

THE MEASUREMENT OF INDIVIDUAL DIFFERENCES IN GENERAL ENGLISH VOCABULARIES

ROBERT H. SEASHORE
Northwestern University

AND

LOIS D. ECKERSON
University of Southern California

I. BACKGROUND OF THE PROBLEM

In attempting to answer the apparently simple question, "How many words do you know," numerous investigations have discovered that individual vocabularies, from childhood to maturity, are much larger than we have previously thought. It is not at all uncommon for an educated adult to estimate his own vocabulary at a figure which is only one to ten per cent of the results obtained from accurate measurement. But in spite of the general agreement on the relatively large size of measured vocabularies, there is an almost equally wide variation in the absolute size of vocabularies studied, even for comparable groups. This suggests the need of a careful analysis of definitions and methods to determine the sources of such variations, and to set up a more reliable method for such measurements.

But beyond the interest in the mere size of vocabularies and their measurement there are many significant relations of vocabulary to problems of psychology, education, and literature. In a figurative way it might even be said that words, or the concepts for which they stand, are the coin of the realm of knowledge. These units of knowledge may be accurately defined, classified, and their numbers counted so that they give an inventory of the raw materials among our intellectual resources. When supplemented by measures of various skills, we may then estimate a person's effectiveness in dealing with various problems, both theoretical and practical.

Some of these relationships between vocabulary and other factors have already been measured, but the real significance of these and similar problems can best be indicated by a careful analysis of the background and methods of the experiments themselves.

A. Analysis of Major Variables

1. *Definition and Classification of Words as Units of Measurement.*
For the purposes of our study we may begin by defining the unit of

measurement, a word, as an item listed in a standard unabridged dictionary, Funk and Wagnalls' *New Standard Dictionary of the English Language*, two volume edition of 1937.¹⁰ The editors of this dictionary state that it includes about four hundred fifty thousand entries, but this includes about eighty-six thousand five hundred seventy alternative meanings which are simply numbered and listed under a single word. If we omit these alternative meanings, in keeping with the ordinary definition of a word, there are only three hundred seventy thousand two hundred sixty words in this dictionary. Alternative spellings (about twelve thousand six hundred seventy), usually bracketed together, are not counted as separate words by the dictionary or by us.

The Funk and Wagnalls dictionary was chosen for this study for two reasons: (1) It is one of the few unabridged volumes having all words in a single alphabetical order, which facilitates the construction of representative short sample lists for the purposes of testing; and (2) it is large enough to give full opportunity for nearly all educated adults to test all of the words which they were likely to know, but does not include the very rare words which would be almost unknown except to the few specialists in languages who might prefer the larger and more authoritative Oxford dictionary. In order to avoid disputes on definitions or usages, persons tested were given credit for knowledge of a word if they mentioned any meaning listed by either Webster's or Funk and Wagnalls'. This degree of accuracy was sufficient for our purposes.

As a rough distinction in the importance of words, the Funk and Wagnalls dictionary lists certain words, which we shall call "basic," in heavier type and next to the margin. "Derivative" terms, including compound terms, are usually listed in medium type and indented under the basic term. For example, the basic word "loyal," has listed under it four derivatives, as follows: Loyal Legion (a Civil War order); loyalism (n.); loyalize (v.); and loyally (adv.). Since some persons may wish to consider the derivatives as only minor variations of the basic words, we have studied the two groups separately so that they may be considered either separately or together. Philologists have shown that most of our modern language words are really derived from a relatively small number of root words, but the evolution has extended back over so long a time and the modifications are in many cases so great that the origin of most words is not recognized by persons who are not language specialists. If we are to judge the separateness

of words by the criterion of *how an individual originally learned them*, whether independently or simply as variants of some earlier known word, the dictionary distinction of listing "basic" and "derivative" words should give us a meaningful basis for our study, and further analyses may be easily extended in later studies. People who are skilled in recognizing derivations beyond these rather obvious variations have usually learned the process of derivation long after they acquired most of their vocabulary as separate words.

An alternative definition of "a word" for testing purposes is that employed by Thorndike³¹ and others in measuring use vocabularies. Thorndike's rules for classifying words as mere variations of a main word include all of the following changes:

- (a) Plurals, *s*, or changing *y* to *ies*.
- (b) Adverbs, *ly*.
- (c) Comparatives and superlatives, *er* and *est*.
- (d) Verb forms *s*, *d*, *ed*, *ing*.
- (e) Past participles adding *n*.
- (f) Adjectives, adding *n* to proper nouns. (Adjectives adding *al* or *ate* are not mentioned.)

These main words correspond roughly to our "basic" words, and by counting the number of words in the dictionary which would be combined with our basic list according to Thorndike's definition we may derive a constant for translating our type of unit into the Thorndike type of unit. Either method is meaningful, and since both have been employed in vocabulary studies, it is necessary to remember their relationship. By counting all of the words in the dictionary which would be grouped with our basic sample according to Thorndike's definition we find that, in proportion, our estimate of three hundred seventy thousand two hundred sixty as the total number of words in the dictionary would be equivalent to a much smaller number of words according to Thorndike's definition.* However, following commonest usage, we shall hereafter use our own definition of a word unless specifically stated otherwise.

A further analysis of the number and kinds of words found in the dictionary is given in Table I, which represents the totals from an

* Such a comparison will be reported in a later study. The problem is complicated by the fact that there are many other types of variations besides those counted by Thorndike's system.

TABLE I.—ANALYSIS OF PROPORTIONS OF BASIC AND DERIVATIVE WORDS FROM SAMPLE PAGES OF THE FUNK AND WAGNALLS, NEW STANDARD DICTIONARY

Vol. I, number of page	"Basic," number of marginal words in heavy type	"Derivatives"					
		Other parts of speech	Com- pound words	Multiple mean- ings	Variants in spelling	Total deriva- tives	Total number of words on a page
101	70	17	33	23	5	78	148
201	34	11	112	18	2	143	177
301	58	6	95	24	2	127	183
401	40	27	91	29	4	151	191
501	62	11	36	40	5	92	154
601	55	9	97	60	8	174	229
701	38	16	32	27	3	78	116
801	40	22	50	24	2	98	138
901	49	19	67	39	5	130	179
1001	89	45	19	28	8	100	189
1101	107	15	28	30	5	78	185
1201	53	39	27	35	7	108	161
1301	47	43	58	18	5	124	171
1401	67	6	16	45	3	70	137
Totals.....	809	286	751	440	64	1551	2358
Average Vol. I.	57.7	20.4	53.6	31.4	4.6	111	168
Average Vol. II.....	62.9						
Average, I + II.....	60.3						
X 2757.....	166,247	56,243	147,775	86,570	12,672*	466,760	
(number of pages in dictionary)		204,018		99,242		minus 12,672*	
		(total derivative words included in our definition of a word)		(not included in our definition of a word)		454,088	Total number items in dictionary.

* Alternative spellings.

analysis of the words on a sample of twenty-seven pages equally spaced throughout the dictionary.

Assuming that these pages are a representative sample of all the words in the dictionary, we find that the total number of words according to our definition is divided into approximately one hundred sixty-six thousand basic words and two hundred four thousand derivative words.

Table II gives further evidence that our sample is representative of the kinds and proportions of various parts of speech in the dictionary.

2. *Methods of Sampling.*—The basic theory underlying the measurement of vocabulary size is that of obtaining a representative sam-

pling of the total material. Obviously it is inexpedient to test a person on all of the words in the dictionary, but even if we had the necessary time and facilities, there would be little point in doing so. Scientists and technicians have verified many times the fact that measurements of minute samples of very extensive material will give results which are almost as accurate as for the total itself, providing that the samples can be known to be representative of the kinds and

TABLE II.—THE PERCENTAGE OF PARTS OF SPEECH OF BASIC WORDS IN SAMPLES FROM THE DICTIONARY

Part of speech	Twenty-seven pages at one hundred-page intervals		One word every other page (our large sample)	
	Per cent	Per cent	Per cent	Per cent
Nouns	70	..	68	
Proper nouns		23		21
Common nouns		39		40
Abstract nouns		8		7
Pronouns	0	..	0	
Verbs	11	..	10	
Particles	18.2	..	22	
Adjectives		16		19
Adverbs		2		3
Prepositions2		0
Conjunctions		0		0
Interjections8	..	0	
	100	..	100	

amounts of materials occurring in the total amount. Instances in which we base our decisions on minute samples could be cited from any field of science, industry, and every-day life. We judge from a sample what the rest is like, and if the sample is carefully taken we are usually accurate in our estimate of the whole.

The dictionary, which includes in a single alphabetical order all of the words to be tested, is very advantageously arranged for the construction of such test samples. We wished to obtain a sample of several hundred words which would have the same proportion of words beginning with each letter of the alphabet as the proportions for the

dictionary as a whole. We, therefore, took the third "basic" or marginal item in the first column of every left hand page in the dictionary, which gave us a total of one thousand three hundred seventy-eight words, of which fifty-eight were prefixes, suffixes, or abbreviations and, therefore, unusable, leaving only one thousand three hundred twenty useful words. Since we wished eventually to construct a number of alternative forms of the test, this list was then divided into four equivalent subgroups, as described later. Annen,¹ in a previous study with Seashore, had shown that the test results from one such sample are reasonably close to those from other samples. Our own basic list Number 1 has since been found to give adult individual test scores for the odd-numbered items which have an average variation of only four per cent from the scores on the even-numbered items, a remarkably small variation when we consider that the complete list included only $\frac{1}{505}$ of all such words in the dictionary.

Our method of sampling is technically known as the method of work limits, since it sets a standard amount of work to be done and measures the number of words correctly defined in unlimited time. In this case it acts as a "power" test since the individual's score is determined by the levels of his word knowledge, but if a relatively short time limit is employed, it emphasizes the speed factors, which are related to, but not the same as, level of ability.

In certain studies of children it is possible to record all of the words used in a variety of situations over a period of days until no new words appear except those learned during the period. This gives a measure of the number of words which the individual did use, but it does not take into account those which he could have used if necessary, nor those which he could only recognize. This method is rarely applied to adults, whose language is too varied in relation to different situations to permit an accurate estimate of the absolute total from any short sampling of time. It can, however, be applied to the measurement of written vocabularies actually used by authors where their entire list of publications can be studied, but even these omit many words which the author probably could have used or at least understood in situations other than those covered in his writings.

3. *Size of Vocabularies in Certain Typical Situations.*—At this point it may be well to indicate a few figures on vocabulary size in representative situations. Among the English writers Shakespeare, for instance, used about fifteen thousand different words (Thorndike

type of units) in his published writings, Milton about eleven thousand different words, while the French author Victor Hugo used about twenty thousand words. Older writers such as Shakespeare had a much smaller English language to draw from, and we know nothing about the additional words which they could have used if necessary, or at least understood. This distinction as well as the difference in units of measurement must be remembered in comparing these figures with those from tests on dictionary samples.

Thorndike³¹ has determined by actual count the twenty thousand words most frequently used in every-day writing and speaking. As would be expected, a relatively small list of words accounts for a large proportion of all the words we use.

For example, in Franklin D. Roosevelt's speech of acceptance of the nomination of the presidency of the United States in 1932, seventy per cent of the words employed are found in Thorndike's list of the five hundred words most commonly used in every-day writing and speaking. Seventy-five per cent of the words fell within the commonest one thousand, and so on, in decreasing increments for successively higher levels.

C. K. Ogden²⁶ has found it possible to select eight hundred fifty general English words (Thorndike units), which together with one hundred fifty additional words from any specialized field such as a business or science are sufficient to enable a foreigner to communicate with anyone who now speaks English. Ogden has simply eliminated most of the synonyms and other equivalent words in favor of this simpler list, and in the same way has eliminated many of the alternative grammatical forms. The combination of simplified vocabulary and grammar is called Basic English,²⁶ and has considerable promise as an international language, since it is merely a restricted sample of the language which is already the most widely used throughout the world.

For the present, however, and as long as we wish to talk to people and to read the works of authors who have not restricted themselves to such a basic list, it is necessary to know more than these commonest words. The recognition vocabulary of an educated adult must be much larger, and, as our measurements will show, is considerably larger than even the vocabulary used by an author such as Shakespeare. In addition there are the specialized vocabularies of each technical field, not fully covered by a general dictionary. Measure-

ment of these is very important as indicators of aptitude or accomplishment in these special fields.

4. *Criteria of Knowledge*.—Here we must distinguish between several common criteria, all of which are meaningful but have different significance.

(a) *Criteria of Recognition*.—Perhaps the commonest definition of the term "knowing a word" is to be able to recognize or identify its meaning. To recognize a word is to refer it to a larger group of words, a genus, and to distinguish its meaning from that of other related species of that genus. Thus we say that lead belongs to the group (genus) of metals and has a certain specific gravity and other physico-chemical characteristics which differentiate it from other species of metals.

Since many words have more than one meaning it is necessary to test vocabulary in several ways. The most important way appears to be a test for knowledge of the commonest meaning, which for our purposes was taken to be the dictionary meaning most frequently given by preliminary groups of university undergraduates. Further research is under way to measure individual differences in the richness of meaning, as shown by knowledge of alternative meanings, and diction, or choice of words which are nearly equivalent but have slightly different connotations. These two separate types of tests may then be compared with that for knowledge of commonest meanings to determine the interrelations of these abilities.

In constructing a test for the recognition of the commonest meaning of a word we have chosen the multiple-response method with four choices. The person being tested selects a synonym or equivalent phrase as the correct answer, distinguishing it from several alternative species of the same genus (the next larger classification of words as given in Roget's *Thesaurus* or a similar work in technical fields) and also from other words which might be confused with the word to be defined because of similarities in sound or spelling. Among professional linguists it might be worth while to insist upon very fine discriminations in alternative meanings, but for our purposes in a general test it is sufficient to discriminate the commonest meaning from those which are definitely incorrect but which might be sources of confusion to non-specialists in the field.

1'. *Effect of Context upon Recognition in Vocabulary Tests*.—In measuring recognition of a word we may choose to test the subject's

knowledge of separate words, or of words appearing in a context. The ability to recognize a word should theoretically be aided by having it appear in a context such as a sentence. The only quantitative evidence so far reported in this field is that of Annen.¹ She found that in the Inglis multiple-response vocabulary test,¹⁷ eliminating the illustrative sentence, which normally appears before the word to be defined, gave no smaller scores than on a subsequent administration of the regular test with the context left in. It is still possible that other types or amounts of context would make some difference, but at least it is quite striking to find that it made none in this case.

2'. *Correction for Guessing in Multiple-response Tests.*—Annen similarly found that on equivalent samples of dictionary words, the average scores for various criteria of knowledge (merely checking a word as known, defining the word in their own terms, giving a sentence to illustrate its meaning) all agreed within eight per cent of the total as estimated from a multiple-response recognition test after correction for guessing. This correction assumes that on a four choice multiple-response test a subject knows certain words so definitely as to require no guessing. On the remainder of the words he would have one chance in four if he simply guessed without any knowledge at all. The actual score (S) which he would make on this type of test would then be made up of the number of words definitely known (K), plus one-quarter of the remaining words (R) in the test. This formula is then written $S = K + \frac{1}{4}R$. From this we may deduce that the actual number of errors (E) is equal to only three-fourths of the words on which it was necessary to guess, and the formula may then be written $K = \text{total words in test} - \frac{4}{3}E$. This assumes that it is proper to count several instances of partial knowledge as equivalent to one instance of complete knowledge, perhaps a debatable point, but as shown by the fact that these corrected recognition scores were not more than eight per cent higher than use scores involving no guessing, the error, if any, is a minor one.

The findings on the relative unimportance of context and the lack of any great difference in difficulty of use or recognition criteria suggest that knowledge of a word is ordinarily "all or none," that is, if you can recognize it at all you can also use it. This might be explained by the fact that the range of difficulty among words is very great. The majority of simpler words have been both recognized and used many times. The most difficult words have neither been recognized

nor used, and so there remains only a limited number of words for which we have partial knowledge.

(b) *Criteria of Use*.—A second criterion of knowledge is that of use, either by defining a word in your own terms or by a synonym, or by using it in a sentence to illustrate its meaning. A test of this type is easy to set up, simply giving the directions and the standards to be used in scoring. In our case the standards were approximately the same (genus-species) as for the recognition tests. This type of count, based upon a dictionary sampling, should be called a test of "possible use," to distinguish it from that of actual use as found for instance in studies of author's vocabularies in published works. A further distinction in the criterion of actual use would be that of use in writing or use in speaking, since we probably are less formal and apt to restrict ourselves to simpler words in our speaking vocabularies.

A modified use test which is practicable along with a recognition test is to have the subject mark a *W* before each item which he has actually used in some previous writing, and an *S* for each item previously used in speaking. The total of the words marked both *W* and *S*, plus those marked *W* or *S* alone gives a sample for estimating the total vocabulary of actual use. In the same way the subject may mark a *P* to indicate that he could possibly use a given word if necessary. In estimates of both actual and possible use, credit should be given only for those words in which the multiple-response recognition answer was also correct.

Annen,¹ using shorter lists than ours, reports moderate positive correlations, about +.40 between tests employing the four following criteria: Checking a word as known, written definitions, use in illustrative sentences, and multiple-response recognition. As mentioned previously, the group average for student vocabularies on these various tests agreed within eight per cent. This study is being repeated by Poindexter with our present list of words.

In addition to the definition of a word and the criterion of knowledge, a measure of vocabulary must also define the standard of accuracy to be demanded in giving credit for knowing a word. Perhaps the best general criterion is to be able to give or recognize a synonym or an equivalent term, and correctly to distinguish this from similar words (species) in the same larger group (genus) and also from words having a fallacious resemblance in spelling or sound. The correct use of a word in an illustrative sentence would involve a similar

type of discrimination and could be judged by the same general criteria.

3'. *Control of Incidental Cues in Recognition Tests.*—In the case of recognition tests our preliminary experiments in the construction of multiple-response tests indicated that the alternative responses must be selected very carefully according to a number of principles in order to avoid either an excessive degree of difficulty, through too fine distinctions in meaning, or the occurrence of "give away" items which would allow credit for incidental cues other than legitimate knowledge of word derivations. Our set of rules was obtained from the leads given by introspective reports of a few well-educated adult subjects who described the manner in which they selected the probably correct answer on each item of the test. After revision of the faulty items in line with these findings, the tests were then given to larger groups of college undergraduates to make sure that each item was actually functioning, that is, receiving some votes, and that the subjects in the highest quartile of the group tended to make fewer wrong answers on each item than did the members of the lowest quartile (the quartiles being selected on the basis of scores from the test as a whole). The easiest items have not yet received this careful analysis because they are too easy for our subjects, being known by practically all college groups, and our youngest subjects in this part of the study included only ages twelve to eighteen years.

The rules adopted for standardizing the difficulty of the discriminations to be called for in our recognition test eventually worked out as follows:

RULES FOR SELECTION OF MULTIPLE CHOICE ANSWERS

- (1) Choose words on parallel levels under the next larger class group, as given by Roget's *Thesaurus* or a similar classification.
- (2) Include words similar in sound or root.
- (3) Balance the length of words or phrases as between correct and incorrect choices, both on the same item and throughout the test.
- (4) Keep the same part of speech for all four items.
- (5) Avoid the use of "give-away" answers such as opposites or similarities in sound or root.
- (6) Keep the level of difficulty of the answers lower than that of the word which is to be tested (judging by Thorndike's *Teacher's Word Book* of twenty thousand words where possible).

Keeping in mind these various definitions of a word, the criteria of knowledge, and the standards of accuracy called for, we may review in Table III the results of previous experiments on the measurement of vocabulary size. Note especially that in these studies the size of the estimated vocabularies is roughly proportional to the size of the dictionary from which the sampling was taken, and that none of these investigators has used a dictionary as large as the Funk and Wagnalls unabridged edition of four hundred fifty thousand words employed in the present study. Our own present choice of that size was due in part to the information from an unpublished study by Hartmann, who showed that such a dictionary is necessary to provide adequate opportunity for measuring the vocabularies of educated adults.

5. *Interrelations with Other Abilities.*—Terman,³⁰ in his construction of the old Stanford-Binet intelligence test discovered that the section comprising a fifty-word vocabulary test was more predictive of the IQ to be obtained from the whole test than any other three parts of the test in combination. In keeping with this finding Seashore, Stockford and Swartz²⁸ found the following correlations between a preliminary form of our present test (Form A, which was one-third longer) and other educational tests:

<i>r</i> vocabulary <i>vs.</i> three speed of reading tests .02 to .04	<i>N</i> = 83
<i>r</i> vocabulary <i>vs.</i> Otis Advanced Intelligence Test Form B .49	
<i>r</i> vocabulary <i>vs.</i> A.C.E. Intelligence Test .57	<i>N</i> = 43
<i>r</i> vocabulary <i>vs.</i> reading comprehension test on psychology .60	<i>N</i> = 100 (approx.)

Evidently, then, vocabulary is closely related to or an essential part of the complex abilities known as general intelligence and reading comprehension, while speed of reading is largely an independent variable, at least within the limited range of college students.

II. OBJECTS OF THE PRESENT EXPERIMENTS

- (A) To construct a set of tests for estimating the total size of general English vocabularies in terms of dictionary entries as previously defined. These tests include:
- (1) A multiple-response recognition test for all commoner basic words, the most important group.
 - (2) A supplementary use test for knowledge of rare basic words where only a few are likely to be known.

TABLE III.—PREVIOUS ESTIMATES OF TOT

Author	Date	Reference*	Basis of estimate	Kind of test	Grammar school			
					2	3	4	5
Dean Farrar		7, 13, 21, 24	"Listened for a long time together to the conversation of three peasants who were „ gathering apples . . .	Use-spoken				
66								
D'Orsey, A. D.		7, 23		Use-spoken				
Bard, W. R.	1891	3	Total words needed in a foreign country to get along	Use-spoken				
Marsh, G. P.	1872	7, 20	Counts of authors' writings	Use-written				
Holden, E. S.	1878	14	Counts of authors' writings and concordances	Use-written				
Kirkpatrick, E. A.	1891 1907	18, 19	Webster's Academic Dictionary 28,000 words	Checking recognition	4,480	6,620	7,020	7,860
Doran, E. W.	1902	7	International Dictionary—Abridged	Checking recognition				
Babbitt, E. H.	1907	2	Unabridged dictionary of 100,000 words					
Bonser, F. G. Burch, L. H. Turner, M. R.	1915	4	Webster's Academic Dictionary 28,000 words Webster's Elementary School Dictionary; 44,000 words	Checking recognition				10,886 Group A Group B
Terman, L.	1916	30	Laird and Lee's Webster's vest-pocket dictionary; 18,000 words	Oral—definition Use		3,600 MA 8		5,400 MA 10
Gertlach, F. M.	1917	11	Funk and Wagnall's New Standard Dictionary; 250,000 words	$\frac{3}{4}$ 4-choice recognition $\frac{1}{4}$ definition Use				
Brandenburg, C. C.	1918	5	Webster's Academic Dictionary 28,000 words	Definition Use	4,000	5,429	6,887	8,207
Neher, H. L.	1918	22	Laird and Lee's Webster's vest-pocket dictionary; 18,000 words	Definition Use				
Holley, C. E.	1919	15				3,144	5,220	6,030
Symonds, P. M.	1926	29	E. L. Thorndike's Test of 1921 based on his <i>Word Book</i>	5-choice recognition— Completion Use			4,000 5,000	6,000 6,900
Gillette, J. M.	1927	12	Used Terman's test Dictionary of 209,000 words	Checking recognition Checking recognition				
Cuff, N. B.	1930	6	Webster's Shorter School Dictionary 35,000 words	4-choice recognition		7,425	10,395	12,460
Seashore, R. H.	1933	1, 27	Funk and Wagnall's Dictionary 145,000 words— derivatives in unabridged	4-choice recognition			

* The number given in the "Reference" column refers to the bibliography at the end of this paper.

AL ENGLISH NON-TECHNICAL VOCABULARY

Grammar school			High school				Fresh- man	Sopho- more	Junior	Senior	Average adults	Superior adults
6	7	8	9	10	11	12						
											100	
											300	
											600	
											3,000 to 10,000	Shakespeare 15,000 Milton 8,000 Italian grand opera repertoire 600 to 800
8,700	10,660	12,000	13,400	15,640	16,020	17,600	18,720	19,000	20,120			
11,634	13,675	17,138	18,865					15,750	19,400	21,500		
								50 to 60,000				
13,216 21,648 17,600	17,038 30,184 26,520	17,052 27,720	18,704 31,120						37,092	35,376		
	7,200 MA 12			9,000 MA 14							11,700	13,500
				66,050				73,650	85,300	86,925	94,075	76,250
9,613	11,445	12,819	13,504	15,340	13,974	14,975						
			9,000	10,080	11,520	11,700						
7,470	8,478	9,378	10,296	10,926	11,565	11,752						
7,900	10,000	11,250	13,200									
8,900	10,600	12,100										
						women men	8,880 9,240	9,325 10,860	10,130 13,040	10,700 12,700		
13,955	14,910	16,800	19,160	20,930	21,245	21,840		127,800				
								65,000			65,000	80,000

- (3) Alternative use or recognition criteria for a small sample of derivative words which are of less importance and for which this is only a preliminary study.
- (B) To determine the average size, range, variability, and norms for such measurements among college undergraduate students.
- (C) To determine the accuracy of such samplings and measurements as shown by the percentage of variation between odd- and even-numbered items of our main test, and by the correlation between odd and even halves of the test.
- (D) To obtain preliminary data on the size of vocabularies of use, including actual use, written and spoken, and possible use.
- (E) To investigate further the interrelations of the size of recognition vocabularies with other intellectual abilities such as reading comprehension, reading speed, and general intelligence, and the attainment of college grades.

III. PROCEDURE IN THE PRESENT EXPERIMENTS

In setting up the tests for measurement of individual differences in knowledge of the commonest meaning of words the following steps were employed:

A. Construction of a sample list of words from Funk and Wagnalls' unabridged *New Standard Dictionary*.

(1) Master sample: Every third basic (heavy type, marginal) item down from the top of the first column on the left-hand pages of the dictionary was taken, giving a total of one thousand three hundred seventy-eight, of which fifty-eight were prefixes, suffixes, abbreviations, and other unusable entries, leaving a list of one thousand three hundred twenty useful words.

(2) Subdivision into test samples: The master sample of one thousand three hundred seventy-eight words entries was then divided into four shorter equivalent forms by placing the first word (alphabetically) in Form 1, the second word in Form 2, and so on in rotation. Since there were slight variations in the numbers of prefixes, etc. which were eliminated, the four forms finally totalled three hundred thirty-one, three hundred thirty-one, three hundred twenty-nine, and three hundred twenty-nine words, respectively. This slight variation in actual length of the four samples is of no importance because we need

to know only the proportion of the sample which can be correctly defined, in relation to the size of the entire dictionary.

In order to arrange the words according to difficulty we next gave the four lists to groups of undergraduate students to define in their own terms or to use in illustrative sentences. All words which were correctly defined or used by at least one person (using a liberal scoring) were included in a list of common basic words, with which we are principally concerned henceforth. The remaining unknown words, together with all proper nouns, obsolete, archaic, provincial, and other words similarly designated by the dictionary were included in a supplementary list of rare basic words for each form.

(3) Preparation of Form 1* for use in testing:

(a') The one hundred seventy-eight common basic words of Form 1† were next put into the form of a four choice multiple-response test, (Part 1), the alternative answers being chosen in accordance with our preliminary findings on the way in which such items function (cf. p. 24). Three successive revisions of the test were tried out in mimeograph form and each was revised in accordance the results of the item analyses as previously described.

(b') The one hundred fifty-eight rare basic words from Form 1 were listed separately as Part 2 of the printed test, and the following directions given for use criteria: "Supplementary List of Proper Names, Geographical Locations and Rare Words. Define, identify, or use in illustrative sentences, being sure to give all meanings known, and compare with usage given in manual, from F. and W. unabridged dictionary."‡ Although the use criterion is slightly more difficult than that of recognition, the average number of correct answers (about three for college students) was so small as to make the error almost negligible in counting this as an addition to the recognition scores on Part 1. In this test the observers were only credited if they gave the particular meaning listed by the dictionary for that particular item,

* Published by the authors, Evanston, Ill.

† In order to insure maximum reliability we first included one-third of Form 4 along with Form 1, the composite being known as Form A (the first printed version). When experiments showed that this form was slightly too long for use in class periods, the additional items from Form 4 were dropped and subsequent printed editions of the test have included only the words from Form 1.

‡ At present the score on this part may be simply estimated from the score on the previous list (Part 1) according to manual of directions.

even if it had other commoner meanings. The directions took account of this by calling for all meanings for all words known. The one hundred seventy-eight words of Part 1 and the one hundred fifty-eight words of Part 2 together constitute $\frac{1}{505}$ th of all such basic words in the dictionary.

(4) Construction of Part 3, test for derivatives: Since it is still an open question as to whether or not derivative terms should be considered as separate words, we conducted only a preliminary study on this list, which included all terms (including compound terms) indented and in medium type under our basic words in the dictionary. Since the full list of words was found to involve too great a burden upon the subjects tested, we further reduced the list by taking only every tenth word, beginning with the tenth. This left a sample of forty-six words which then constituted Part 3 of our test. This sample represented $\frac{1}{4450}$ th of all such derivative words in the dictionary. The directions for Part 3 were as follows: "Derived Terms. (Abbreviated list.) For each word or phrase in this list see if you could think of (or write, if requested) a specific instance in which it would be correctly applied so as to illustrate its meaning. In compound or technical terms you should know the main points, (especially those underlined in the scoring manual) not merely a general idea of its meaning. Mark each item plus or minus, then correct this scoring after comparison with the definitions from the manual."

The rating criterion was justified by the fact that the authors had previously shown that multiple choice scores of basic and derivative words in an earlier test were correlated to the extent of $r = .60$, so that if only a moderate degree of accuracy is desired we may simply estimate a person's score on Part 3 from his score on Part 1 (as given later in Table V).

A slightly higher degree of accuracy may be obtained by having the subject rate his knowledge for each word in comparison with the definition later read to him, or where still greater accuracy is desired, he can write out the definitions or illustrations to be scored by the experimenter from the list of correct definitions.

By having our Ohio State University group of two hundred twenty-seven undergraduate students mark their status at ten-, twenty-, and thirty-minute time limits we have obtained preliminary norms for scoring the vocabulary test under conditions emphasizing speed. In this scoring the number of words attempted in a given time minus

1.33 times the number of errors equals the number of words correctly identified without benefit of guessing.

TABLE IV.—DECILE RANKS FOR TWENTY-MINUTE TIME LIMIT
 NO. OF WORDS KNOWN IN PART I
 (ATTEMPTS—1.33 × ERRORS)

	DECILE
128-141	10
122-127	9
118-121	8
114-117	7
111-113	6
107-110	5
103-106	4
98-102	3
92- 97	2
50- 91	1

IV. RESULTS AND DISCUSSION

The results of the three parts of the test are summarized in Table V, which gives the means, standard deviations, ranges, and decile ranks as obtained from various groups of college undergraduates. Parts 1 and 2 are based upon tests of two hundred thirty-seven Ohio State University students in an undergraduate course in educational psychology, and one hundred sixteen Northwestern undergraduate students, both groups including principally freshmen and sophomores. Part 3 has not yet been adequately standardized, the norms given being based upon only one hundred twenty Northwestern freshmen and sophomores who employed the self-rating criterion. Another similar class of sixty students, including more freshmen, had previously employed both the self-rating and the written definition criteria, knowing that they would be compared. This group gave practically equal results for the two criteria, but the average (78,000) was lower than the average for the larger group (100,000) which employed only self-ratings. This may have been due in part to two other factors; the considerably larger proportion of freshmen in the smaller class, and the implication of a more stringent standard of accuracy. Further studies are under way to provide more representative norms for this part of the test.

Preliminary work with children in the first eight grades of school has indicated the following points:

(1) All such children can be tested by the criterion of use, either definitions or sentence illustrations, on our list of words in Form 1,

TABLE V.—ESTIMATION OF COMPLETE VOCABULARY FROM PART I SCORES (UNLIMITED TIME)

Part I, errors (not corrected for guessing)	Part I, basic common	Part II, basic rare	Part III, derived	Total	Decile
21-31	69,900	2,525	120,150	192,575	10
32-34	65,200	2,020	113,475	180,695	9
35-37	63,200	1,768	106,800	171,768	8
38-39	61,500	1,768	102,350	165,618	7
40-42	59,800	1,515	97,900	159,215	6
43-44	58,100	1,263	93,450	152,813	5
45-47	56,500	1,010	87,220	144,750	4
48-50	54,500	758	80,100	135,358	3
51-55	51,800	505	72,535	124,840	2
56-78	49,800	0	62,300	112,100	1

TABLE VI.—DECILE RANKS FOR ACTUAL SCORES ON PARTS II AND III (UNLIMITED TIME) WHEN NOT MERELY ESTIMATED FROM PART I

Raw score (correct) Part II	Raw score (correct) Part III	Decile
5	27-35	10
4	25-26	9
3.5	24	8
3.5	23	7
3	22	6
2.5	21	5
2	19-20	4
1.5	18	3
1	16-17	2
0	10-15	1

which is arranged in order of difficulty so that the whole list need not be employed for younger children.

(2) In giving Form 1 as a recognition test for children below the fourth grade and for very dull children above that grade, the test as a whole or at least any difficult portions should be given orally. Where children tend to give the last answer or the simplest word rather consistently it is probable that this portion of the test is too difficult for them, and the examiner may then read off the next list of

words to be defined, testing only those on which the subject indicates that he knows something of them. In this case the number of words actually tried is substituted for the total of one hundred seventy-three in the scoring formula for basic words. (Total common basic vocabulary of Part 1 = Number of Words attempted minus $1.33 \times$ errors, times 505.) Scores on use for rare words in Part 2 are almost completely absent at this level except for occasional geographic names of especial interest to the group.

(3) In the case of certain very young or very dull children the tendency to be misled by superficial similarities among the wrong multiple-choice answers in a recognition test was great enough to reduce their corrected score to zero, even though they indicated by use criteria that they definitely did know a number of the easier words. In order to avoid emphasis upon this negative factor of misinformation it is, therefore, recommended that use criteria be employed whenever such results are encountered. It seems probable, however, that the majority of children above the fourth grade can take the multiple-response test by omitting the fifth page of Part 1. Further details for younger children must await more intensive investigation, but it is entirely reasonable to believe that their vocabularies will be described in thousands rather than in hundreds of words.

In summary of the results for college students, it may be pointed out that the average college undergraduate in our groups recognized (after correction for guessing, as previously described) thirty-five per cent of the common basic words, one per cent of the rare basic words, and forty-seven per cent of the derivative words. The proportionate figures for a test on the whole dictionary would be 58,186; 1,662, and 95,888, or a grand total of 155,736 words according to our dictionary definition of a word. The range of total scores for college undergraduates was 112,100 to 192,575.

In comparing the scores from odd and even items of Part 1 basic words it was found that the average difference in the absolute scores of the two halves was only four per cent, and, of course, the full test should be even more reliable. For our restricted range of college undergraduates the corresponding reliability coefficients have also been determined.

As measures of reliability we may cite the following correlations on two hundred twenty-six to two hundred thirty-seven undergraduate students in educational psychology.

r odd-even, unlimited time, Part 1, Form 1 .71 \pm .02; corrected* .83 \pm .02
 r odd-even, twenty minutes, Part 1, Form 1 .87 \pm .01; corrected .93 \pm .01
 r odd-even, thirty minutes, Part 1, Form 1 .73 \pm .02; corrected .84 \pm .02

Intercorrelations of various scoring systems on the test are as follows:

r 1st ten minutes *vs.* unlimited time .45 \pm .04
 r 1st twenty minutes *vs.* unlimited time .67 \pm .02
 r 1st ten minutes *vs.* 1st twenty minutes .84 \pm .01

All three of the above intercorrelations are, of course, partly artificial since the score for longer time includes the score for shorter time. These are given only to illustrate the effect of various scoring systems, and it is obvious that speed and power are to some extent different variables. Either emphasis could be used if desired. Decile ranks for twenty-minute time limits are given in Table IV.

This apparently low reliability is due not only to the restricted range of ability among our subjects, but also to the fact that the first two pages of the test are too easy for college students, being rarely missed except through carelessness by anyone above the first decile. The students, therefore, spend a fairly large proportion of their time on items which were non-discriminative for this group, and only the next two pages were within their range of difficulty, with a few of the abler students making a fair showing on page 5 of the test. It would involve no serious error for college students to be credited with correct answers to all words on pages 1 and 2 without requiring them to take these portions of the test. Pages 1 and 2 are necessary for use with younger or less highly selected groups, and page 5 is necessary for able students even among undergraduates.

When the previously mentioned small group of sixty undergraduates was asked to mark each item on Part 1 as having been previously used in writing (*W*), or speaking (*S*), or both (*WS*), and also if they could possibly use it (*P*), the averages for the various types of use vocabularies were as shown in the table on page 35.

To these figures we may add the previous findings on use criteria of rare basic words in Part 2 (1,662) and derivative words in Part 3 (95,888), which together with the above figures for Part 1 gives an average use vocabulary of 153,202, almost as large as the average recognition vocabulary (155,036).

* Corrected by Brown Spearman prophecy formula for full length of test.

For fifty Northwestern freshmen the vocabulary score in unlimited time correlated $r = .61$ with the reading comprehension section of the Ohio State University entrance examination when given in twenty minutes instead of the standard thirty-six-minute time limit. For the same students the correlation with the Iowa English Examinations

Both written and spoken	42,622
Written only	1,465
Spoken only	1,768
	<hr/>
Total, actual use	45,855
Possible	9,797
	<hr/>
Grand total, all use criteria	55,652

(language usage) also = .61. Both of these values are about the same as that found between such tests or general intelligence tests and college grades.

On one hundred twenty Northwestern freshmen and sophomores Form 1 vocabulary scores (unlimited time) correlated $r = .62$ and $r = .19$ with the Iowa Reading Comprehension Test, Forms A and B, respectively. The correlations of Form 1 with college grades in elementary psychology was $r = .57$ ($N = 120$).

V. SUGGESTIONS FOR FURTHER RESEARCH

Although the field of vocabulary testing is a fairly old one, it has not been developed either intensively or extensively. Among the promising lines of further investigation which have been opened by the present study the following would seem to be immediately feasible:

- (1) The standardization of the present test for children of school grades I to XII, and comparison with previous results by other methods on development of vocabulary at the pre-school age.
- (2) Longitudinal studies in which the vocabulary development of the same children is followed over a period of years, or by Laswell's method of interlapping age ranges for shorter periods of time.
- (3) The adaptation of the test for special age or talent ranges by determining which words may be omitted for determinations of absolute size as being either entirely too easy or too difficult. A similar project has already been tried for the selection of the one

hundred most diagnostic words in the test for the measurement of relative differences only. This test is being tried out in various personnel uses, *e.g.*, the prediction of success in acquiring technical training through a correspondence school.

(4) The measurement of "depth" or richness of meanings, including multiple meanings of words (now under way). This may or may not be related to our measurements of single commonest meanings. Similarly the ability to discriminate fine differences between series of closely related words may or may not be a separate factor.

(5) The measurement of effectiveness of various types of word training, *e.g.*, vocabulary drills, word games, wide reading, extensive writing of different types, etc. Form II of the test is now being prepared for use in such measurements of initial and final status in reference to various types of learning.

(6) Similar vocabulary tests of technical knowledge in specialized fields are being studied by Rexroad* and others as measures for prediction of scholastic success, and as measures of attainment in given courses of study. These may be constructed for any given field and are certain to be of interest to teachers.

(7) In the analysis of individual differences in verbal abilities the vocabulary test can be employed either as a straight power test for level of ability, or as a speed test. Intercorrelation and factor analysis with other verbal tests such as Thurstone's battery of primary abilities would be desirable.

BIBLIOGRAPHY

- †1. Annen, Sister Ida, O. S. B.: *The Construction, Analysis, and Evaluation of a Vocabulary Measure*. Unpublished Masters Thesis, University of Oregon, June, 1933.
2. Babbitt, E. H.: "A vocabulary test." *Pop. Sci. Mon.*, Vol. LXX, 1907, pp. 378.
3. Baird, W. R.: "The study of languages." *Memory*, May-June, 1891.
4. Bonser, F. G., Burch, L. H., and Turner, M. R.: "Vocabulary Tests as measures of school efficiency." *School and Society*, Vol. II, 1915, pp. 714-718.
5. Brandenburg, G. C.: "Psychological aspects of language." *J. Educ. Psychol.*, Vol. IX, 1918, pp. 313-332.
6. Cuff, N. B.: "Vocabulary Tests." *J. Educ. Psychol.*, Vol. XXI, 1930, pp. 212-220.

* Unpublished study.

† References 1, 7, 8, and 20 contain much more extensive bibliographies.

- *7. Doran, E. W.: "A study of vocabularies." *Ped. Sem.*, Vol. xiv, No. 4, 1907, pp. 401-438.
- *8. Eckerson, L. E.: *The estimation of individual differences in the total size of English recognition vocabulary*. Master's Thesis, University of Southern California Library, Los Angeles, 1938.
9. Funk and Wagnalls: *New Standard Dictionary of the English Language*. New York: Funk & Wagnalls Co., 1929.
10. Funk and Wagnalls: *New Standard Dictionary of the English Language*. New York: Funk & Wagnalls Co., 2nd Vol., Unabridged edition, 1937.
11. Gerlach, F. M.: *Vocabulary Studies—False Definition Test*. Colorado College Studies in Educ. and Psychol., No. 1, 1917.
12. Gillette, J. M.: "Extent of personal vocabularies and cultural control." *Sci. Mon.*, Vol. xxxix, 1927, pp. 451-467.
13. Harlan, C. F.: "The technical vocabulary of psychology." *J. Educ. Psychol.*, Vol. xvii, 1926, pp. 554-557.
14. Holden, E. S.: "On the number of words used in speaking and writing." *Bull. of the Phil. Soc. of Wash.*, Vol. ii, 1874-1878.
15. Holley, C. E.: *Holley Sentence Vocabulary Scale—Grades 3-12*. Bloomington, Ill.; Pub. Sch. Publ. Co., 1919.
16. Horn, E.: *A Basic Writing Vocabulary—10,000 words most commonly used in writing*. Coll. of Educ., Univ. of Iowa, n.d., No. 4.
17. Inglis, A.: *Inglis Test of English Vocabulary*. Boston: Ginn, 1923.
18. Kirkpatrick, E. A.: "The number of words in an ordinary vocabulary." *Science*, Vol. xviii, 1891, pp. 107.
19. Kirkpatrick, E. A.: "Vocabulary test." *Pop. Sci. Mon.*, Vol. lxx, 1907, pp. 157-164.
- *20. McCarthy, D.: Language Development of the Pre-school Child. In *A Handbook of Child Psychology*. Worcester, Mass.: Clark Univ. Press., 1931, pp. 278-315.
21. Müller, F. Max: *The Science of Language*. Two volumes. New York: Scribners, 1891.
22. Neher, H. L.: "Measuring the vocabulary of high school children." *School and Society*, Vol. viii, 1918, pp. 355-359.
23. Nice, M. M.: "On the size of vocabularies." *Am. Speech*, Vol. ii, 1926, pp. 1-7.
24. O'Connor, J. and Inglis, A.: *Worksample No. 95*. Boston: Ginn.
25. Ogden, C. K. *The Basic Dictionary*. 2nd Edition. London: K. Paul, Trench and Trubner & Co. Ltd., 1932, p. 106.
26. Ogden, C. K.: *Basic English. A General introduction with rules and grammar*. K. Paul, Trench and Trubner & Co. Ltd., London, 6th Edition, 1937, p. 95.
27. Seashore, R. H.: *Qualitative Aspects in the Improvement of Science Teaching*. Univ. of Oregon, 1935.
28. Seashore, R. H., Stockford, L. B. O., and Swartz, B. K.: "A correlational analysis of factors in speed of reading tests." *School and Society*, Vol. xlvi, 1937, pp. 187-192.
29. Symonds, P. M.: "Size of recognition and recall vocabularies." *School and Society*, Vol. xxiv, 1926, pp. 559-560.

* References 1, 7, 8, and 20 contain much more extensive bibliographies.

30. Terman, L. M.: *The Measurement of Intelligence*. Boston: Houghton Mifflin, 1916.
31. Thorndike, E. L.: *A Teachers Word Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading For Children and Young People*. New York: Bur. of Publ., Teachers Coll., Columbia University, 1932.
32. Williams, H. M.: "Some problems of sampling in vocabulary tests." *J. Exper. Educ.*, Vol. 1, 1932, pp. 131-133.