

## A TENTATIVE STANDARDIZATION OF A COMPLETION TEST

WALTER R. MILES AND JULIAN E. BUTTERWORTH

*The study supervised and the report edited by Professor Irving King, University of Iowa*

The aim of the following study was to determine as far as possible the relative difficulty of the various elisions of one of the widely used completion tests, "Where the Dandelions Went." It will often be desirable to have such an evaluation of the separate blanks, if the test is used for anything but the roughest of measures. If a time limit is set for the test which will not permit of its being finished there must be some method of scoring what is done in terms of the difficulty of the various parts. The correction of errors also requires a definitely determined series of weights for the various elisions or blanks.

This particular test was selected, not as being superior to others, but simply to illustrate a method of evaluating which may be applied to any test of this character.

Hitherto, most of the users of such completion tests have tended to assume, in practice, at least, that each blank space is of equal value with every other blank. Ebbinghaus, for instance allowed 1.0 for each elision properly filled and for each failure to fill an elision .5 was deducted. He also deducted 1.0 whenever an elision was improperly filled. Terman and Childs who used a similar completion test (The Strength of the Eagle) in their "*Tentative Revision and Extension of the Binet-Simon Measuring Scale of Intelligence*"<sup>1</sup> recognized on the other hand that all parts of the selection are not of equal difficulty. They therefore modified it by printing all the blanks of equal length, thus giving no suggestion of the length of the word or part of word to be supplied, and further, they divided the selection into approximately four equal parts, the first part mutilated to the extent of 33 1-3%; the second, 45%; the third, 54%; and the fourth, 66 2-3%. In scoring, each elision, if correctly supplied, was given a credit of 6, 8, 10, or 13, accordingly as it occurred in parts I, II, III, or IV of the test.

The present investigators felt that the difficulty of any particular portion of this test is not so much dependent upon the

<sup>1</sup> JOURNAL OF EDUCATIONAL PSYCHOLOGY, Vol. III, p. 198ff.

degree in which words are omitted as upon the nature of the omissions. For instance, elided verbs are less easily supplied than elided adjectives or adverbs.

The reader may here raise the objection that it is scarcely worth while to evaluate accurately every omission in a completion test; that it should be sufficient to rate a set of papers according to the fractional part of the whole performed. If the test is expected to yield only a very rough measure, such a procedure would undoubtedly do well enough. But the results of the test have seldom been interpreted as merely rough measures. Usually the investigator attempts to assign to each paper a definite numerical grade. If, however, the omissions are not of equal difficulty it would seem that comparisons or correlations based on grades determined on the basis of the equal difficulty of all elisions, should be open to serious question.

Whether it is worth while or not to attempt the careful evaluation of a completion test, such as is here proposed, this study will show at least two things: First, that it is possible to assign fairly accurate values to the different blanks of such a test, and second, that a numerical grading means very little if it does not take into account the relative difficulties of the different elisions. We might also mention as a third point that our study shows how complicated even the so-called simple tests are.

The data for this study were collected in the spring of 1912 from twenty-five observers, nearly all of whom were men, and who were students or members of the faculty in the State University of Iowa. If twenty-five such selected observers seem to anyone to furnish an unsuitable basis for the standardization of such a test, it is hoped that the number will at least be sufficient to furnish a suggestion as to a method of approach which may be feasible in a more extended inquiry.

The subject in each case sat at a desk conveniently lighted. Nearby one of the experimenters had a muffled telegraph key for signalling to his colleague who was attending a suitable time recording apparatus in an adjoining sound proof room. The subject, after reading a set of written instructions,<sup>2</sup> was given a sample mutilated text, which served for him both as instruction and practice. The real test blank was placed face down on the desk and on a signal was turned over and the blanks were filled in by him.

<sup>2</sup> Instructions are given in the appendix.

The sentence was used as the working unit. There are eleven of these sentence units in this test, and the average of the times of the twenty-five subjects necessary to complete each unit are as follows, in percentages of the time required for the whole test:

Value of unit	1	12.8%	6	10.8%
	2	11.7	7	9.2
	3	8.5	8	7.8
	4	11.0	9	8.3
	5	11.2	10	5.2
			11	3.4

As might be expected, the mean variation from the average time taken for each unit is in every instance rather large, emphasizing the need of a larger number of observers. The following table gives the percentage in the case of each unit which the average deviation was of the average time.

Unit	1	51%	6	50%
	2	55	7	29
	3	47	8	47
	4	39	9	55
	5	33	10	42
			11	32

In order to determine the average difficulty of the different words within each unit the subjects were asked, as each unit was completed, to distinguish three degrees of difficulty in the words supplied. They were asked to do this by underlining once and twice those words which seemed of more than average difficulty and those which seemed especially difficult respectively. The time required for doing this was not large but it was measured and deducted from the total time for the test. The relative difficulty of each elision was then estimated by the frequency with which it was marked by all subjects as of one of the three degrees of difficulty. There were, of course, considerable differences among the observers as to the relative difficulty of the various words. The first five words of the test are typical and were marked thus by the twenty-five observers:

Elided word	1	2	3	4	5
Average difficulty.....	21	21	24	19	24
More than average difficulty.....	3	3	1	4	1
Very difficult.....	1	1	0	2	0

The question which was next considered was what value should be assigned to a word of more than average difficulty and what to a word of great difficulty in proportion to the value of the word of average difficulty. At first thought, one may be

tempted to assign such arbitrary values as 1, 2, 3, or 1, 2, 4, or to say that very difficult words are so many times as hard as words of average difficulty, and that words of more than average difficulty should be assigned a value about midway between the two extremes. Introspection, however, seems to indicate that the words of great difficulty are more than three or four times as hard as the average words. It usually happened that the observer took as much time to determine a single hard word as to arrive at a solution of all the rest of the unit, and it frequently happened that even more time was spent on one hard word than on all the rest of the unit.

Now, if most of the words in the whole test had been marked as *very difficult*, only a small value, on the basis of 100% for the whole test, could have been given each word, no matter how much time it took the observer to settle upon it as the right word. On the other hand if few words were so marked, a relatively higher value could and probably should be assigned to each word marked as very hard. In other words, these three degrees of difficulty are purely subjective degrees determined by each observer, and hence the less frequently words are marked as of more than average difficulty, the more difficult they probably are compared with words of average difficulty, and the more value they should be assigned in comparison with the easier words. Hence the conclusion that the best method of assigning the values to the different blanks of the test should be based on the frequency with which each degree of difficulty was specified by the different subjects.

In answer to the possible criticism that this method is invalid because each subject's standard of difficulty would be likely to change in the course of the test, we may say that while it is probably true that few words of great difficulty were of exactly equal difficulty, it must be remembered that the subject could always tell when the solution went easily, when it was very difficult, and when it was moderately difficult. At least it seems that this method should give us more nearly the correct values than those we should get if all elisions were assumed to be equally difficult.

Acting, then, on the assumption that our method of assigning values to the three grades of difficulty is admissible, we next combine the records of all our subjects and have the following percentages of frequency:

83.2% of the elisions were marked as of average difficulty.  
 9.3% of the elisions were marked as of more than average difficulty.  
 7.5% of the elisions were marked as very difficult.

On the basis of frequency we may say that the very difficult elisions are 11.1 times as difficult as words of average difficulty, ( $83.2 \div 7.5$ ), and that words of more than average difficulty are 8.9 times as difficult as words of average difficulty, ( $83.2 \div 9.3$ ). For convenience these proportions should be reduced to fractions of unity, as follows:

Elisions of average difficulty	.048
Elisions of more than average difficulty	.427
Elisions of greatest difficulty	.533

Now the frequency with which each elision in the first unit was indicated by the observers as of the various degrees of difficulty was as follows:

Elision No.	1	2	3	4	5	6	7	8	9	10	11	12
Average difference.....	21	21	24	19	24	24	21	14	20	18	18	17
More than average difficulty...	3	3	1	4	1	1	3	5	3	4	2	5
Very difficult.....	1	1	0	2	0	0	1	6	2	3	5	5

If now we substitute for these frequencies the values for each degree of difficulty as indicated above, we have the following results

Elision No.	1	2	3	4	5	6
Average difficulty.....	1.008	1.008	1.152	.912	1.152	1.152
More than av. difficulty...	1.281	1.281	.427	1.708	.427	.427
Very difficult.....	.533	.533	.000	1.066	.000	.000
Totals.....	2.822	2.822	1.579	3.686	1.579	1.579

  

Elision No.	7	8	9	10	11	12
Average difficulty.....	1.008	.672	.960	.864	.864	.816
More than av. difficulty..	1.281	2.135	1.281	1.708	.854	1.281
Very difficult.....	.533	3.198	1.066	1.599	2.665	2.665
Totals.....	2.822	6.005	3.307	4.171	4.383	4.762

Therefore, the percentage of value (computed on the basis of 100% for the unit) due each elision of this first sentence is: 7.14; 7.14; 3.99; 9.35; 3.99; 3.99; 7.14; 15.19; 8.38; 10.56; 11.09; 12.05. But, since this first unit of the test is entitled to only 12.8% of the credit of the entire test, it is desirable to reduce the above determined values to a basis that will render them directly comparable with the values of all the other units of the test. This we have done by taking 12.8% of each of the values belonging to this unit. We thus obtain as the real values for

each of the twelve elisions of this unit the following: .9; .9; .5; 1.2; .5; .5; .9; 1.9; 1.1; 1.4; 1.4; 1.6.

The procedure which has been given in detail above for the first sentence of the test, was followed in getting the real values of the elisions in each of the other units. The values thus computed are given below inserted in the blanks of the test itself.

Wh.9 Willy .9 two .5 old he 1.2 .5 .5 red farm-h.9 1.9th 1.1 yard 1.4 front .14 1.6. The dan.7 were 1.5 th1.8 there; so that 1.6 2.0 lo2.2 yellow instead of 1.9.

One bright .4 Willy's m.6 put 1.7 1.6 straw .7 .9 sent him .7 .6 .6 yard to .7. She knew 1.2 1.5 had .8 high 1.3; and .4 could not o1.0 .4 gate; so he 2.2 2.3. W.5 it .3 time .5 him .3 8. .3 nap and 1.0 went .5 c.7 him, she noticed that .4 gr1.3 ma1.6 of the da.9 1.1 1.5. She won.7 where 1.3 1.1; but, as 2.2 1.2ld not talk much, 1.1 did not 1.5 him 1.7 them.

A short .6 after, while .4 was 1.1 .7 his crib, his mamma went .9 1.6 1.8 1.2 wa1.0. When .5 buck.7 came .9 full 1.0 2.2 1.6 top was all y.4 with .5. Look.8 down into .8 1.4 .6 could .6 no 1.1 at all, on1.3 1.7. 1.1 1.1 .9 wonder, then, where .5 blos.4 had 1.2. Willy .3 been .4 busy try.3 .3 fill .5 .7 .9.

If the reader will take the trouble he may compare his own introspective judgments of the difficulty of the various elisions with the values here assigned. He should remember, however, that the judgments of a single individual will frequently differ from these values which represent the combined judgments of twenty-five individuals.

It was the original purpose of the writers to correlate the degree of difficulty with the degree of correctness in the assignment of final values to the various elisions. While both of these elements must be considered in the marking of any paper, the writers decided, finally, to leave these two elements, difficulty and correctness, separate. There seemed to be no way of determining the time that would be required to make each observer's work absolutely correct. This being true, it seemed safer to base the value of the elision only upon the combined introspections of the subjects as to their difficulty than to assume some arbitrary relationship between difficulty and correctness.

Since accuracy should be considered in using the table of values here presented, we suggest two possible methods of doing it. First, give no credit for what is not correct or as good as correct. Second, give full credit for what is correct or as good as correct, and part credit for what gives sense but is not strictly correct. In neither case would credit be given for elisions left unfilled or for those filled so as to have no connection with the rest of the thought. We believe that the second method is the

fairer. An elision or a group of elisions may be so filled as to be not entirely unrelated to the rest of the thought, and yet not be sufficiently correct to warrant full credit. Thus the seventh unit, when correctly filled in, should read:

"She wondered where they were; but, as Willy could not talk much, she did not ask him about them." One observer filled it in to read thus: "She wondered where they were; but, as Willy could not talk much, *they* did not *call* him *before* them."

Then again, what proportion of the value of an elision should be deducted for words such as these in italics, which give sense but are not correct? Either of two methods might be followed: (1) an arbitrary value, say, one-half, might be deducted; (2) a credit proportional to the accuracy of the words could be assigned. The aim of the first method is to average all degrees of correctness, assuming that in the long run one word will be as much better than this arbitrary value as another is less. The second method aims to make the assignment of value a matter of individual judgment. If the papers contain many errors, we believe that the second method of evaluating would be the fairer. However, in the case of the papers used in this study there were so few errors that the first method of dealing with mistakes was adopted. The difference in the results produced by the two methods applied to our papers would be very slight. It is easily possible, however, that with some papers the results would be very different, and where this is the case the second method should be selected as the better.

We believe that the more accurate results which may be secured by using a table of individual values for the various elisions based upon the method herein described is sufficient ground to justify the little additional labor that will be required to use or apply them.

It must be remembered that the values of the various sentence units and of the various elisions which are here suggested are based on the reactions of scholastically advanced adults. It is a question whether the proportions of difficulty would remain the same if less mature subjects were tried out in the same way with the test. We have one bit of evidence to support the view that difference in maturity and in scholastic standing would yield much the same relative values. One of the investigators, Dr. Butterworth, the following winter gave this same test to one hundred first year normal school girls as a mass test. They had tried on the previous day the completion test called "The Strength

of the Eagle," to give them some familiarity with this sort of exercise. The units were separated by red lines and each subject was provided with a piece of notched card-board so that she could cover up all of the test except the unit on which she was engaged. All observers began on each unit at the same time and as each finished she raised her hand and was given her time by the experimenter from a stop-watch. Although some time was lost in the raising of the hand and the getting of the time, it was very nearly equal for all subjects. The times needed by the hundred subjects for each of the units were thus secured and averaged and percentage values of the various units were computed in precisely the same way as those given on page 4, above. The values last secured from the hundred girls are strikingly like those obtained by more accurate methods from the twenty-five more advanced subjects in the University of Iowa. Both sets of values are here presented for comparison.

	University subjects	First year normal school girls
Unit 1	12.8	10.9
2	11.7	14.8
3	8.5	9.2
4	11.1	11.7
5	11.2	12.6
6	10.8	8.5
7	9.2	9.6
8	7.8	6.4
9	8.3	7.9
10	5.2	4.9
11	3.4	3.6

## APPENDIX

Instructions given to all subjects to read before beginning the test.

### THE COMPLETION TEST

The observer will supply all missing words or portions of words. The length of each dotted line shows approximately the length of the word or portion of a word to be supplied.

Try first to make sense, second to fill in every omission, third, to work as rapidly as possible.

Finish one unit or sentence at a time. These are marked off by red lines.

Be rather certain before beginning to write.

After finishing each sentence, immediately underline the words which gave difficulty. Place two under any word or words which gave you noticeably the most difficulty in the sentence.

The title of this story in which you are to fill omitted words is "Where the Dandelions Went."