



Intelligence Examinations for College Entrance

Edward L. Thorndike

To cite this article: Edward L. Thorndike (1920) Intelligence Examinations for College Entrance, The Journal of Educational Research, 1:5, 329-337, DOI: [10.1080/00220671.1920.10879060](https://doi.org/10.1080/00220671.1920.10879060)

To link to this article: <http://dx.doi.org/10.1080/00220671.1920.10879060>



Published online: 15 Dec 2014.



Submit your article to this journal [↗](#)



Article views: 2



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

Journal of Educational Research

Editorial Office, Urbana, Illinois

Editor-in-chief B. R. Buckingham, University of Illinois

VOLUME I

MAY, 1920

NUMBER 5

INTELLIGENCE EXAMINATIONS FOR COLLEGE ENTRANCE¹

EDWARD L. THORNDIKE

Teachers College, Columbia University

For an ideal examination of the intelligence of candidates for college entrance we might set the following specifications:

I. *Significance*.—The score should correlate³⁴ as closely with future achievement in college as is possible. This maximum possible correlation will not be 1.00, since achievement in college is due in part to health, to freedom from personal worries, and to various moral qualities as well as to intellect. Also, the magnitude of the correlation coefficient will depend on the range of the intellect of candidates, being smaller as that range is restricted. If all the eighteen-year-olds in the country were educated for college, tested, and given a trial in college, we might perhaps expect a maximum correlation as high as 0.75 to 0.85. Within the restricted range of those who complete a high-school course and actually become candidates, we may expect as a maximum 0.55 to 0.65, possibly more. A correlation above 0.50 is probably an improvement over the attainment of standard systems of accrediting high schools or of entrance to college by examination in school subjects.

II. *Continued significance*.—The correlation should maintain itself year after year after the nature of the intelligence examination is known in spite of whatever coaching candidates undertake. This implies, in theory, that such

¹ A paper prepared for presentation before the National Association of Directors of Educational Research at its meeting at Cleveland, Ohio, February 26, 1920.

special coaching shall give training that is nearly as useful for future college achievement as for success in the test, or that the tendency to undertake, and the ability to profit by such special coaching shall be very closely correlated with ability to achieve in college. In practice, it means that the examinations year after year should be totally different in content until a great many of them are available; and should differ in form where there is danger that adaptation to the form of the examination may enable a dull boy to secure a good score.

III. *Comparability of standards and results.*—(a) The examinations given year after year should be closely equal in difficulty; (b) the scoring should be by standard keys so that the same paper scored by different examiners will receive substantially the same score; (c) the administration of the examination should be simple and independent of the personality of the examiner.

IV. *Probable error of the examination itself.*—For the examination's main purpose, the probable error is already taken account of in the correlations with college success. For that purpose, in fact, a large probable error would be encouraging, for it would mean that by repeating the examination, using a different form, the correlation with college success could be notably raised. For certain special purposes such as comparing individuals or groups, and for the justification of shortened forms of the examination, the smaller the probable error the better.

A half-year's experience with the Thorndike Intelligence Examination for High School Graduates may increase our confidence that these specifications may be met to a rather close approximation. This examination was devised at the request of the faculty of Columbia College, where it is an optional means of entrance for boys suitably recommended. Three forms of it have been used to date by Columbia, one in June 1919, one in September 1919, and one in February 1920. About seven hundred candidates have taken it, though over half of these were already legally admitted by the content examinations of Columbia University or the College Entrance Examination Board or the New York Regents. It has also been used for various purposes by a number of other institutions.

SIGNIFICANCE

Four years must elapse before an exact accounting on this point can be made, but the following facts may be noted:

Of eleven boys at Columbia, reported to the dean's office for inability to do college work in the early weeks of the year, all had notably low scores in the intelligence examination. Of a score or more so reported as a result of the mid-term records, all but two had low scores. The defective college work of these two was by common consent not due to intellectual defect.

In the case of a group of eighty odd candidates in another college, the correlation between the score in the Thorndike examination and the first term's college record was over 0.65. The highest 40 percent in the examination showed 74 percent of high grades in college work and 26 percent of low grades. The lowest 60 percent in the examination showed 27 percent of high grades and 73 percent of low grades. Of the 16 percent receiving the lowest grades in the examination, none received any college grade above C and 86 percent failed in five hours or more of their college work.

In a third institution, where the entire student body was tested, the members of the three upper classes showed no scores below 60 in the examination, whereas a third of the freshmen (admitted by certificate) scored below 60. The distributions given in the accompanying table (Table I) give a correlation of approximately 0.55 between test score and continuance beyond the freshman year. The returns received from the first and third of these institutions are a random selection from trials of the examination. Those from the second institution were reported to the writer. They may have been reported because of the positive evidence of the success of the examination, and so may be unduly favorable.

CONTINUED SIGNIFICANCE

If we take the best of the test material that has demonstrated high correlations, such as selected series of opposites, mixed relations, and cleverly devised problems and completions, and give it without fore-exercise, we can undoubtedly get high correlations, but we may be unable to maintain them.

Our opposites, relations, and problems of later years will probably not be so ingenious as those which we chose first. Moreover, part of the first high correlation may have been due to the testing

TABLE I. THE SCORES, IN PERCENTAGES, OF STUDENTS IN
A WESTERN PROFESSIONAL SCHOOL IN THE THORNDIKE
INTELLIGENCE EXAMINATION FOR HIGH
SCHOOL GRADUATES

(This institution admits on certificate from accredited schools but finds that a considerable percentage of the students so admitted do not satisfy its standards, the elimination being rather large. The figures below show that apparently an ability denoted by a score of 60 or more in the Thorndike examination is required for meeting the standards of this institution. None of the sophomores, juniors, and seniors fall below 60, whereas about one-third of the freshmen do.)

Score in Thorndike Examination	Freshmen Percent	Sophomore Junior Senior Percent
30	2.6
35	2.6
40	12.9
45	2.6
50	2.6
55	10.3
60	15.4	5.5
65	10.3	16.3
70	12.9	10.9
75	5.1	8.1
80	7.7	19.0
85	10.3	10.9
90	7.7	8.1
95	10.9
100	8.1
105	0
110	0
115	2.7

of ability to understand quickly the nature of the tasks. When the nature of the examination becomes a matter of public knowledge to candidates, the best features of it may be lost. It will be specially disastrous if some candidates do and others do not know about it.

Consequently, thorough precautions must be taken so that the examination will be as effective for twenty years as it is in its first trials. Our precautions are:

1. All test material is made up in quantities sufficient to provide for the entire series of examinations into which it enters, and is distributed without prejudice among them.² The selection for September 1929 will be nowise inferior to the selection for September 1919.

2. Three-fourths of the candidates' time is spent on paragraph reading, completing sentences, tests of information, arithmetical and algebraic computations and problem solving, common-sense questions, and following directions. The nature of these tasks is understood in a second or two by everybody of the intelligence level of a high-school graduate. In examining hundreds of papers, I have never seen a case of misunderstanding a single one of these tasks.

3. With two exceptions,³ fore-exercise is given for the other tasks, and two trials at each task are allowed. They include such tests as the disarranged sentences, opposites, number-series completions, verbal analogies, pictorial analogies, geometrical relations, and absurdities. The nature of these is, in general, such that any unfair advantage to a boy whom some unscrupulous tutor might train in test-taking would be slight. Experiments are planned to measure this gain exactly. If it is appreciable, the examination will be extended to include "traps" to penalize such improper preparation.

4. The content of the examination is entirely different in each issue. The form of the examination in the case of those tests where an unfair advantage from coaching is a possibility, will change every two or three years.

5. Finally, no new issue will be used regularly until it has been tried out sufficiently to insure its giving a prophecy substantially equivalent to the previous issues of the examination.

COMPARABILITY OF STANDARDS AND RESULTS

Since each element of the examination is selected at random as to difficulty, the total examination, comprising over twenty different tests, will in any one issue of the same general type, be closely equal in difficulty to any other (with the exception of the

² This is not true of the June 1919 series, which had to be arranged in part before the total material could be planned.

³ In the later series it is planned to provide fore-exercise for these, if any evidence appears that it is needed to equalize matters for those who have and those who have not previous acquaintance with the examination.

June, 1919 series which, for the reason given above, could not be fully standardized and is in part somewhat harder than the others). Each issue is further checked in this regard by actual trial. For example:

Eighty college students took Forms A and F, Part I, of the examination in that order. Eighty others took Forms A and G in that order. Letting a = score in A, f = score in F, and so on, we find that Median of $(f-a)$ = Median of $(g-a)+1$. That is, Form G is 1 point, or less than 1%, harder than Form F.

Thirty-six college students took Forms A and H in that order. Median of $(h-a)$ = Median of $(f-a)+1\frac{1}{2}$. That is, Form F is $1\frac{1}{2}$ points, or a little over 1%, harder than Form H.

One hundred fifty-four college students took Forms A and J in that order. Median of $(j-a)$ = Median of $(f-a)+2$. That is, Form F is 2 points, or about $1\frac{2}{3}$ % harder than Form J.

$$\text{Thus } f=g+1=h-1\frac{1}{2}=j-2$$

Part II Form B and Part II Form C were each given to 85 college students, the two groups of 85 being random halves of a group of 170. The two groups were tested at the same time in the same room, by the same examiner, all instructions being identical. The median score for Form B was 81.5; for Form C, 79.0 or a difference of 2.5. The 25-percentile score for Form B was 58.3; for Form C 57.5 or a difference of -0.8 . The 75-percentile score for Form B was 102.7; for Form C 101.9 or a difference of -0.8 . In general, the two distributions are closely similar.

We may, therefore, conclude that Part II C is probably a trifle harder than Part II B (about $1\frac{1}{2}$ percent).⁴

Thirty-three college students took both III B and III C, sixteen having one form first and seventeen having the other form first. The median of $(c-b)$ = $1\frac{1}{2}$ points or about $2\frac{1}{2}$ percent.

Two examinations made up respectively of:

I G	I J	II B	and III B
I F	I H	II C	and III C

would thus be almost exactly equal in difficulty for a college group.

Twenty-three adults took both II B and II D, twelve having Form B first and eleven having Form D first. The median of $(b-d)$ = 4. That is, Form D is 4 points harder than Form B.

⁴ Test 7 was omitted in both cases, but we know from other evidence that II B 7 and II C 7 are very closely similar in difficulty.

So, when II D is used in place of II B or II C, the other parts of the examination should be chosen to be a trifle easier than I G, I J, III B, or I F, I H, III C.

The difference in difficulty between any two issues can thus be known and allowed for by transmuting the score, or, if it is at all large, it can be reduced by alterations in the examination and re-checking. This process of equalization can be applied to the 10-percentile or 20-percentile or any other ability levels as well as to the general tendency of the group.

This checking is extremely laborious and probably unnecessary, since differences of 3 and 4 or even of 5 and 6 percent in the difficulty of different issues are of little practical consequence in any uses to which the examination is likely to be put. However, their reduction to small amounts and an exact knowledge of these amounts make an instrument available for very precise comparisons.

The scoring of all save the reading and completion tests is, with very slight exceptions, done by absolute keys with no requirement of judgment on the part of the scorer. The scoring of the reading and completion tests requires some judgment in the use of the keys, but the correlation between the ratings given by two scorers of the ability of a college instructor or assistant is very high. The correlation between two such scorers is 0.98. The probable difference between two such scorers was found to be only five points out of a total score which ranges from 7 to 96 among the actual candidates whose papers were scored. These scorers used only the stock keys supplied with the tests. By having one person score one test of the six, another score another, and so on, the probable error for the total result of one such group of six would obviously be unappreciable. This is the practice recommended where the examination is used with groups of any considerable size.

The administration of the examination consists simply of giving out the blanks, and instructing the candidates to go ahead at certain times even if they have not finished the work to date. In general, a candidate does test after test without awaiting instructions.

IMPROVEMENT OF THE EXAMINATION

As soon as adequate records of success in college are available, the correlation of each element of the examination therewith will

be computed together with the partial correlations where these are instructive. In the meantime, data are being gathered concerning the value of different available tests as shown by their correlations with a composite score obtained in seven or eight hours of work by adults of the general intelligence level of college candidates. Some of the data so far obtained are of value to anybody who is working with group tests for adult intelligence, and are presented here. Lack of space forbids adequate description of the tests. The following definitions apply to captions appearing in Tables II and III.

1. "Part I (first half)" means the first 30 minutes work of the standard examination as now given.
2. "Part II" means Part II of the standard examination.
3. "Part III" means Part III of the standard examination.
4. "New Part I (first half)" means a new form of Part I to go into effect in 1922.
5. "Information" means a test of the type used in the Army Alpha, but harder.
6. "Name part" means a test in writing a word that means a part of the thing named by the given word, and also begins with a certain letter.
7. "Name opposite" means a test in writing a word that means the opposite of the given word, and also begins with a certain letter.
8. "Name thing of which is part" means a test in writing a word meaning something of which the thing named by the given word is a part. The word written must also begin with a certain letter.
9. "Name thing made of," "synonyms" and "species of genus" are like the above. The word written has to fulfill the requirement of being in the stated relation to the given word and also of beginning with a certain letter.
10. "Sum, difference, product" means a test in giving two numbers which are defined by certain relations (e.g. two numbers whose product is 1 and whose sum is 2).

TABLE II. CORRELATIONS OF VARIOUS FORMS OF GROUP TESTS OF ADULT INTELLIGENCE LEVEL WITH A COMPOSITE SCORE FOR SEVEN HOURS OF TEST WORK, IN THE CASE OF 33 COLLEGE STUDENTS

Test	Time in Minutes	Correlation
Part I (first half).....	30	0.75
Part II.....	60	0.85
Part III.....	40	0.65
Part III.....	120	0.81
New Part I (first half).....	30	0.84
Information.....	15	0.62
Name part.....	8	0.40
Name opposite.....	12	0.80
Name thing of which is part.....	13	0.72
Name thing made of.....	13½	0.61
Synonyms.....	16	0.66
Species of genus.....	12½	0.61
Sum, difference, product.....	20	0.83
Total Time.....	390	

TABLE III. INTERCORRELATIONS OF VARIOUS FORMS OF GROUP TESTS FOR ADULT INTELLIGENCE LEVELS 33 COLLEGE STUDENTS

	Name Part	Name Opposite	Thing of Which Is Part	Thing Made of	Synonyms	Species of Genus
Information.....	0.03	0.56	0.53	0.54	0.48	0.52
Name part.....		0.25	0.27	0.30	0.30	0.33
Name opposite.....			0.61	0.46	0.83	0.49
Thing of which part.....				0.49	0.64	0.39
Thing made of.....					0.49	0.55
Synonyms.....						0.21

Facts like those shown in Tables II and III are being collected from individuals of the intelligence level of college students as fast as time and facilities permit, so that no test need ever be used in the examination until its value has been demonstrated.