

FRESHMAN TESTS AT THE STATE UNIVERSITY OF IOWA

IRVING KING AND JAMES M'CRORY

State University of Iowa

In the autumn of 1916 practically all the freshmen entering the College of Liberal Arts at the State University of Iowa were given a series of mental tests. There were 276 women and 268 men.

The tests selected were the following:

1. The Courtis Standard Arithmetic Test, Series B, covering problems in the four fundamentals. This test was given exactly as prescribed by Courtis. A single grade for speed was obtained by adding together all the "attempts." The total number of "rights" was divided by the number of "attempts" for the total percentage of accuracy.

2. A list of twenty-four "mixed relations" or analogies of a fairly difficult type (a set designed by Whipple).

3. Two tests of "opposites." The stimulus words were four hard lists used by Simpson and reported in his "Correlation of Mental Abilities" (1911). These words were rearranged into two lists of forty words each, in order from the easiest to the most difficult, according to the standard of difficulty determined by Gold and King, and published in the JOURNAL OF EDUCATIONAL PSYCHOLOGY, October, 1916.

4. The first three paragraphs of the Completion Test (Mutilated Prose) used by Simpson.

5. Visualization; the three inch cube test, each of the nine questions being weighted on the basis of previous experimentation and assigned a percentage value.

6. Range of Information; Whipple's one hundred technical words. The subjects marked the list to indicate first, which words they could define or explain, and secondly those words with which they were roughly familiar. They were then asked to write the definition or explanation of the first five words which they had indicated as being able to explain. On the basis of these definitions the grade for the entire test was com-

puted, one-fifth of the entire grade being deducted for each wrong definition.

7. Logical Memory, using the paragraph given in Whipple's Manual entitled "The Dutch Homestead." The paragraph was read to the subjects and they were then asked to write as much as they could recall. The papers were graded on the basis of the number of ideas they were able to reproduce according to the standard given in the above mentioned Manual.

All of these tests were given to the Freshmen in groups of from ten to twenty-five. They were timed with a stop watch in each test except in the last three as listed above. The grades for "Opposites," "Analogies," and "Completion" were computed by first marking the errors on each paper and then adding to the time score in seconds, 5 percent. of the time for each error. For instance if it took a student three minutes and twenty seconds to complete a test his score would be 200 which is the number of seconds. If he made on error it would add 5 percent. of that which would make his score 210.

SCORING OF RESULTS

In any investigation of this sort, where there are no standards established, it becomes necessary to set up arbitrary standards for grading. This was found to be the case with nearly every one of the tests given. There is a basis, however, for each method used.

Since each of the different tests had a system distinctly its own, there had to be a common basis so that the average of all the tests could be computed and so that the tests could be compared with each other. The following system was devised for such comparison. The median of each distribution of grades was found for each sex; the deviation of each individual from that median was next found; that deviation was multiplied by a factor obtained by dividing the median into 100. This method gave us the grades as multiples of 100. Zero was, of course, the middle case. The algebraic sum of these corrected deviations was then found and divided by the number of tests, which gave the average for all the mental tests. These averages were half above zero and half below, or in other words they were half positive and half negative.

Distribution of Medians and Ranges in the Various Tests

Test	Med. Girls	Med. Boys	% Boys Reaching Med. of Girls	Range Girls	Range Boys
Arith. Speed.....	50	52	60	19-97	23-97
Arith. Accuracy.....	74	77	55.4	16-100	24-100
Completion	335	378	36	770-159	841-154
Analogies.....	228	213	53	480-105	720-85
Opposites Easier.	350	384	41	840-144	1128-171
Opposites Harder.....	900	823	64	1512-318	1538-263
Opposites Comb.	1239	1215	54.1	2055-486	2658-551
Information.	25	28	61.6	0-59	0-78
Logical Memory	43	36	32	21-78	9-65
Visual Imagery.	33.5	42.5	68	0-100	0-100
Ages.....	19-62	19-61	50	16-30	17-25

In examining the third column of figures we observe that the boys slightly excel in several of the tests, the exceptions being the completion, easier opposites and logical memory. When the two opposites tests were combined, we find that more than 50 percent. of the boys reach or excel the median grade of the girls. In visual imagery test the boys excel the girls in the greatest degree. We note, however, that the median score made by the girls was only 33.5 out of a possible 100. In this test there were no girls who made a score of zero while there were eight who made a perfect score of 100. Using the median of the girls' scores as a basis and taking the gain of the median of the boys' score over this, we find that the boys accomplished 27 percent. more than the girls.

<i>Variability</i>		
Test	Boys	Girls
Arithmetic—		
a. speed.....	.17	.22
b. accuracy.....	.17	.16
Completion.....	.20	.18
Opposites.....	.13	.18
Analogies.....	.18	.23
Logical Memory.....	.22	.17
Information.....	.25	.32
Visual Imagery.....	.41	.37

We see from these results that the statement that boys are more variable than girls is not borne out as far as these tests are concerned. If we should add the total variation we find that the boys have 173 and the girls 183 which would make the boys less variable if that one figure were to be taken as a measure. When we divide the tests we find the boys are more variable in four and the girls more variable in the other four.

The above coefficients of variability were computed by dividing the standard deviation by the median.

DISCUSSION OF RESULTS

Arithmetic

The Courtis Arithmetic Test, Series B, which we used, has been given to more than a million school children. It is a test in the four fundamentals, and each of the four is usually graded separately. The writers pooled the four grades into a single index. The total number of attempts was taken as speed and the total number of right answers divided by the total number of attempts was taken as the index of accuracy.

Correlations

	Girls	Boys		Girls	Boys
Speed with Accuracy....	41	38	Accuracy with Speed...	41	38
Completion.....	22	25	Completion.....	16	13
Opposites ..	03	Neg.	Opposites.....	01	Neg.
Analogies.....	17	04	Analogies ..	22	18
Information.....	06	Neg.	Information ..	13	Neg.
Vis. Imagery.....	13	Neg.	Vis. Imagery.....	20	Neg.
Log. Memory.....	11	Neg.	Log. Memory..	13	Neg.
Test Average.....	51	56	Test Average.....	48	41
Uni. Grades.....	31	36	Uni. Grades ..	40	36

The above coefficients are all Pearson "r's". They will be used throughout the article.

The writers believe this test should be omitted from any series of tests designed for use with university freshmen. In the first place it takes too much time. Twenty-six minutes were given to this test, which means that by the time the instructions were given, practically thirty-five minutes of the two hours had been taken by this one test. Not only does it take too much time to give, but it also requires a great amount of time to score. There were 97 problems in the four processes and although the answer card is made as convenient as possible, nevertheless it is bound to take more time than the results warrant in this particular kind of testing. Another reason is that given by some of our critics, namely that it tests one particular function under conditions which are not normal. The writers feel that some arithmetic test should be given and possibly the Woody scale in addition would satisfy the conditions. This could be easily graded as it is a scale and not a test. It is what McCall would term a power test and not a speed test. We feel that the reason for including some kind of arithmetic test might be termed a social one. We find in reading the newspapers, magazines, books, or any current literature, that a certain amount

of mathematical knowledge is needed. Not mathematics as such, but the interpretation of statements mathematically given. Addition, of course, does not take care of all these cases and indeed it would be difficult to even list all of them, but since we cannot give a long test to cover all the cases we feel that addition will come the nearest of any of the processes to fulfilling the conditions.

The highest correlation we were able to get between arithmetical speed and any other one test was .25. Bell, at the University of Texas, got an inter-test correlation of .20 between addition and alternatives. Simpson got a raw correlation of .79 between adding and hard opposites. King, at the University of Iowa, tested 56 freshmen engineers. He got a correlation of .75 for attempts or speed and scholastic rank. He got .48 for accuracy and scholastic rank. McCall used 88 6B pupils in New York City. He got an "r" of .20 for adding and school mark. His "r" for arithmetic was .30. His arithmetic test consisted of six selected examples, but details as to their precise nature are lacking.

Completion

We find the completion test one of the two tests in which less than 50 percent. of the boys reached or excelled the median grade of the girls. We also find the boys grades distributed over a wider range than the girls.

Correlations

	Girls	Boys
Completion with Speed.....	.22	.25
Accuracy.....	.16	.13
Opposites31	.79
Analogies.....	.46	.58
Information.....	.21	Neg.
Vis. Imagery06	Neg.
Log. Memory.....	.26	Neg.
Test Average.....	.60	.77
Uni. Grades.....	.22	.41

The lowest correlation among the tests for the girls was with visual imagery and we also find a negative correlation for the boys. The P. E. for these coefficients is about .026. As far as these coefficients can be used, it would seem that for a single test the completion is better adapted to the boys. Although they have three negative correlations with the other tests, we find they have a higher coefficient both with the average of all the tests and the average university marks.

McCall in *School and Society* for Jan. 6th, 1917, gives some coefficients which are of interest in this connection. His article is entitled, "Correlation of some psychological and educational measurements, with special attention to the measurement of mental ability." One of his conclusions is as follows: "Meaning by mental ability a composite of all the measurements, the omnibus and completion tests correlate with it 1. and .96 respectively. That is to say, a perfect measure of an individual by omnibus or completion would be a substantially true index of his mental ability." McCall's subjects were 88 6B grade public school pupils in a typical elementary school in New York City. For his completion test he used the Trabue language scale. McCall got a coefficient of .21 with arithmetic (adding), while our coefficient with Series B is practically the same, .22 and .16 for speed and accuracy respectively for the girls and .25 and .13 for the boys. His coefficient was .52 for teachers' marks while ours was .22 for the girls and .41 for the boys. The difference in the age and mental attitude of these two groups must be taken into consideration when any comparison is made. From our results we are not ready to recommend the completion test for use in judging the university student in as thorough-going fashion as McCall does for the sixth grade pupil when he says "The completion test will be a substantially perfect measure of his mental ability."

Bell found in his study of the freshmen at the University of Texas that the completion test showed the highest correlation of any of the test scores with the class marks. The writers found that for the boys two other tests correlated higher than the completion tests with class marks and two others practically the same, while for the girls, the coefficient for completion was next the lowest of any of the tests with class marks.

Dr. King found when he tested a group of 16 Juniors in the College of Applied Science and correlated the results with the combined judgments of three instructors in that department, that he got a coefficient of .35 between the completion test and these combined judgments. The judgments were given simply as to the ability of the men in question.

The writers' results show no important sex differences in the completion test. No reason is assigned for the girls making the better score, on the whole, than the boys.

Opposites

In the easier opposites we find only 41 percent. of the boys reaching the median for the girls, but in the harder opposites we find that 64 percent. reached the median for the girls. In order to obtain a single index for the opposites we added the scores made in each test and called that the score for the opposites combined. When we distribute those scores we find 54.1 percent. of the boys reaching the median for the girls.

Correlation

Opposites with	Girls	Boys
Speed.....	.03	Neg.
Accuracy.....	.01	Neg.
Completion.....	.31	.79
Analogies.....	.52	.77
Information.....	.24	.56
Visual Imagery.....	.07	.56
Logical Memory.....	.32	.38
Test Average.....	.51	.88
University Grades.....	.45	.84

We see from the above coefficients that the opposites correlate rather highly with all the other tests for the boys, but not so highly for the girls. We do find, however, that the opposites correlate the most highly of any of the tests *with university marks* for both boys and girls.

We believe that in giving both the easier and harder lists of opposites we obtain a measure which would not be possible by using only one of the tests. Following is the opening paragraph of an article in the JOURNAL OF EDUCATIONAL PSYCHOLOGY for October, 1916, by Irving King and Hugo Gold, of the University of Iowa.

"Among the various mental tests the test of ability to give opposites has found much favor with a number of investigators. Such tests are supposed, if the words are easy, to measure, in some degree, the rapidity of controlled associations reactions; if the words are hard, including such words as "unless," "suave," "ignorant," etc., the test measures in some degree logical keenness in selecting the word which will express most nearly the contrary idea."

Simpson says that the easy opposites is a good test for *readiness for controlled associations*. King got an "r" of .24 between hard opposites (two of Simpson's lists of 20 words each) and

ability of the students as ranked by the three instructors mentioned before in this article. In the case of the 56 freshmen engineers King got a coefficient of .26 for hard opposites and scholastic rank.

Waugh at Beloit College found a correlation of .84 between class standing and opposites. .

Kitson, at the University of Chicago, found a correlation of .84 and .86 for the easy and hard opposites respectively, when correlated with the combined index of all his tests. These coefficients were obtained by the product-moments method. The coefficient, using the method of rank differences, as the writers did, was .53 when correlated with standings in the net score.

Kitson's word lists were very easy, his hard list being much easier than the writers' easy list. He defends the use of these easy lists by attempting to eliminate all words that are ambiguous. According to his opinion "these ambiguous words are productive of long pauses and incorrect responses because of conflict of impulses." This, it seems to the writers, is one of the things we are trying to test. Surely one who is able to make a quick, accurate, selective response has better mental ability than one who allows his mental attitude to be upset every time he is called upon to make such a decision. According to Kitson's results, 21 out of 40 made a perfect score of 100 percent. in the hard list, while 31 out of 40 made a perfect score in the easy list. Can we really compare these people when so many of them make a perfect score? If we are trying to determine the percentage of accuracy in that manner, then it seems it would be possible to select a still easier list or in other words, one of zero difficulty.

The writers take practically the opposite stand when selecting words for the opposites test. The answer card which we used was standardized to such an extent that it was fairly indicative of what the majority of the answers were likely to be. We feel that if we were to choose any *one* of our tests as a sign of the mental ability of the individual being tested, it would be this combination of the two opposites tests.

Logical Memory

Memory, perhaps, was the first function of the mind to undergo a test. Nearly everyone who has attempted to make a test

series has had in it some test which had to do primarily with memory. We used for our memory test a passage which was intended to test logical memory or memory for ideas in their logical order.

Correlations

	Girls	Boys
Logical Memory with		
Speed.....	.11	Neg.
Accuracy.....	.13	Neg.
Completion.....	.26	Neg.
Opposites.....	.32	.38
Analogies.....	.32	Neg.
Information.....	.31	.13
Visual Imagery.....	.31	.12
Test Average...	.53	.65
University Grades.....	.35	.40

Simpson found a correlation between estimated intelligence and memory of words of .93 and memory of passages of .35.

Kitson found a correlation of .29 between logical memory and the net score of all the tests. This was for the forty students in the College of Commerce in the University of Chicago.

In the group of Junior engineers mentioned elsewhere, King found a correlation of .17 between logical memory and estimated intelligence. He found a correlation of .20 between logical memory and completion while ours was .26 for the girls and slightly negative for the boys. He found a correlation of .27 for logical memory and opposites, while ours was .32 for the girls and .38 for the boys.

The writers feel that some test of memory should be incorporated in any series of tests. There does not seem to be sufficient difference between the results of immediate and deferred memory to warrant taking the time to give both tests. If these tests are to be practical at all they must have the so-called mechanical features refined to such a degree that it will minimize the effort and economize the time of those concerned in both the giving and the taking of the tests. The test for immediate memory would be better than the deferred as far as the time factor is concerned.

Visual Imagery

We find this test to be the one in which the boys excel the girls score to the greatest degree. We find the range for each of the groups to be the same, namely, 0 to 100. There were

nine in each group that made zero. There were eight girls who made a perfect score while only five boys made a perfect score. We do find, however, that 68 percent. of the boys reached or excelled the median made by the girls.

<i>Correlations</i>		Girls	Boys
Visual Imagery with			
Speed.....	.13		Neg.
Accuracy.....	.20		Neg.
Opposites.....	.07		.56
Completion.....	.06		Neg.
Analogies.....	.20		Neg.
Information.....	.23		Neg.
Logical Memory.....	.31		.11
Test Average.....	.72		.62
University Grades.....	.32		.21

Analogies

W. S. Miller, director of the university high school at the University of Minnesota, has probably done more than any other one person in using the analogies test. He gets a correlation of .79 with the analogies and class standing. He gave the tests individually, however, while ours were given in groups.

<i>Correlations</i>		Girls	Boys
Analogies with			
Speed.....	.17		.04
Accuracy.....	.22		.18
Completion.....	.46		.58
Opposites.....	.52		.77
Visual Imagery.....	.20		Neg.
Information.....	.28		Neg.
Logical Memory.....	.32		Neg.
Test Average.....	.72		.74
University Grades.....	.14		.40

This is a test in logical relation. For some reason it has not found favor with those who have attempted to test college freshmen. The Binet-Simon tests for general intelligence and the Yerkes Bridges point scale use it, though only for adult measurements. It seems to the writers that these tests measure something that is not found in other tests used in this series, and that it would be a mistake not to include them in a series of tests intended for use with college freshmen.

Range of Information

We find 61.6 percent. of the boys reaching the median of the girls in this test. We also find a wider range among the boys,

the highest grade being 78, while the highest grade of the girls was 59. This test, as we gave it and scored it, is at best a sort of a guess affair, the only check being the definitions which the student was asked to write on the back of the sheet. It would be perfectly possible for the student to check each of the 100 terms and then happen to know the meaning of the first five, in which case he would get 100 percent. This case, while possible, is not very probable as is seen when we look at the median performance of these people and find it was only 25 words out of 100 for the girls and 28 for the boys. Unfortunately no one else has published any results of this test computed as ours were so that figures for comparison are not available.

Correlations

Information with	Girls	Boys
Speed.....	.06	Neg.
Accuracy.....	.13	Neg.
Completion.....	.21	Neg.
Opposites.....	.24	.56
Analogies.....	.28	Neg.
Visual Imagery..	.23	Neg.
Logical Memory..	.31	.13
Test Average.....	.60	.65
University Grades41	.44

The chief value of these tests lies in the supplementary light they cast in disciplinary cases. By means of these results the Dean of Women and the Adviser of Men were able to determine whether a student failing in his work, was failing from lack of mental ability or for other causes. Below will be given the records made by certain individuals. These are not typical cases, but are selected from among the best and the poorest records in the *mental tests*. It is noticeable that they have good and poor records in their academic work as well.

Case No. 1.

This is a boy who made the best record in his University work. He was the only member of his class of some 500 students who made four "A's." He is carrying English, German, Mathematics, and Zoology. He was below the median in the mental tests in only one test that being speed in arithmetic. He had an accuracy of 88 percent., but he attempted only 51 out of a possible 97. *His rank in his academic work was first while in mental tests it was sixth.*

Case No. 2. This is the record of a boy who had next to the best record in both mental tests and in his academic work. He received two "A's" and two "B's," his course being English, German, Mathematics, and Economics. He was above the median in all the mental tests and was able to complete all the 97 examples in the arithmetic test, being one of three in both the boys' and girls' groups to complete the entire list.

Case No. 3. This is the record of a boy who ranked first in mental tests and who received next to the highest grade in his academic work. He ranked first in the tests, although he attempted only 45 of the 97 examples in arithmetic and was only 65 percent. accurate.

Case No. 4. This is the record of a boy who made zero for his university work, although he took only two of the four examinations. He ranked 259 in mental tests out of 268 boys. He was above the median in only one test, that being range of information. He is a boy who was classed by those who examined him most carefully as below normal. He was in the hospital for some time, having adenoids and tonsils removed. Later a hernia operation was performed. This boy left home when he was 15. His father is a veterinary surgeon and from all indications the home environment was good. When he left home he went west, then served in the navy for four years. During that time he went around the world and naturally acquired a great deal of information the average boy of his age would not have. He was advised not to re-register for the second semester.

Case No. 5 is a boy who ranked 267th out of a possible 268 in mental tests and 234th in academic work. He also was asked not to re-register. The trouble with his academic work was ostensibly ill health, as he had 26 absences excused on that account. He came to the university with very poor preparation, and showed little resistance to distracting influences. He was a constant cigarette smoker, and was consistently reported by his professors as on the delinquent list.

Case No. 6 was a boy who ranked 258th in mental tests and 214th in his academic work. He was utterly lacking in initiative, had no ambition, and failed in all his academic work.

Case No. 7 was a boy who ranked 245th in the tests and 233d in academic work. His trouble, perhaps, is poor preparation,

as he is the fourth student to come to the university from his high school who had to be sent home on account of poor work. He was reported as delinquent from the very first. He was overwhelmed and showed no ability to adjust himself to the situation. He was asked not to re-register for the second semester.

Case No. 8 was a boy who ranked 253d in the tests and 234th in academic work. There was a distinct lack of preparation.

We were not able to obtain as much personal information about the girls as about the boys. The tests tend to pick them out, however, quite as accurately as the boys.

Case No. 9 is a girl who took first rank in the mental tests and also first in her academic work. There were seven girls who made the same grades in academic work, *viz.*, two "A's" and two "B's." This girl is seventeen years of age.

Case No. 10 is a girl who ranked second in mental tests and made the best grade in her academic work. She is sixteen, the youngest member of the freshman class. She has never been reported on the delinquent list and her work is of a generally high order.

Case No. 11 is a girl who ranked fourth in mental tests and received next to the highest grade in her academic work. She is an exceptionally strong student and has never been reported on the delinquent list.

Case No. 12 is a married woman who was the oldest member of her class. She ranked 276th in mental tests, the lowest rank. She left on account of illness before the end of the semester, but was doing failing work when she left.

Case No. 13 is a girl 20 years old who ranked 268th in mental tests. She failed in all her academic work and was asked not to register for the second semester. She was utterly lacking in ability, as she had tried the work the previous year and failed.

Case No. 14 is a girl who ranked 273d in mental tests, and failed in all her academic work. She also was doing the work the second time, and had been reported as failing from the start. She had tried a private tutor to help her with her work, but was simply lacking in ability.

Students Who Were Asked Not to Re-register

There were thirteen girls and fifteen boys who were asked not to re-register for the second semester on account of failing work

during the first semester. In each of these lists we find the median falling in the lower 25 percent. of the ranks in mental tests. The highest rank of those who were asked to leave was a girl who had an average of 13.3 points above the median for her class, and whose rank was 75. This girl failed in more than half of her academic work and was constantly reported to the Dean of Women as on the delinquent list. She was appealed to through her mental test record but to no avail. It was simply a case of too many dates with young men friends, and too many movies. She was perfectly capable, according to the testimony of the Dean, but her attention was on other things.

Distribution of Averages and Ranks of the Boys

	Mental Test Ave.	Mental Test Rank
Case No. I	3.3	111
II	1.2	126
III	.6	129
IV	— 7.6	164
V	—12.1	186
VI	—13.2	191
VII	—13.3	194
VIII	—13.7	199
IX	—15.2	206
X	—17.1	212
XI	—22.7	231
XII	—31.4	245
XIII	—32.0	250
XIV	—35.1	253
XV	—41.0	258

Total Cases 268

Range of Ranks in this Group 111—258

Median Rank 199

Girls

	Mental Test Ave.	Mental Test Rank
Case No. I	13.3	75
II	10.8	83
III	1.3	128
IV	— .1	142
V	— 3.3	162
VI	— 7.9	182
VII	—10.5	193
VIII	—11.2	199
IX	—13.8	208
X	—25.0	242
XI	—37.3	258
XII	—41.8	263
XIII	—45.6	268

Total No. Cases 276

Range of Ranks in this Group 75—268

Median Rank 193

CONCLUSIONS

I. This series of tests show fairly good correlations with academic work for the first semester.

II. Inter-test correlations show that we are testing a variety of mental functions.

III. Results show that this series, with the possible exception of the Courtis Arithmetic Test, Series B, is suitable for testing freshmen.

IV. These tests pick out those of poor mental ability much more accurately than those of good mental ability.

V. Certain mental tests would be valuable for colleges to use in dealing with disciplinary cases.

VI. The median performance of the boys slightly excels that of the girls in six of the eight tests.

VII. Sex differences in mental functionings are negligible as far as these mental tests are concerned.