, 1929). First of all, it involved an analysis of the nature and frequency of business terms in investment literature written for the layman. We found 2,276 terms and tested the 200 most common ones using a four-choice multiple choice test.

What is a word and what determines its meaning?

In this study we have assumed that a word is a unit of meaning presented in a dictionary with its semantic variations. Some words in the dictionary are given only one clear-cut meaning, other words may have ten to twenty meanings. In The World Book Dictionary, for example, there are 23 meanings of good. The Shorter Oxford English Dictionary presents more than one hundred variant meanings of the word go.
The definitions used for the words presented here came chiefly from *The World Book Dictionary*. Many dictionaries, however, were consulted to find the number of different meanings of certain words and to note especially simple, clear-cut definitions. Some word meanings are our own, sometimes involving a shortened version of the longer dictionary entry.

**When is word known?**

Knowledge of a word can be placed on a continuum, starting with “I never saw or heard the word before.” For example, the reader of this introduction does not know the words *bittles, polentulate, or fundular* because they do not exist as meaningful words. Second, we may say, “I know there is such a word but I don’t know what it means.” Such words might be *hugger-mugger, adnoun, adit, P.E. ratio, serendipity*. A third stage is “the contextual placing of the word. You know you make the *welkin* ring and that *hustings* has to do with elections, but what do these words mean specifically? You *bask* in the sun. Can you *bask* in the shade? What are *kith and kin*? What is a *pied piper*? Was President Andrew Johnson *impeached*? (He was.) These words are in our “twilight zone.” Fourth, and finally, we reach the stage where we have pinned the word down. We know it. We would recognize it again if we saw it, and we are likely to remember it. This is the level at which we have tried to test our words.

**What about multiple meanings of words?**

When the list of entries was established, the next problem was the selection of important multiple meanings. The high frequency words have many meanings. The selections of the different meanings to be tested were usually made by the late Lytton Beeler, a student of words, who analyzed various dictionaries and selected the multiple meanings of key words. He spent much of his time—for ten years—working on this task. For example, the word *point* has more than 40 separate definitions in *The World Book Dictionary*, and with compound words and idioms occupies a full page. We tested the following meanings:

4-78% point................a compass direction  
4-82% point................a place  
6-81% point................important idea  
4-83% point................a mark or dot  
8-72% point................a certain stage  
4-71% point................to show with finger  
4-78% point.............to aim  
6-73% point...............a unit of scoring  
4-77% pointed...............sharp-ended  
12-47% needle-point........a kind of lace

Few major studies of the frequency of the occurrence of multiple meanings of a word have been made. Professors Lorge and Thorndike, with the aid of a staff of 270 persons, counted the frequency of the occurrence of different meanings of a sample of about four and a half million words. (Irving Lorge, “The English Semantic Count,” *Teachers College Record, 39*:65-77, October 1937. Irving Lorge and Edward L. Thorndike, *A Semantic Count of English Words*. New York: The Institute of Educational Research, Teachers College, Columbia University, 1938.)

**What about the inflections of words?**

The standard for testing inflections was this: If the inflection is likely to provide a variant score, we should test it. We usually did not test regularly inflected words—the addition of *s, ed, ing, ly*, etc. But we considered retesting if the visual pattern of the word was irregularly changed in the inflecting, for example, *substance* and *substantive*. Lengthening the word did tend to make it “harder” and in the lower grades inflection may make the word less familiar since the affixes might not be known.
Persons and Places?

When we discuss the vocabulary of an individual we are usually thinking about common nouns, not proper nouns. Yet there is much significant learning tied to the names of persons, trademarked products (cereals, automobiles), the states in the Union, the Presidents of the United States, the countries of the world, characters in literature. In the past 25 years we have been busy learning new words such as Zaire, Botswana, Namibia, Malagasy, Malawi—all in Africa and increasingly in the news.

Most scores of persons and places were low. We discovered that names of persons and places will only be known when they are specifically taught in school or have appeared repeatedly in news over time. We tested only a few trademarked items. A limited number of proper nouns appear in this list. However, we have additional test scores on about five thousand such words and may publish them later.

How were the words tested?

We gave up hope of developing an inexpensive, quick method for determining word knowledge and turned to a multiple-choice test. We chose a three-choice multiple-choice test for several reasons. Amos Tversky has pointed out in the *Journal of Mathematical Psychology* (1:386-91, University of Michigan, 1964) that: “Whenever the amount of time spent on the test is proportional to its total number of alternatives, the use of 3 alternatives at each choice point will maximize the amount of information obtained per time unit . . .” (p. 390). Clearly, in the writing and testing of some 44,000 different items this factor of saving time is critical. Tversky also says: “There exists some empirical evidence, based on the study of auto-instructional items, which indicates that 3-alternative test items are indeed optimal . . ." (p. 390). Taken together with Garner’s conclusions concerning human capacities to process multi-dimensional information, these results suggest that the use of three levels per dimension may be the most efficient way to code and process information (p. 391).

How did we know what words to test at what levels?

We had access to Gates’ *Spelling Difficulties in 3,876 Words*, which gave us a familiarity score on each of those words. (Bureau of Publications, Teachers College, Columbia University, New York: 1937.) We also had access to the unpublished study by Paul Diedrich and Osmond Palmer. They did multiple-choice testing of over 4,700 words in grades 11 and 13. These were chosen from the Thorndike word list. Out of these 4,700 words in the 11th grade list about 1/3 were known (67% or better). However, over 1/2 of the 13th grade words were known (67% or better). In addition, we made extensive use of the graded vocabulary list included by B. R. Buckingham and E. W. Dolch in *A Combined Word List* (Ginn and Co., 1936).

Our aim was to secure a score on each word of at least 67% and not more than 84%. If a word got a score of 66% in the 4th grade, we tested it again in the 6th grade. If a word in the 6th grade got a score of 85% or above, we retested it in the 4th grade. Thus most words in this list have scores of 67 % or above. However, we did not retest all words in grade 13 which got less than 67%—usually from lack of time.

We tested in grades 4, 6, 8, 10, etc. and did not test in the intervening grades—5, 7, 9, 11—because this would have almost doubled the cost of the project. However, the scores in the missing grade levels can be inferred from the fact that between the 2 tested grades there is usually an average difference of 20% in word scores. Hence scores between single grade levels are likely to differ by about 10%. Testing words in the 5th and 7th grades would have been useful because the child is then learning to read well and his vocabulary is growing rapidly. Many of the harder words at the 12th, 13th, and 16th grade level have scores below 67%. Some are being retested as a part of our continuing word testing program.
We offer a note of caution about words with scores below 50%. Mere chance would give a score of 33%, which might mean no knowledge of the word at all. Thus, a word with a score of 50% or less is generally a hard word.

Should we test the word in some kind of context?

Unfortunately, context can vary from no help to complete help in revealing the meaning of a word. So when someone says, “Test the word in context,” we must ask: “What context?”

Victor H. Kelley compared the scores received when he correlated the scores of students who used the word in a sentence and those from matching or multiple choice tests, and concluded: “The multiple choice and matching tests appear to be the best technique with no real difference existing between the two.” (Victor H. Kelley, “An Experimental Study of Certain Techniques for Testing Word Meanings,” Journal of Educational Research, Vol. 27, December 1933, pp. 227-282.)


Construction of Multiple-Choice Items

In constructing the three-choice tests, the critical problem lies in selecting the two incorrect choices, the distractors. We set up these rules for selecting them but art and skill are important in choosing these two distractors.

1. The order of the correct choice was based on a table of random numbers, hence is one entirely of chance.

2. The incorrect choices are assumed to be on the same level of familiarity as the word being tested. The correct definition must not be harder than the word being tested. For example, we would not test garbage by the noun refuse or uncertain by the word dubious. When we began testing we did not know for certain how hard the distracting choices were but as the testing continued we used our own data to determine the difficulty level of words used as distractors. Distractors should not include terms ordinarily not known beyond the grade at which the word is tested. That is, of course, a counsel of common sense, but persons who make tests are mature adults with superior vocabularies and they may use distractors which the child doesn’t know.

3. The test usually occupied only one line.

4. We aimed to make all choices about the same length.

5. In general, if the word being tested is a negative one, the choices should be negative. If affirmative, the choices should be affirmative.

6. There is a tendency among test writers to use distractors which require spelling discriminations. The student is then tested not on the meaning of the word but on his spelling ability. So we would not test dual (twofold) with the distractor “to fight with swords,” or test bier (coffin stand) as the distractor for “an alcoholic drink.” True, the ability to spell well may be required to tell the difference between the two words, but meaning, not spelling, was the aim of this testing. Inability to spell correctly may cause reading errors.
7. Care must be taken so that the word tested cannot be guessed by easy elimination. For example, if we were testing *countersign*, we would not use (a) answer to password (b) a small dog (c) a child, but rather: (a) answer to password (b) rear flag on a ship (c) inside advertisement. Below is a current list of actual test samples and scores ranging from the fourth grade to the college senior level.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Score</th>
<th>Word</th>
<th>Sample Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>58%</td>
<td>etymon</td>
<td>(a) study of insects (b) figure of speech (c) a root word</td>
</tr>
<tr>
<td>16</td>
<td>68%</td>
<td>evoke</td>
<td>(a) call upon for help (b) take back again (c) call forth</td>
</tr>
<tr>
<td>13</td>
<td>67%</td>
<td>magma</td>
<td>(a) moon rocks (b) molten earth (c) sea bed</td>
</tr>
<tr>
<td>13</td>
<td>55%</td>
<td>bicameral</td>
<td>(a) two-horned (b) with two lens openings (c) having two houses</td>
</tr>
<tr>
<td>12</td>
<td>79%</td>
<td>cardinal number</td>
<td>(a) 1, 10, 20 (b) 1st, 2nd, 3rd, etc. (c) 20th, 30th, 40th</td>
</tr>
<tr>
<td>12</td>
<td>72%</td>
<td>macho</td>
<td>(a) cruel (b) speed (c) marly</td>
</tr>
<tr>
<td>12</td>
<td>78%</td>
<td>adrenalin</td>
<td>(a) gland secretion (b) harmful disease (c) stiff fabric</td>
</tr>
<tr>
<td>10</td>
<td>79%</td>
<td>gala</td>
<td>(a) bitter-tasting liquid (b) a celebration (c) a star cluster</td>
</tr>
<tr>
<td>10</td>
<td>67%</td>
<td>medley</td>
<td>(a) interference (b) illness (c) a mixture</td>
</tr>
<tr>
<td>8</td>
<td>79%</td>
<td>nitrogen</td>
<td>(a) 4/5 of the air (b) medicine (c) explosive</td>
</tr>
<tr>
<td>8</td>
<td>79%</td>
<td>architectural</td>
<td>(a) bowed (b) dartlike (c) about buildings</td>
</tr>
<tr>
<td>6</td>
<td>68%</td>
<td>gas guzzler</td>
<td>(a) requires much fuel (b) a stomach problem (c) drinks too much</td>
</tr>
<tr>
<td>6</td>
<td>85%</td>
<td>method</td>
<td>(a) good excuse (b) loud noise (c) plan for doing</td>
</tr>
<tr>
<td>6</td>
<td>74%</td>
<td>meteor</td>
<td>(a) rocket ship (b) falling star (c) crater on moon</td>
</tr>
<tr>
<td>4</td>
<td>71%</td>
<td>actual</td>
<td>(a) a kind of act (b) busy (c) real</td>
</tr>
<tr>
<td>4</td>
<td>74%</td>
<td>capable</td>
<td>(a) not needed (b) having great value (c) able to do</td>
</tr>
</tbody>
</table>

Making the Tests

When we tested one form of a word it was helpful to have the earlier test scores on other forms, e.g., vary, variant, variable, variation, invariant, and the like. Hence the more we tested, the more data we gathered on the scores of related words.

We studied the scores of tested words to see whether they agreed with previous scores on inflected words or agreed with our personal knowledge of familiarity of words. The computer gave each word a score noting the differences between the score in the upper 2% of the group tested, and the lowest 25%. We could quickly spot and revise weak tests on which the less able students did as well or better than the able students.

Furthermore, the average correct score on our tests ran around 75%, hence we were working with words which are, on the whole, fairly well known. This was not true, however, at the 13th and 16th grade levels. We also had computer data on the % choosing each distractor which helped us determine their power or weakness as distractors. Words were retested when we concluded they tested too high or too low. But sometimes it was hard to discover why certain distractors overpulled or underpulled.

The Nature of the Sampling

In a study of the vocabulary knowledge of students in the United States the question of sampling immediately arises: What students and how many were sampled from the total population?

Here are some questions asked about the sample of persons:

1. Why was a minimum of 200 students used for testing?
2. How were the students chosen for the study? Was it a random sample? Are these students representative?
Most of our tests were randomly sampled in cities and public school districts as found in the Education Directory of the U.S. Office of Education, Public School Systems.

School Systems

We had a few refusals to cooperate, but no evidence that the schools selected the more able classes. The few schools which could not give our tests usually indicated that they already were engaged in extensive testing programs. To avoid sectional bias we sent 50 tests or less of a single form to a single school district. We tested in all sizes of school systems—both rural and urban.

From 1954 to 1960 we used a three-choice test to determine the familiarity of 17,360 words in grades, 4, 6, 8, 10, and 12. The sampling was nationwide and included 41 states and 283 schools.

The familiarity scores and the names of all cooperating schools were presented in Children's Knowledge of Words (Edgar Dale and Gerhard Eichholz, The Ohio State University, 1960, out of print). Excerpted from Encyclopedia Buying Guide. Reprinted with permission of the R.R. Bowker Company. Copyright © 1981 Xerox Corporation.

After 1960 we tested about 320,000 students throughout the United States, approximately 1,600 tests. We present here familiarity scores on more than 44,000 word meanings in grades 4, 6, 8, 10, 12, 13, and 16. Not less than 200 subjects were tested on each word. Keep in mind, however, that many words were tested more than 200 times. For example, the word adrenalin (gland secretion) was tested 200 times at the 8th (42%); 200 times at the 10th (65%), before it attained a passing score of 78% at the 12th grade level (the score in this book). The word cagey (crafty) was tested at the 6th grade (46%), the 8th grade (57%), and at the 10th grade (72%)—its final score to date.

Our computerized data furnished a Kuder-Richardson reliability coefficient for each test. These correlations were rarely below .90. As noted earlier, we found that a sample of 200 students (per word) would provide stable averages. Arthur I. Gates of Teachers College found in his testing vocabulary that: "In some cases, it was necessary to obtain two hundred or more records from each of the six grades; in others, fewer grades and fewer cases were needed to stabilize the placement for the several percentages." (Ibid., pp. 2-3.)

Validity

The ultimate test of the word list is its value in preparing reading materials. Does the use of the list enable writers to anticipate the difficulties students are likely to have in the reading material? Do the writers and editors use The Living Word Vocabulary when they have easy access to it? Here is one answer:

For many years the editorial staff of The World Book Encyclopedia has used the familiarity scores in their writing and editing. This is what James Shacter, Chief Copy Editor of The World Book Encyclopedia, says:

The value of your readability testing has certainly been proven again by the results on the articles recently returned to us. The correlation between the words cited as difficult by students and/or teachers, and the grade levels on the Dale word list, is indeed remarkable.

Has the use of this list by a major publisher resulted in clear-cut benefits? Here is the judgment of Kenneth F. Kister, Editor, Encyclopedia Buying Guide, A Consumer Guide to General Encyclopedias in Print,* who points out:
... For many years World Book has paid stricter attention to vocabulary control and precise levels of readability than any other general encyclopedia on the market . . . (p. 333).

... New and technical terms are italicized and defined in context. This approach to vocabulary is based on 44,000-word graded list developed by Dr. Edgar Dale, a leading authority on readability and special consultant to World Book. In point of fact, Dr. Dale and his staff review all new and significantly revised material in each edition of World Book prior to publication to make sure that the vocabulary is geared to the encyclopedia's intended readership. This procedure ensures a high degree of comprehension of material as well as assists in vocabulary development . . . (p. 333).

To elaborate briefly, World Book is unquestionably one of the best designed and meticulously edited American encyclopedias ever made. As the digest above indicates, the encyclopedia passes every major critical test with flying colors. It is constructed to serve a very broad readership, ranging from students in the upper elementary grades to adults seeking basic information or reliable overview material on both academic and practical subjects (p. 337). The set's authority is unimpeachable, its contents thoroughly accessible, and its up-to-dateness and breadth of annual revision impressive. The text is clear, direct, usually interesting, and, when appropriate, the articles are written in pyramid style (i.e., from simple to complex). In addition, the vocabulary is strictly controlled, and, again when appropriate, articles are written to grade level . . . (p. 337).

We shall not complete our testing with this list of more than 44,000 words. We shall continue to test new words and check the scores of previously-tested words. Like Samuel Johnson, we trust that there are “few wild blunders” and few “risible absurdities.”

May 1981

Edgar Dale, Ph.D.
Joseph O'Rourke, Ph.D.