THE VOCABULARIES OF SCHOOL PUPILS

EDWARD L. THORNDIKE

Teachers College, Columbia University

It would seem to be a very simple matter to find out how many words pupils of a given age or school grade know, but there is a very wide diversity among the estimates made by those who have studied the matter. In this article I shall review the previous investigations, point out certain principles of method, and show the limits between which the truth probably lies, with special reference to the word-knowledge of pupils in Grade 9.

Kirkpatrick, Doran, Babbitt, Bonser, Gerlach, Brandenburg, Terman, Neher, and Holley have estimated the number of words known by testing individuals with a sample list chosen from some dictionary, and then multiplying the number that the individual knew by $\frac{n \text{ in dictionary}}{n \text{ in sample}}$. The tests used have been (a) to check the words known, (b) to define them, (c) to use them in sentences, and (d) to select the right meaning from alternatives.

Kirkpatrick (’07) used a list of 100 words selected from Webster’s Academic Dictionary. The subjects were given a printed list and were asked to mark the words they knew with a plus (+) sign, those they did not know with a minus (−) sign, and doubtful ones with a question mark. There was the added direction to “count as known all words that you would not, as to their meaning, need to look up in the dictionary if you saw them in a sentence.” The resulting estimates are shown in Table 1.

Doran (’07) reports that he used word lists varying from one thousand to several thousand words, and required written or oral definitions. He selected all words on a page, except such unusual words as hardly anybody would know, and selected the pages either at random, or more often in a certain order, as every twenty-fifth or fiftieth page. In school No. 1
<table>
<thead>
<tr>
<th>Grade</th>
<th>Kinke</th>
<th>Doran</th>
<th>Bonser</th>
<th></th>
<th></th>
<th></th>
<th>Brandenburg</th>
<th>Gerlach</th>
<th>Neuber</th>
<th>Holiday (Estimated)</th>
<th>Teman</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4480</td>
<td></td>
<td></td>
<td>10856</td>
<td>D</td>
<td></td>
<td>4000</td>
<td>5429</td>
<td>4000</td>
<td>3144</td>
<td>M. A. 8</td>
</tr>
<tr>
<td>3</td>
<td>6620</td>
<td></td>
<td>13216</td>
<td>21648</td>
<td>D</td>
<td></td>
<td>6887</td>
<td>5220</td>
<td>6030</td>
<td>3144</td>
<td>M. A. 10</td>
</tr>
<tr>
<td>4</td>
<td>7020</td>
<td></td>
<td>15038</td>
<td>30184</td>
<td>D</td>
<td></td>
<td>8207</td>
<td>5220</td>
<td>7470</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7860</td>
<td>4192</td>
<td>11634</td>
<td>26520</td>
<td>D</td>
<td></td>
<td>9613</td>
<td>5220</td>
<td>8478</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8700</td>
<td>3849</td>
<td>13075</td>
<td>27720</td>
<td>D</td>
<td></td>
<td>11445</td>
<td>6030</td>
<td>9378</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>10660</td>
<td>7937</td>
<td>17138</td>
<td>31120</td>
<td>D</td>
<td></td>
<td>12319</td>
<td>6030</td>
<td>9378</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>18400</td>
<td>13400</td>
<td>18865</td>
<td></td>
<td>D</td>
<td></td>
<td>13504</td>
<td>10296</td>
<td>M. A. 14</td>
<td>9000</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>15640</td>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td></td>
<td>15340</td>
<td>10926</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1H</td>
<td>16020</td>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td></td>
<td>13974</td>
<td>11565</td>
<td></td>
<td></td>
<td>Av.</td>
</tr>
<tr>
<td>2H</td>
<td>17600</td>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td></td>
<td>14975</td>
<td>11700</td>
<td></td>
<td>Superior</td>
<td></td>
</tr>
<tr>
<td>3H</td>
<td>15720</td>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td></td>
<td>76250</td>
<td>11752</td>
<td></td>
<td>Adult</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>15750</td>
<td></td>
<td>D</td>
<td></td>
<td>19400</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>School</td>
<td>15750</td>
<td></td>
<td>D</td>
<td></td>
<td>21500</td>
<td></td>
<td></td>
<td>Superior</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>College</td>
<td></td>
<td></td>
<td>D</td>
<td></td>
<td>J. 35376</td>
<td></td>
<td></td>
<td>Adult</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td></td>
<td>S. 37092</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 11700 and 15400 for low-grade classes in this normal school.
and in the normal school, he used Webster’s High School Dictionary; in school No. 2, he used Webster’s International. In the latter school the tests were given at intervals for two or three months. He estimated by multiplying in the same fashion as Kirkpatrick, allowing for the unusual words. His estimates appear in Table 1.

Babbitt used Kirkpatrick’s methods, with the exception that the dictionary was an unabridged one, containing over 100,000 words.

Bonser (’15) used Kirkpatrick’s list in his first testing of the Speyer School. In his second test of the pupils at the Speyer School, and in his testing of pupils at Edgewater, Horace Mann, Teachers College, and elsewhere, he used a list of 150 words selected from Webster’s Elementary School Dictionary, which contains 44,000 words. In both cases he used Kirkpatrick’s methods. His estimates appear in Table 1.

Brandenburg (’18) used a list of 200 words, one for every 140 in Webster’s Academic Dictionary. The instructions were as follows: “In the space after each word that you know, write a sentence using the word correctly. Place a cross before each word that you do not know.” His estimates appear in Table 1.

Terman (’16) used a list of 100 words, selected from Laird and Lee’s Vest Pocket Dictionary, containing 18,000 words, and tested pupils individually by oral definition. He estimated the number of words in the subject’s vocabulary by multiplying his score in the test by 180. His estimates are for Stanford Mental Ages.

Neber (’18) used Terman’s methods and also took his words from the dictionary from which Terman chose the Stanford list.

Gerlach (’17) used “the first word of the third column of every third page and thirtieth page” of Funk and Wagnalls’ New Standard Dictionary, making 1000 in all. “Biographical and geographical terms were omitted.” In the case of 400 of the least familiar of the thousand words, the subjects were asked to define those which they knew. Each of the remaining 600 words were supplied with four definitions, only one of which was correct. The subjects were asked to
check the correct definitions. He estimated that his list was representative of 250,000 words. He deducted from the number of the 600 which were correctly marked, one third of the number wrongly marked, and added the number of the 400 which were correctly defined. The resulting number he multiplied by 250.

Holley ('19) used Terman’s list of 100 words, the task being to select the right word from four, as in:

A gown is a ... string ... animal ... dress ... plant.
An orange is a ... dress ... animal ... fruit ... hornet.

He did not estimate total vocabulary, but his list is, as stated, the same as Terman’s and his averages multiplied by 180 give the numbers of Table 1.

The discrepancies in the results of Table 1 taken at their face value are obviously very great. For the average first-year high school pupil, for example, we have a range from 9000 to 66,650. The same investigator reports 7937 and 17,138 for the average pupil in Grade 8, according to the school and dictionary used. Bonser, using Kirkpatrick’s method, estimates 18,704 and 32,120 for Grade 8 in the same school, according to the dictionary used, whereas Kirkpatrick’s result was only 12,000 or 13,400 (Grades 8 and 9, respectively, in an elementary school with a ninth-grade course).

These discrepancies are partly due to the dictionary used. For example, Terman’s estimates are of how many correct definitions an individual would give if he were tested with the 18,000 words of the Laird and Lee book. They would be enormously increased if plurals, comparatives and superlatives, past tenses, and familiar derivatives are to be counted as words known. As to root words and unfamiliar derivatives, his estimates are too low if the individuals know such words that are not contained in the Laird and Lee dictionary. This is almost certainly true of the superior adult. Up to the mental age of twelve, this source of error is probably nearly negligible. For his primary purpose of a test of intelligence, of course, the Laird and Lee list is satisfactory.

The discrepancies are partly due to the method of selecting
words in a dictionary. To take a random sample of 100 words from a dictionary containing, say, 28,000, we should take words 1, 281, 561, 841, etc., or words 2, 282, 562, 842, etc., or the equivalent. Nobody, it appears, has done this. For example, Gerlach took the first word on every third page and on every thirtieth page. This or any similar method will overweight the sample with the more important words, unless the dictionary used gives equal space to all words, since the more space a word has given to it in a dictionary, the more chance it has of being the first word on a page. For example, suppose that certain words have each a half page devoted to them, and that certain other words have each one fiftieth of a page devoted to them. Then the chances are one in two that one of the first set will be the first word on any page where it is, while the chances are only one in fifty for one of the second set. Nobody, except possibly Terman, seems to have noted this factor of error, nor the corollary that in the long run the larger the dictionary used, the more the overweighting will be, since the bigger dictionaries show the greater variations in the amount of space given to different words.

Gerlach’s list shows very clearly the effect of this error, since his list of a thousand words supposed by him to be a random sampling of about 275,000, contains nearly a third as many words from the first ten thousand for importance as from the entire remaining 265,000 words of less importance.

Brandenburg states that his words were selected “at equal intervals throughout the dictionary.” This probably means equal intervals of space, since if he had counted off equal numbers of words defined, he would probably have given himself credit for the labor involved, and stated the reason for the expense of time. In his list of 200 words, 57 are from the

---

1 This assumes that “being the first word on a page” means being the first word any part of whose definition forms the beginning of a page. If it means being the first word the beginning of whose definition is the first beginning of any definition on the page, this calculation of chances does not apply; but there is still a heavy overweighting, because important words often appear each as several words listed separately in the dictionary.

2 It should be noted that for Gerlach’s main purpose, the comparison of the sexes, the error was not important.
first 5000 for importance—approximately, and only 163 from the remaining 23,000.

It is to be hoped that future workers who estimate range of vocabulary will consider the nature of the list which they sample and the method of sampling more thoroughly than has been the case in the past.

The discrepancies are partly due to the method of the test. Pupils may quite honestly think they know a word and so check it for Kirkpatrick or Bonser, who could not define it well enough to pass Terman's test on it. Pupils may be unable to define “majesty” who could check the correct word in “Majesty, refers to ... dresses ... kings ... countries ... climates.”

Both questions, “What is meant by knowing a word?” and “How many words are there in the English language?” are really extremely complex. We can define what by degree of knowledge we mean by the test we frame for any given word and the key for scoring the test. Terman, Gerlach, and Holley do so, and Brandenburg presumably had a key for scoring the sentences that would serve the same purpose. The selection form of test tends to represent a rather vague, inadequate, and loose knowledge, but it can be made to represent a very exacting standard, as when five distinctions are required or temptations to error are introduced, as in

habitat dweller bodice prodigality habit home
cord dispute harmony famine firewood string

In spite of the apparent discrepancies in the estimate of Table 1, a critical study of the original reports enables us to make an approximate estimate of the word-knowledge of the average pupil at the beginning of Grade 9, or the first year of a four-year high school course.

“To know a word” shall mean for us “to be able to define it passably, or at least to recognize a definition of it among three or four wrong definitions.” 1 We shall not count plurals in s, verbs in s or d or ing or n, past tenses in d, comparatives and superlatives in or, est, r, and st, or adverbs in ly. We shall

1 This is lamentably vague, but is the best we can do at present.
consider first the estimates made as a result of tests, not the opinions of individual pupils as to what they know. Terman’s mental age of fourteen may be taken as nearly equivalent to the average first-year high school pupil, since the Terman mental age fourteen is probably well above the chronological fourteen-year-old of the country as a whole. His estimate of 9000 is low, since some of the pupils would know some words not in the Laird and Lee. Neher’s 9000 for first-year high school pupils is low for the same reason. Holley’s 10,296 tends to be low for this same reason, but tends to be high because his selection tests are so easy. Brandenburg’s 13,500 is high because of the selection error described above, and because his criterion of use in a sentence is not at all exacting.

Gerlach’s 66,650 shrinks greatly because of this error, and because the 275,000 which he tried to sample included probably many of the forms which we have ruled out. The average first-year-high school pupil scored 266.6 (median 267) words on Gerlach’s test of 1000. If we assume that he knew the most important 266 of the 1000, his vocabulary would be about 11,000, since 240 of the Gerlach 1000 words are in the Thorndike list of the 10,000 most important words.\footnote{More than 240 are in the list, but some are used in the test with very rare meanings and so are not counted by us; e.g., bear, to force down the price of.} Allowing generously for the fact that he probably knew more words outside the Thorndike list than he failed on within the list, his vocabulary still would be under 14,000. So it seems fairly certain that the number of words (excluding the forms mentioned above and compounds which are obviously understandable from their component words) which are known well enough to define them passably, is between 9500 and 12,500.

Doran’s results cannot well be used, because we do not know just how to treat his allowance for the very unfamiliar words.

Our own results with the Thorndike tests are in accord with this. The median score is 65 in a test of 100 words, 76 words of which are from the Thorndike 10,000 and 24 of
which are from the next 10,000 for importance, the task of selecting synonyms being deliberately made hard by the requirement of some fine distinctions. If the words known were the most important, they would correspond to about 9000 words as a total vocabulary. Allowing for the fact that the pupils would know more outside the 1000 than they would fail on within the 10,000, we have from 10,000 to 11,500 as an estimate.¹


¹ There are many matters to be considered in such estimates, such as the allowance for chance successes in the selection form of test, the case of recognizing a definition in comparison with framing one, and the allowance for the proper names which make up part of the Thorndike list of 10,000. We have had those in mind, but make no claim to rigorous or exact treatment of them. Such a treatment is hardly worth while with the data at hand: When we have 20,000 or more words rated for importance, and tests for each thousand or two thousand, and a rigid scoring, and perhaps tests sampling adequately the different meanings of the same word, much greater refinement of treatment will be in order.