VOLUME III FEBRUARY, 1912 NUMBER 2

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

A TENTATIVE REVISION AND EXTENSION OF THE BINET-SIMON MEASURING SCALE OF INTELLIGENCE.

LEWIS M. TERMAN,
Assistant Professor of Education, Stanford University,

AND

MR. H. G. CHILDS,
Assisted in the Tests by Katherin Romer Kip and Edith M. Bushnell.

PART I.[1]

INTRODUCTION.

In the light of the numerous recent studies of the progress of children through the grades it would appear that the time had arrived to determine more definitely than has yet been attempted the qualitative nature of the phenomena of retardation and acceleration. This determination should proceed along two lines. It should ascertain as far as possible the native intellectual endowment of each pupil, and it should undertake to establish with reference to acceptable standards his exact pedagogical status.

Though the latter task is by far the easier—is, in fact, quite feasible for most kinds of school work—we have barely set about its accomplishment. Except in handwriting and arithmetic, there are no generally acceptable standards of school

[1]This is the first of a series of articles. Part II will set forth the results of additional tests and Part III will give the author's revision and extension of the scale.

performance which enable us to make a reliable comparison of different pupils and schools. The average superintendent of schools accepts on faith the accomplishments of his own system and looks with suspicion on those of other schools, as is proved by the retardation which pupils usually suffer when they change from one school system to another. Such chaos cannot continue indefinitely now that the standarization has begun, and it does not seem unreasonable to expect that within a few years we shall have measures which will permit the expert determination of the efficiency of a school's efforts in most branches of the curriculum.

Measurements of intelligence are far more difficult to make, but even more important than those of pedagogical status. The degree of intelligence is a factor in every performance. Unfortunately, however, its influence cannot be isolated from that of training. Intelligence cannot work without ideas, and the quality and number of the latter are largely dependent upon accidents of environment. In our efforts to measure intelligence, therefore, we always measure intelligence plus a certain kind and amount of training. The only way to make sure that our measurement will arrive at any approximation to the former is to choose for our tests types of performance which will not be too greatly influenced by such differences in training and experience as ordinarily exist among the children of civilized people living under fairly uniform conditions of home life and educational advantages. While the social inheritance cannot, of course, be wholly identical for any two individuals, however similar their environment, there is nevertheless a limited social inheritance common to practically all who are possessed of a mentality sufficiently normal to make it their own. The environment of two children living in the same country under anything but exceptional conditions probably resembles a great deal more than it differs. The mere fact of a 12-year-old child's ignorance of the names of the month is not, *per se*, a reflection upon the child's intelligence. The child may be deaf, or he may have been reared by feeble-minded parents, or he may never have attended a school. But if the 12-year-old has had the usual home and school advantages

common to a civilized people such a deficiency as that noted becomes decidedly significant. Even then, however, we should not be justified in pronouncing the child subnormal upon this evidence alone. Gaps may occur in the education of anyone. It is only when the child's intelligence has been triangulated from many positions, and always to its disadvantage, that we can reasonably infer the presence of a subnormality marking the subject as incapable of acquiring the common elements of our social inheritance.

Surveys of this nature are possible only in so far as we have age norms for different lines of performance, few of which have yet been established. Many studies have been made showing age and grade tendencies for certain kinds of mental ability, but for the most part these have either been too limited in scope, or have not utilized non-selected subjects, or else have been made under conditions impossible to duplicate. Worst of all, the experimentation has been of haphazard nature instead of being guided by some directing idea like that of establishing definite age or grade standards. The Binet tests of 1908 are the only set hitherto devised covering any considerable variety of functions or directed by any comprehensive plan. The invaluable contribution of Binet consisted in the simple device of arranging the tests in a series of groups according to their difficulty as determined by age differences in performance. In spite of the limited data which served as the basis for his scale, the 1908 tests of intelligence immediately became serviceable and already seem to have proved their worth to all who have experimented with them. The series can readily, by extensive and thorough trial, be indefinitely extended, refined and adapted to different conditions and purposes. The following preliminary and imperfect study is offered as a slight contribution to this end.[2]

The purposes of this study are (1) to determine the adaptability of the Binet tests to American conditions; (2) to discover the changes in serial arrangement that may be necessary

[2]This study was completed before the publication of Binet's 1911 revision, and before the appearance of Goddard's and Bobertag's criticisms.—AUTHORS.

to make them reliable; (3) to try out certain other tests designed to supplement the Binet series, and (4) to arrive at some conclusion as to the psychological and pedagogical value of the tests individually and collectively.

Sources of Data and Conditions of the Study.

Between February 1 and May 1, 1911, the writers, with the assistance of Miss Katherine Kip and Miss Edith M. Bushnell, gave the Binet tests to 396 non-selected children.[3] The schools selected for the tests were a district school on the Stanford University campus, attended almost equally by the children of college professors and of laborers (30 children); the public schools of the village of Mayfield, Cal., where about one-half of the entire number were tested (72); the city schools of Palo Alto (150); a rural school near Stanford University (20), and two public kindergartens at Long Beach, Cal. (124). In the campus school, rural school and kindergartens practically all the pupils were tested; in Palo Alto the tests were given to the entire third and fourth grades of one building, to one-half the A and B fifth grades, and to several in each of the first, second, sixth and seventh grades. Where not all the pupils of a room were tested, as was the case in the Mayfield and some of the Palo Alto work, care was taken to secure representative case by requesting teachers to send only pupils whom they considered of average intelligence, or, in case a pupil of either exceptionally high or low ability was submitted, to follow this by another varying in the opposite direction from the normal. It is believed that the material is about as representative, with the exceptions later to be designated, as could be secured. As far as home advantages and social milieu are concerned the pupils are probably not very different, on the whole, from what would be found in the better class villages

[3]The tests were all made under the supervision of Dr. Terman, who himself applied the tests to about 90 children. Two hundred and twenty-two pupils were tested by Mr. Childs, and 124 by Miss Kip and Miss Bushnell. About 40 of the cases tested by Dr. Terman were a selected retarded group, the data from which could not be included in this study. The writers are greatly indebted to Miss Kip and Miss Bushnell for their very able and exceedingly careful assistance.

and towns from New England to California. The age and grade distribution was as follows:—

Age.	Number.	Grade Distribution.
4 years	29	27 (Kgn), 2 not in school.
5 years	83	81 (Kgn), 2 not in school.
6 years	26	17 (Kgn), 9 (I).
7 years	29	13 (I), 12 (II), 4 (III).
8 years	43	7 (I), 15 (II), 16 (III), 5 (IV).
9 years	49	2 (I), 2 (II), 16 (III), 26 (IV), 3 (V).
10 years	33	6 (III), 13 (IV), 11 (V), 3 (VI).
11 years	44	1 (II), 3 (III), 13 (IV), 18 (V), 7 (VI), 2 (VII)
12 years	35	6 (IV), 11 (V), 2 (VI), 7 (VII).
13 years	17	1 (IV), 5 (V), 4 (VI), 6 (VII), 1 (VIII).
14 years	6	3 (VI), 3 (VIII)
15 years[1]	2	1 (VI), 1 (VIII).

[1]Age here means age at last birthday, *i. e.*, 4 years includes children between 4 and 5, etc.

The tests were conducted in hallways or vacant rooms of the various school buildings, practically free from distractions. It was impossible to observe any uniformity in regard to time of day, the time ranging from 8.30 A. M. to 4.30 P. M. It is doubtful, however, whether this factor vitiates the results very materially, inasmuch as the tests almost never failed to enlist the interest of the child. The experiments of Thorndike and others indicate that undoubtedly existent fatigue does not materially influence a brief intellectual performance which holds the S's spontaneous attention. The time required for the tests was usually between 30 and 45 minutes, depending upon age, the number of tests covered and the promptness of responses.[4] Extreme care was taken to win the confidence of the child and to rid him of any embarrassment before beginning the tests. Occasionally several minutes were consumed in this way.

The kindergarten children took all the tests from the third to the eighth year, inclusive. The others, with a few exceptions, began the series a year below the S's chronological age, or further back if necessary, and continued as follows: The first and second grade pupils generally through the ninth-year

[4]The testing of seven to ten pupils constitutes a day's work of five hours if the examination is made reasonably deliberate and thorough. In the opinion of the authors, tests carried through at the rate of 20 to 30 per day are sure to give unreliable and misleading results.

group, and in many cases through the tenth year. In the grades above the second, testing usually began with the eighth-year group, but farther back if necessary, and proceeded as far as the pupil could go. In some of the earlier tests a few were not taken over as large a range as was later covered, and some of our data, therefore, may not be absolutely complete. Discarding some of the more questionable records has left the data here included, the writers believe, not seriously deficient from this cause.

Too much emphasis cannot be placed upon the importance of securing uniformity of conditions and procedure in applying the tests. An apparently insignificant change in the wording of a question or of a preliminary explanation may influence the response very appreciably. It is inevitable that, in spite of all precaution, a preliminary study of this kind will contain some errors arising from this source, considering that even variations in the vocal inflection of different E's will influence the result. Elaborate instructions, however, were placed in the hands of each person who proposed to assist in the tests, and these were carefully studied before the work began. Moreover, the methods of different workers were occasionally checked up by comparison during the progress of the tests.

The detailed instructions used are too bulky for reproduction here, but were practically identical with those in Whipple's Manual, except on the following points:

(1) The test of "making change" involved the purchase of a 9-cent article and payment with a 25-cent piece.

(2) For the test of ability to name pieces of money the following eight coins were used: Penny, nickel, dime, quarter, half-dollar, dollar, five dollars (gold), ten dollars (gold).

(3) *Palo Alto*, *river* and *money* were the three words used for sentence construction.

(4) The following tests of "comprehension" were used in year five:

(*a*) What's the thing to do when you feel sleepy?

(*b*) What's the thing to do when you feel cold?

(*c*) What's the thing to do if it's raining when you start to school? (Two out of three correct.)

The following were used in year ten:

(*a*) What ought you to do when you have missed a train?

(*b*) What ought you to do when you have been struck by a playmate who did not do it purposely?

(*c*) When you have broken something which does not belong to you?

(*d*) When you have been detained so that you will be late for school?

(*e*) Before taking part in an important affair?

(*f*) Why does one excuse a wrong act committed in anger more readily than a wrong act committed without anger?

(*g*) What ought you to do when asked your opinion about someone whom you only know a little?

(*h*) Why ought you to judge a person more by his acts than by his words?

(Five out of eight correct.)

(5) Numbers 1, 2, 4 and 5 of Whipple's absurdities (see Whipple's Manual, p. 509) were used, and in addition the following: "There was a railroad accident yesterday, but it was not serious. The number of dead is only 48."

(6) In the tests of memory for sentences and also for digits one

(7) In the rhyming test "day" was substituted for "obey." Year 12.)

(8) The following "problems of fact" were used in year 12:

1. "There was a little boy who had never been to the city. When he was six years old his father took him to San Francisco. As soon as the boy saw the electric street cars for the first time, he said . . ." "What do you think the boy said?"

2. "My neighbor has just been having queer visitors: first a doctor, then a lawyer, then a priest. What's happening there?"

3. "An Indian, coming to town for the first time, looked very intently at a white man riding up the street. When the white man had passed the Indian said to his companion: 'White man lazy; him walk sitting down.' What was the white man sitting on?"

(1 and 3 are substituted for the rather too gruesome problems of Binet.)

An S's test age was computed by assigning him to the lowest year in which he passed all the tests, or all but one, and then crediting him with a half year additional for each three tests satisfactorily passed beyond this. In the seventh year four tests were considered as the equivalent of a half year. From the ninth to the twelfth years, inclusive, five credits were counted as equivalent to a year, and a remainder of three as a half year. This seems to be the only way of giving the tests of any year approximately equal value with those of any other year if Binet's method is to be adhered to at all. Even by this method the child often loses credit for one or two tests passed beyond the half year to which he has been assigned, but on the

other hand he often gains one credit where he fails in but one test of a given year. Binet's method of assigning credit is crude at the best. The changes here introduced below the ninth year tend rather to draw the results nearer to the Binet norm than away from it. A much more accurate method will be suggested in the last article of this series. ,

Another difference between the method here employed in reckoning test age and that followed by Binet and others touches a more fundamental point and deserves special consideration. In deriving the age norms for the various tests it seems that Binet reckoned chronological age by even years: that is, by 8-year-olds, he means all children between 8 and 9 years; by 12-year-olds, all between 12 and 13, etc. In other words, the median age for the so-called 8-year group is about 8½ years. This is the most common procedure, is the one used also in our tests, and is perfectly legitimate, provided the facts are kept in mind. This means that what Binet designates as a characteristic performance of 8 years is really that of 8½ years. It follows, therefore, that when an S (whatever his chronological age) passes as far as to the end of the 8-year group, and no farther, or if his performance is equivalent to this, then such S should be assigned a test age of 8½ years. This was the method of computing used in this study, one which obviously throws our test ages one-half year in advance of those which would have been assigned by the method others have used. *This should be held in mind when comparisons are made of our results and those obtained by other studies.* The only logical way to avoid the necessity of adding on this half year is to group differently the S's to whom tests are applied for the sake of securing age norms. They can be grouped either by half years or in year groups separated by the half-year point. If we define the standard for 8-year mentality as the characteristic performance of non-selected S's lying between 7½ and 8½ years, then we can continue to use the Binet method of reckoning. Otherwise it is necessary to rename his standards as those for 6½, 7½, 8½, 9½ years, etc., instead of calling them, as Binet does, norms for 6, 7, 8, 9 years, etc. To illustrate the working of this, an S completing all the tests to the middle of the 8-year group would be ranked as testing 8

years. If he passes three additional tests either in this or the following years, he would be ranked as testing 8½ years.

Results from Tests of 396 Pupils.

Although, generally speaking, it is desirable to publish all the raw data from investigations of this nature, it is doubtful whether a preliminary excursion of this kind into a field where conditions and procedure have not been fully standardized warrants the expense and trouble of publishing the twenty or more pages which would be required for our material. We reserve this, together with a more thorough mathematical treatment of the data, for a later and more extended study.

The following table shows the distribution of *test ages* resulting from the application of the scale to the 396 subjects:

TABLE II.

Age.	No. tested.	Variations of test age from chronological age														Median age	Median test age
		+3.0	+2.5	+2.0	+1.5	+1.0	+0.5	0.0	−0.5	−1.0	−1.5	−2.0	−2.5	−3.0	−3.5		
4	29		8	13	4	3	1									4.75	6.5
5	83		3	14	31	24	5	5	1							5.5	7.0
6	26				2	9	7	7		1						6.37	7.0
7	29		1	1	6	6	4	8	2		1					7.5	8.0
8	43	2	1	3	2	6	6	7	6	7	3					8.5	8.5
9	49			5	6	5	6	13	8	3	3					9.5	9.5
10	33			1	3	5	5	7	5	3	4					10.5	10.5
11	44					1	3	10	7	11	5	5		2		11.46	10.5
12	35							3	2	10	7	7	3	2	1	12.33	11.0
13	17								1	2	1	8	1	3	1	13.42	11.5
14	6										1	3		2		14.58	12.5
15	2														2	15.2	12.0

Our data, so far as the Binet tests are concerned, are not at all conclusive for the thirteenth year, since we did not get the brightest pupils of this age for the reason that our tests were not extended into the eighth grade. The number of cases at fourteen and fifteen years is altogether too small to warrant any conclusions relative to these ages.

The following table shows a brief comparison of results by years and by grades:

TABLE III.

Age.	Median age.	Median test age.	School grade.	Median grade age.	Median grade test age.
7	7.5	8.0	I	7.5	7.5
8	8.5	8.5	II	8.17	8.3
9	9.5	9.5	III	9.33	9.5
10	10.5	10.5	IV	10.0	10.0
11	11.46	10.5	V	11.42	10.7
12	12.33	11.0	VI	12.17	11.5
*13	13.42	11.5	†VII	12.5	12.5

*Based on rather limited data. 17 cases.
†Based on rather limited data. 15 cases.

The following chart represents graphically the relation between the median chronolgical and the median test ages, and throws into relief the periods of rapid, normal and slow mental growths as measured by the Binet scale. The points in the upper horizontal line represent chronological and median ages, and in the lower line test ages and median test ages. The connecting lines join the corresponding medians. When there is an identity in the two medians the connecting line is vertical. If an age group tests ahead of the chronological age, the connecting line slants to the right as it goes down; if the group tests behind the age, to the left. It will be noted that the chronological ages cover a range of nine years; the test ages but five years. In other words, the scale is far too easy at the lower end, while at the upper end it is too difficult. An S at the upper range is also at a great disadvantage in that he is not given opportunity to try tests beyond thirteen years.

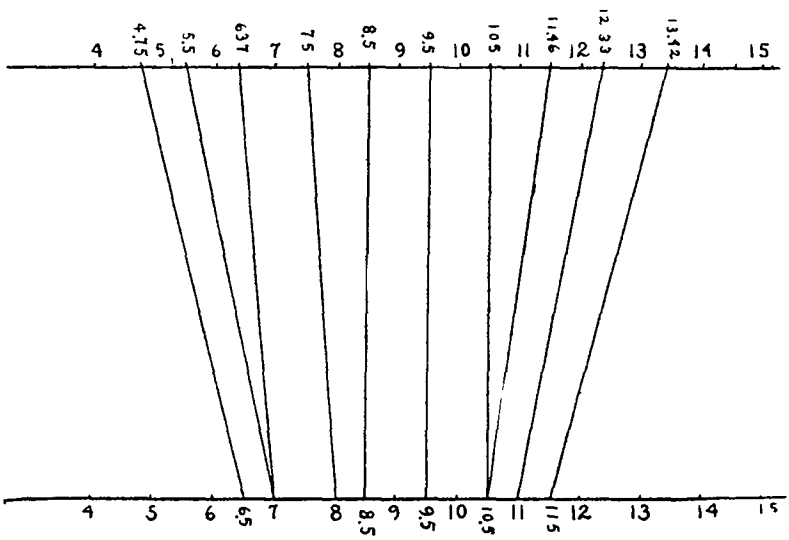

FIGURE I.—Upper line, chronological ages: lower line, test ages.

Charting the test ages as Dr. H. H. Goddard has done, by lumping all the ages together (see *The Training School* for

May, 1911), conceals, of course, the very facts we wish to know. From the above it is seen that the number of younger pupils testing ahead is about balanced by the number of older ones testing behind. What we want to know is how nearly accurate the scale is at every point. The following chart shows graphically how badly the scale fits California children of five and twelve years:

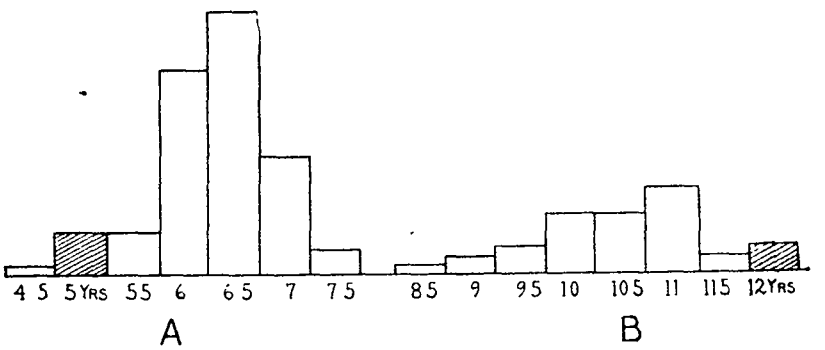

FIGURE II.—"A" shows the distribution of "test ages" for 83 five-year-old children; "B," for 35 twelve-year-olds. In each case the number testing normal is indicated by the barred column.

Table IV shows the number of children tested at each age and the per cent. of those tested at each age who pass each test taken. A comparison with Binet's scale will show how erroneously some of these tests are placed for use with American children. Our results indicate that for California children, at least, some of the tests will have to be moved as much as three years.

In the table where at any given year the number tested is but a small part of the whole number of that age the results should not be given serious consideration, for these represent either retarded older pupils who were given the lower year tests or the precocious younger children who were carried to a point far in advance of the majority of their age.

Further criticisms and suggestions are reserved for a later article.

TABLE IV.

This table shows the per cent. of those trying a test at any particular age who passed it. The figures in heavy type show the number of pupils of each age who tried each test.

Age.	IV. No. of cases.	1	2	3	4	5	V. No. of cases.	1	2	3	4	5	6	VI. No. of cases.	1	2	3	4	5	6	7
4	**29**	100	93	97	100	93	**29**	97	86	90	90	51	97	**29**	41	80	31	86	83	72	65
5	**83**	100	100	100	100	97	**83**	98	91	88	91	83	96	**83**	73	88	53	91	81	79	84
6	**18**	100	100	100	100	83	**20**	95	85	95	90	85	95	**24**	75	83	62	83	87	96	87
7	**0**	...	...	...	...	...	**4**	100	100	75	100	75	100	**20**	90	90	80	90	95	95	100
8	**0**	...	...	...	...	...	**2**	0	100	50	100	100	100	**14**	93	93	72	93	100	100	100
9	**0**	...	...	...	...	...	**0**	...	...	...	...	...	...	**3**	100	100	100	100	100	100	100
10	**0**	...	...	...	...	...	**0**	...	...	...	...	...	...	**0**	...	...	...	...	...	...	...
11	**0**	...	...	...	...	...	**0**	...	...	...	...	...	...	**0**	...	...	...	...	...	...	...
12	**0**	...	...	...	...	...	**0**	...	...	...	...	...	...	**0**	...	...	...	...	...	...	...
13	**0**	...	...	...	...	...	**0**	...	...	...	...	...	...	**0**	...	...	...	...	...	...	...
14	**0**	...	...	...	...	...	**0**	...	...	...	...	...	...	**0**	...	...	...	...	...	...	...
15	**0**	...	...	...	...	...	**0**	...	...	...	...	...	...	**0**	...	...	...	...	...	...	...

VIII. No. of cases.	1	2	3	4	5	6
29	0	0	65	0	0	34
83	0	3	74	1	0	47
26	0	11	84	7	7	47
29	20	58	96	62	79	55
43	51	60	100	69	83	65
45	89	73	100	95	100	89
20	90	85	100	100	100	100
14	86	72	100	93	100	100
6	100	83	100	100	100	100
0	...	...	...	...	...	...
0	...	...	...	...	...	...
0	...	...	...	...	...	...

VII. No. of cases.	1	2	3	4	5	6	7	8
29	34	14	0	23	17	44	34	17
83	68	33	10	53	50	47	55	38
26	57	73	42	61	50	42	77	53
25	68	92	100	92	72	60	100	72
35	82	97	100	91	74	60	97	88
12	83	100	100	92	83	83	100	92
1	100	100	100	100	0	100	100	100
2	100	100	100	100	100	100	100	100
0	...	...	...	...	...	...	...	...
0	...	...	...	...	...	...	...	...
0	...	...	...	...	...	...	...	...
0	...	...	...	...	...	...	...	...

TABLE IV.—Continued.

This table shows the per cent. of those trying a test at any particular age who passed it. The figures in heavy type show the number of pupils of each age who tried each test.

IX.

No. of cases.	1	2	3	4	5	6
0	...	...	..	...	..	...
0	...	...	. .	...	...	...
0	...	...	...	...	...	..
27	25	85	7	3	0	25
38	52	85	26	29	13	29
49	81	98	42	24	46	55
31	77	100	35	58	58	54
39	87	100	67	61	65	77
29	93	100	62	69	62	55
12	92	100	83	83	100	75
2	100	100	100	100	100	50
0	...	...	...	...	...	...

X

No. of cases.	1	2	3	4	5
0	..	.	...	...	..
0	.	...	..	..	...
0	.	..	. .	...	.
16	62	18	18	6	31
28	64	50	42	17	50
46	91	69	60	37	65
33	81	88	84	57	92
43	81	96	86	64	71
35	93	87	93	77	78
17	94	94	100	89	89
6	100	100	100	83	100
2	100	100	100	100	100

XI.

No. of cases.	1	2	3	4
0	...	...	...	...
0	...	...	...	...
0	...	...	...	...
4	25	50	0	0
20	50	35	10	10
42	40	57	2	19
33	52	67	18	24
43	55	83	16	32
34	73	82	26	62
17	88	94	36	66
6	100	67	33	83
2	50	100	100	50

XII

No. of cases.	1	2	3	4
0	..	..	.	..
0	...	...	...	...
0	...	...	. .	..
3	0	100	0	0
14	28	85	42	7
27	29	74	29	18
28	46	92	35	35
37	43	81	37	40
34	38	82	41	41
17	36	89	54	48
6	83	100	67	100
2	50	50	100	100

XIII.

No. of cases.	1	2	3
0	...	...	...
0	...	..	...
0	.	..	.
0	..	..	...
0	..	...	...
0	...	..	...
0	...	...	..
8	25	. .	25
19	15	..	30
9	22	...	55
5	40	..	60
2	0	...	50

EXPLANATION OF TABLE IV.

The Roman headings in the above table designate year group; the Arabic designate the individual tests in the following order:

IV—1. Sex. 2. Names familiar objects. 3. Three digits. 4. Comparison of two lines. 5. Sentences of ten syllables.

V—1. Weights. 2. Copies square. 3. Divided rectangle. 4. Counts four. 5. Fourteen syllables. 6. Comprehension questions.

VI—1. Right hand, left ear. 2. Aesthetic comparison. 3. Sixteen syllables. 4. Defintions of familiar objects. 5. Three commands. 6. Knows age. 7. Morning and afternoon.

VII—1. Omissions from pictures. 2. Number of fingers. 3. Writing from copy. 4. Diamond. 5. Five digits. 6. Description of pictures. 7. Thirteen pennies. 8. Four coins.

VIII—1. Reading for two memories. 2. Value of stamps. 3. Four colors. 4. Counting 20 to 0. 5. Writing from dictation. 6. Comparison of things.

IX—1. Names date. 2. Days of week. 3. Makes change. 4. Definitions superior to use. 5. Six memories. 6. Weights.

X—1. Names months. 2. Nine pieces of money. 3. Three words in sentence. 4. Comprehension questions. 5. Six digits.

XI—1. Absurd statements. 2. Sixty words in two minutes. 3. Defines abstract terms. 4. Disarranged sentences.

XII—1. Seven digits. 2. Rhymes. 3. Twenty-six syllables. 4. Problems of fact.

XIII—1. Draws from design in folded paper. 2. Reversed triangle. 3. Distinctions between words.