THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

MEASUREMENT OF THE DRAWING ABILITY OF TWO THOUSAND ONE HUNDRED AND SEVENTY-SEVEN CHILDREN IN INDIANA CITY SCHOOL SYS- TEMS BY A SUPPLEMENTED THORNDIKE SCALE

H. G. CHILDS

Indiana University

During the winter of 1914, the writer, in coöperation with his class in School Administration, decided to use in a limited way the newly constructed Thorndike Scale to measure achievement in drawing for the purpose:

1. Of determining growth in ability from grade to grade.
2. Of determining standards or norms of ability for each grade.
3. Of determining limitations of the scale in actual use.
4. Of determining what administrative problems would be brought to light which might be desirable for supervising authorities to know.

METHOD OF DERIVING THE SUPPLEMENTED SCALE

A preliminary consideration of the scale and its application to a limited number of drawings convinced us that for our purposes it would be desirable to supplement the scale so that there might be a greater degree of uniformity in the composition represented by the different units of the scale. As several units

on the Thorndike Scale and many of the supplementary drawings represented snow scenes with action, it occurred to us to construct a scale every unit of which should represent a snow scene with human figures in action, placing houses, landscapes, etc., as accessory features.

For the purpose of supplementing the scale, drawings were chosen from the Thorndike supplementary sheet, numbered from one hundred one to one hundred seventeen, eleven of which represent snow scenes with human figures in action.

Seventeen judges were secured, consisting of teachers, graduate students, and seniors in Indiana University, to rank these supplementary drawings.

Tables I and II show the results of the comparisons.

TABLE I.
Rank Assigned Each Sample.

Judge.....	1	2	3	4	5	6	7	8	9	10	11
A ..	116	106	107	110	111	108	104	103	101	105	102
B ..	116	106	107	110	111	108	104	101	103	105	102
C ..	110	116	107	106	111	108	101	103	104	105	102
D ..	110	107	116	106	108	111	104	101	103	105	102
E ..	116	107	106	110	111	108	105	101	103	104	102
F ..	116	107	110	106	111	108	104	103	101	105	102
G ..	116	107	110	106	111	108	104	103	101	105	102
H ..	116	110	107	106	111	108	104	103	102	105	101
I ..	116	106	107	110	111	108	104	103	101	105	102
J ..	116	110	107	106	111	108	104	101	103	105	102
K ..	110	107	116	108	111	106	104	103	101	105	102
L ..	116	107	110	108	111	106	104	105	103	101	102
M ..	110	111	116	106	107	108	104	103	101	105	102
N ..	116	107	110	111	108	106	105	103	101	104	102
O ..	116	107	110	106	111	108	104	101	103	105	102
P ..	110	106	116	107	108	111	104	101	105	103	102
Q ..	116	107	106	110	111	108	104	101	103	105	102

1, 2, 3, etc.-11 represent the relative values from the poorest to the best

From a comparison of the rankings of Table I, the percentage relationships expressed in Table II were derived.

TABLE II.

88%	of the judges	rated	107	as better than	116
71%	"	"	106	"	107
76%	"	"	111	"	106
76%	"	"	108	"	111
100%	"	"	104	"	108
82%	"	"	101	"	104
82%	"	"	105	"	101
94%	"	"	102	"	105

Samples 110 and 103 were omitted because they were given almost equal value with 107 and 101 respectively.

Considering the difference detected between two samples by seventy-five per cent. of the judges as unity, the samples in Table II have the following relative values:

TABLE III.

Sample 107	is	1.7	units	better	than	sample 116
" 106	"	.8	"	"	"	107
" 111	"	1.0	"	"	"	106
" 108	"	1.0	"	"	"	111
" 104	"	3.5	"	"	"	108
" 101	"	1.4	"	"	"	104
" 105	"	1.4	"	"	"	101
" 102	"	2.3	"	"	"	105

Having secured this table of relative values for the supplementary samples used, the next step in perfecting the supplemented Thorndike Scale was to fix, by comparison, the value of one, or better yet, of several of these samples on the Thorndike Scale. For this purpose ten of the seventeen judges who made the first set of comparisons were secured. Table IV indicates the results:

TABLE IV

73%	of the judges	ranked sample	2	T	6.5	as	better	than	sample	116
90%	"	"	"	"	111.	"	"	"	"	T 8.6
90%	"	"	"	"	T 10	5	"	"	"	111
63%	"	"	"	"	108	"	"	"	"	T 10.5
80%	"	"	"	"	T 13.5	"	"	"	"	104
80%	"	"	"	"	101	"	"	"	"	T 13.5
70%	"	"	"	"	T 14	4	"	"	"	101
80%	"	"	"	"	T 16	"	"	"	"	105
70%	"	"	"	"	102	"	"	"	"	T 17

Reducing these per cents. to differences expressed in terms of the unit difference, or that detected by seventy-five per cent. of the judges, we have the results indicated in Table V.

From Tables III and V the final scale as given in Table VI was derived. In many cases a combination of differences indicated in Tables III and V was used in fixing the value of any given sample on the supplementary scale.

¹ E. L. THORNDIKE. *The Measurement of Achievement in Drawing*. Teachers College Record, November, 1913, p. 25, Table 4.

² T 6.5 means the sample on the Thorndike Scale having a value of 6.5.

³ Odd per cents. are due to some judges giving the samples equal rating.

TABLE V.

Sample T	6.5	is	.9	units	better	than	sample	116
"	111	"	1.9	"	"	"	"	T 8.6
"	T 10.5	"	1.9	"	"	"	"	111
"	108	"	.5	"	"	"	"	T 10.5
"	T 13.5	"	1.2	"	"	"	"	104
"	101	"	1.2	"	"	"	"	T 13.5
"	T 14.4	"	.8	"	"	"	"	101
"	T 16	"	1.2	"	"	"	"	105
"	102	"	.8	"	"	"	"	T 17

The reader should bear in mind that the number of judgments upon which the foregoing tables are based is insufficient for accurate scientific work, but for obtaining approximate values for supplementary points on a scale already constructed the writer believes the results obtained are reasonably adequate.

TABLE VI.

The Supplemented Scale as Used.

Sample	Value	Sample	Value	Sample	Value
T 0.	0 0	106	8 0	T 13 5	13 5
T 2.4	2.4	T 8 6	8.6	101	14 0
T 3.9	3 9	111	9 3	T 14.4	14.4
116	5.5	T10.5	10.5	105	15 2
T 5 7	5.7	108	11 0	T 16.	1 .0
T 6.5	6 5	T11 8	11 8	T 17	17 0
107	7 2	T12 6	12 6	102	17 6
T 7 8	7.8	104	13.0		

The objection may be raised that this scale has too many fixed points. Those who used the supplemented scale report that a variety of samples of approximately equal value aids greatly in rating drawings. Both the Thorndike and Ayers writing scales make use of this principle by providing three samples under each quality on the scale.

There are still lacking samples of the general scale in the units 1 to 1.99, 2 to 2.99, 3 to 3.99, 4 to 4.99, and 12 to 12.99.

A FURTHER REVISION OF THE SCALE.

Because of this lack of samples in the lower range of the scale and because of some criticisms made relative to the values assigned to samples 107 and 108, the following supplementary comparisons were made subsequent to the completion of the tests in systems A and B. Eleven judges compared samples T10.5, 108, 104, T13.5; ten judges compared samples 107 and T8.6; and twelve judges compared samples A, B, C, with 116.

Results of the Supplementary Comparisons

(1).	82%	of the judges rated	108	as better than T	10.5
	64%	" " " "	104	" " "	108
	55%	" " " "	T 13.5	" " "	104

Combining these judgments with those of Tables II and IV we have the following:

73%	of the judges rated	108	as better than T	10.5 or .9 better
86%	" " " "	104	" " "	108 " 1.6 "
67%	" " " "	T 13.5	" " "	104 " .7 "

These combined comparisons still give 104 approximately its former relative position with respect to T13.5 and T10.5, its value being reduced to 12.9, but the position of 108 is changed from .5 to .9 better than T10.5 and is thus given a value of 11.4.

What effect would these new values of 104 and 108 have upon the results from systems A and B? Whether the exact or unit group averages be taken, the surface of distribution more nearly approaches the normal probability surface in the ranges of the scale below 6 and between 11 and 14.

(2). 60% of the judges rated T8.6 as better than 107 or .4 better.

This gives 107 a value of 8.2. Combining this with the value 7.2 assigned as the result of comparisons between 116, 107, and T6.5, Tables III and V, a resultant of approximately 7.6 represents the new value for 107. If we substitute this value for sample 107 and reject sample 106 (value 8) from our scale, the results found in systems A and B will not be appreciably changed.

(3). Samples A, B, and C, together with several others, were chosen by the writer from the poorest drawings in system A for comparison with sample 116, the lowest on our supplemented scale.

75%	of the judges rated	116	as better than C or	1.0 better
92%	" " " "	116	" " "	B " 2.0 "
100%	" " " "	116	" " "	A " 3.5 "

As a result of this comparison samples A, B, and C, given below, are recommended for use with values of 2.0, 3.5, and 4.5 respectively.

The following scale is suggested for rating drawings of snow scenes with human figures in action. With one or two exceptions this is a scale of approximately unit differences, and includes all the human figure samples of the original Thorndike scale.

Sample 2

SAMPLE C.—QUALITY 45

TABLE VIa

Sample	Value	Sample	Value	Sample	Value
T O	0.0	T 6.5	6.5	104	12.9
A	2.0	107	7.6	T 13.5	13.5
B	3.5	T 8.6	8.6	T 14.4	14.4
C	4.5	111	9.3	T 16	16.0
116	5.5	T 10.5	10.5	102	17.6
		108	11.4		

METHOD OF CONDUCTING AND RATING THE TEST.

The drawing exercise was given to one thousand four hundred and five pupils in the elementary schools of system A on February 16 and 17, 1914, and to two hundred eighty-five pupils in the high school, May 26, 1914. As the high school seniors were excused from attendance at this time, this study shows very limited data from them. The test was given to four hundred eighty-seven pupils in one school of system B, April 11, 1914. In system A the exercise was conducted directly by my students in some cases and more frequently by the teacher regularly in charge of the class. In every case the following directions were observed.

In the school of system B the same directions were given under the direction of the drawing supervisor and the teacher.

Directions

1. Materials:
White drawing paper, 6 x 9, drawing pencil, charcoal or crayon-black only.
2. Preliminary data on the back of the sheet.
City, school, teacher, pupil, grade, age (years only), date.
3. Ask pupils to do the best they can and not to hurry.
4. Subject:
Scene or picture with snow on the ground and boys or girls doing something as snowballing, coasting, etc.
5. Be sure pupils understand just what they are to do and the time they are to have for doing it.
6. Time:
Ten minutes, exclusive of all directions and recording of preliminary data.
7. Collect papers; tie papers of each grade in a package and label with subject, city, school, teacher, and grade.
8. No assistance shall be given any pupil aside from interpreting the instructions and helping to fill in the data on the back of the sheet if necessary.
9. Distribute materials before giving the directions to the pupils.
10. Teachers are not to know in advance the nature of the test to be given.

All papers were rated by my pupils in administration, five in number, the papers being so assigned that the final rating of any grade represented the composite judgment of at least three of the judges and generally four. Each paper was rated by two judges and the ratings averaged.

My instructions to the judges were to determine how well the pupil had done what he was directed to do: viz., to draw a snow scene with human figures in action. The suggestion was made that emphasis should be placed as follows: human figure and action, grouping, landscapes, houses, trees, etc. The scorers were asked to judge largely on the representative features of the drawing.

CONDITIONS UNDER WHICH DRAWING IS GIVEN IN THE SCHOOLS TESTED.

During the year 1913-1914 the schools of system A had no drawing supervisor, the work being carried on by each teacher from outlines previously used. It should be noted that previously to this year the drawing had been supervised. The time given weekly varied from thirty to seventy-five minutes, but fifty and sixty minutes per week seemed to be the length of periods most frequently found on the programs.

In the departmental grades (6A to 8A) and in the high school no time was given to drawing as such.

In system B about ninety minutes per week were devoted to drawing in grades 3A to 8A and seventy-five minutes per week in grades 1B to 3B.

The work was well supervised.

THE RESULTS.

The results can be most adequately represented by the tables of distribution, the curves of progress, the central tendencies, and variabilities by grades.

In tabulating results, the exact averages of the two judges on any given paper were not taken, but merely the unit group within which the average fell, e. g., a paper rated 7.2 and 8.0 was assigned to the 7 to 7.99 group.

TABLE VII
Distribution of Drawing Ability—System A, 1690 pupils

	Number of Pupils by Grades																	Per cent. of Pup- ils Receiving each Score	High School					Per cent. Receiv- ing each Score	
Score	1B	1A	2B	2A	3B	3A	4B	4A	5B	5A	6B	6A	7B	7A	8B	8A	Total		9	10	11	12	Total		
0.....	1																1	.07	1				1	.35	
1 to 1.99.....	1		1			1											3	.21	1				1	.35	
2 to 2.99.....	2		4			1											7	.50	0				0	.00	
3 to 3.99.....	13	16	18			3			1			1					57	4.05	3				3	1.05	
4 to 4.99.....	17	7	7	3	1	2			2		1	0			1		42	3.00	1		1		2	.70	
5 to 5.99.....	11	16	14	3	5	10	1	0	6	3	3	1		0			73	5.20	5	3	2		10	3.51	
6 to 6.99.....	27	34	18	23	14	18	15	6	15	13	6	3		2		1	198	14.09	6	8	3		17	5.96	
7 to 7.99.....	8	22	7	30	25	39	33	10	12	18	12	6		7	2	5	7	243	17.30	18	14	11		43	15.09
8 to 8.99.....	6	19	12	33	39	40	23	57	23	48	42	17	39	24	23	17	462	32.88	20	32	27	4	83	29.12	
9 to 9.99.....	1	1		2	7	7	7	13	10	12	9	13	13	31	9	6	141	10.03	19	18	20	0	57	20.00	
10 to 10.99....				1	3	3		5	8	9	17	31	7	34	8	13	139	9.89	11	12	15	2	40	14.00	
11 to 11.99....								2	0	0		5	1	1	1	5	15	1.07	5	4	6	1	16	5.61	
12 to 12.99....								1	0	0		3	1	3	0	0	8	.57	2	2	2		6	2.11	
13 to 13.99....									2	1		3			4	1	11	.78	3	2	1		6	2.11	
14 to 14.99....												2			3		5	.36							
15 to 15.99....																									
Total	87	115	81	100	94	124	79	95	79	104	90	85	71	98	53	50	1405	100.00	95	95	88	7	285	100.00	
Median....	5.91	6.56	5.78	7.50	8.06	7.69	7.73	8.54	8.17	8.38	8.56	10.06	8.67	9.66	8.96	9.00	8.17		8.65	8.72	9.00	8.75	8.80		
P. E..	1.17	1.11	1.54	.86	.76	.85	.66	.18	1.13	.68	.63	1.07	.55	.72	1.01	1.12	1.04		1.25	.95	.97	1.09	1.04		

TABLE VIII
Distribution of Drawing Ability—One School of System B. 487 Pupils

Score	Number of Pupils by Grades																Total	Per cent. of Pupils Receiving Each Score
	1B	1A	2B	2A	3B	3A	4B	4A	5B	5A	6B	6A	7B	7A	8B	8A	1-8	
0...							1										1	.21
1—1.99...							0	1									1	.21
2—2.99...		1	1				1	0					1				4	.82
3—3.99...		12	2			3	3	0					0				18	3.69
4—4.99...		9	4		1	0	0	0					0	1			15	3.08
5—5.99...		15	8		2	3	0	3	1		1	4	0	2			39	8.01
6—6.99...		2	14	9	5	11	9	5	2	4	0	3	1	4	1		70	14.37
7—7.99...		1	7	8	11	11	5	7	9	11	13	4	0	8	7	2	104	21.35
8—8.99...			2	2	7	5	8	7	15	14	18	10	19	16	13	3	139	28.54
9—9.99...				3	3	3		4	3	5	4	10	8	3	5	6	57	11.70
10—10.99...						1			1	3	1	2	2	2	1	7	20	4.11
11—11.99...										1		2		2	1	3	9	1.85
12—12.99...												1		1	0	2	4	.82
13—13.99...												1		1	1	2	5	1.03
14—14.99...																1	1	.21
Total	0	40	38	22	29	35	27	27	31	38	37	37	31	40	29	26	487	100 00
Median ...		4 83	6 32	7.31	7 64	7 27	7.00	7 71	8.27	8 32	8 28	8.80	8.74	8.34	8.54	10 36	7 92	
P. E.		89	82	66	77	76	98	1 03	61	.73	56	1 05	.49	.78	63	1.12	1.09	

Below is given a summary of the combined results of Tables VII and VIII.

TABLE IX

Summary of Tables VII and VIII.

Grade . . .	1B	1A	2B	2A	3B	3A	4B	4A	5B
Median Score . .	5 5(?)	6 1	5 95	7.46	7 96	7 60	7 54	8 35	8 20
Grade . . .	5A	6B	6A	7B	7A	8B	8A	All	
Median Score . .	8 36	8 48	9 68	8 69	9 28	8 81	9 47	8 09	

From this it would seem that no distinctive norm can be established for each grade, but rather by groups as indicated below.

TABLE X.

Group	Grades	Characteristic Scores
I.	1B	4 5 (?), 5.5
II.	1A, 2B	5.5, 6 5
III.	2A, 3B, 3A, 4B	7 2, 8 0
IV.	4A, 5B, 5A, 6B	8.0, 8.6
V.	6A, 7B, 7A, 8B, 8A	8 6, 9.3

8 to 8.99 is the modal score (the one most often made) by pupils from the 4A to 8A grades inclusive.

No doubt a few periods of practice upon the type of drawing represented by this scale would yield higher scores in all grades if tested by the scale.

ADVANTAGES AND LIMITATIONS OF THE SCALE IN ACTUAL USE.

There is no longer any serious question as to the advantages of objective scales for measuring school achievement. Repeated trials have demonstrated that such scales limit the subjective factor in judgment and narrow greatly the range of variability in rating achievement.

It seems to the writer also that it is desirable that the teacher or supervisor of drawing know what quality of achievement the pupils of any grade should attain, e. g., that pupils of the 1B grade should draw snow scenes with action equal in quality to sample 116, and in grade 3B equal to sample 106, etc.

On page twenty-nine of *The Teachers College Record* for November, 1913, Professor Thorndike indicates the most obvious limitation of the scale, viz., that it does not admit of easy and accurate comparison of all types of drawings. This can be remedied only by devising a series of scales covering the various types of drawing, such as human figures, animals, landscapes, flowers, designs, etc.

Table XII. Distribution of drawing ability. - - - -System A
Based on tables VII and VIII _____ System B
All grades 1 to 8.

Teachers and supervisors in the schools where these tests were given protested that it was unfair to judge the quality of drawing achievement by a scale based upon a type of drawing which was given but little emphasis in their courses of study. Landscapes and object drawing receive much attention in most systems, but the landscape effects in the Thorndike scale below quality 14.4 are so crude as to afford but little basis for comparison. This feature in the drawings of systems A and B was very pronounced. Landscape effects, even with lower grade children, often compared favorably with the best on the scale, but the figures were inferior.

Many pupils would, no doubt, have made a higher score if the time had not been limited. This objection applies to the conditions of administering the test and not to the scale itself. Here, as in any other field of human achievement, efficiency must be measured both in terms of product and time. Just where to strike the balance must be determined by experimentation.

Another limitation of the scale already referred to is that the scale has no snow scenes with action of quality, 2 to 2.99, 3 to 3.99, or 12 to 12.99, and no samples of any sort, 1 to 1.99 or 4 to 4.99. As a result there are fewer ratings of 4 to 4.99 than there are of 3 to 3.99, whereas a normal distribution of ability should give more. In like manner there are more ratings of 13 to 13.99 than of 12 to 12.99, which should not be with a normal distribution of ability. For verification of the above facts consult Table VII. This shortcoming of the scale may be improved by the use of the scale suggested in Table VIa.

SUGGESTIONS PERTAINING TO THE USE OF THE SCALE AND THE INTERPRETATION OF RESULTS.

Those interested in using the scale will find it helpful to sort the papers to be rated into piles such that the papers of any one pile shall be of approximately equal value and perceptibly different from those of the piles of nearest value. This provides the reader with a double set of comparisons; first, of drawings with each other; and second, of each with the standard scale. Then, too, one can work more rapidly, having to use but a limited range of the whole scale in rating any one pile or sample. If teachers are encouraged to use the scale to measure the prog-

ress of their pupils from time to time, more definite standards of achievement and progress will result. Supervisors will find it advisable to do some preliminary work with teachers in the use of the scale. This can best be done with a few selected drawings by calling the teachers' attention to the features to be emphasized in the scale and the drawings. Practice will naturally lead to facility in the use of the scale.

Smoothness of the curve and a marked rise at its end, together with a greater range of gain during the eight years and a smaller range in variation of the product of each grade, are indicative of good supervision. These features are clearly apparent in the curves and table of distribution of system B. The curve of system A is characterized by more ups and downs and a decline from the 6A level of achievement to the end. The range of progress from the 1A to the 8A grades is 2.44 units as compared with 5.53 units in system B, and from the 1A to the 6A grades, 3.50 units as compared with 3.97 units in system B. Systematic supervision, no doubt, accounts for much of the difference found in the two curves of progress. It might be noted, however, that the actual achievement in system A is slightly above that in system B in nearly every grade from 1A to 8B.

Our tables of distribution and curves show that the average child develops more ability in drawing before entering school at the age of six or seven years than he does in the entire eight years of his elementary school course. This statement is based upon the data of Table VII and the fact that 1B drawings were made within four weeks of the beginning of the term.

In both curves the long plateau subsequent to the level reached in grade three is clearly noticeable. This is completely in accord with the studies of children's drawings by Barnes, Lukens, Burk, and Götze, who report a plateau of non-development from the ages of nine or ten on to adolescence.

In the light of these studies and the development of the same fact by the use of the objective scale, it seems pertinent to inquire whether our administrative practice is sound in assigning the same time to drawing in grades four and five as in earlier or later grades where natural development seems to be more marked. In a system having three or four sections each in the fourth and fifth grades, why not assign one section thirty minutes weekly for drawing, another sixty, another ninety, another one hundred

twenty? By adequate testing at the beginning and end of a half year or a year, the progress of each group may be determined.

If thirty minutes weekly produces as good results as ninety, why not assign drawing a minimum of time during these years and appropriate the extra time, now generally given to this subject, to other subjects in which experience has shown that pupils do make rapid progress with the added time? To quote the late Professor James: "Strike while the iron is hot."

Present achievement should be determined by the supervisor for the purpose of planning future improvement. Tables of distribution and curves of progress should stimulate the supervisor to prevent relapses and to eliminate plateaus, to keep the curve ever slanting upward.

The limited findings of this study should not be considered as establishing norms of ability in drawing because they represent only one phase of drawing achievement in a very limited number of schools. These findings are not intended to be final or conclusive in any respect, but it is hoped that these results may be suggestive and helpful to other teachers as to methods of experimentation and as a basis for comparison of achievement.