A SCALE IN ANCIENT HISTORY¹

LEROY W. SACKETT

University of Texas

The material used in this study of the range of information in ancient history and the scaling of difficulty of ancient history questions was taken from Myers' General History and the ancient period was defined as closing in the year 800 A. D. The basis for selecting the questions was to include only questions which every pupil who takes a course in ancient history must encounter. Topics which were treated with equal emphasis could not be selected as that depends upon the pupil's interest and the teacher's method, but every student being tested was expected to feel at every point a sense of familiarity in the presence of every question. Even where there could be no adequate recall they should feel that they had known it at some time but had forgotten. To do this the test material was submitted to a number of teachers of ancient history and many valuable suggestions were obtained. Nothing was included which did not meet with general approval from these teachers. It was tried in a preliminary way among University Freshmen and Sophomores coming from a wide range of schools and from classes just completing ancient history in the high school. These students were asked to point out any features which they had never encountered in a study of ancient history. After inspecting their answers to the questions and their comments upon them, the following test material was adopted:

Test No. I. For what are the following men noted? (1) Hannibal. (2) Khufu or Cheops. (3) Demosthenes. (4) Darius. (5) Solon. (6) Charlemagne. (7) Attila. (8) Constantine. (9) Mithradates. (10) Justinian.

Test No. II. Name a man noted in ancient history for each of the following: (1) Orator. (2) Painter. (3) Sculptor. (4) Historian. (5) Law-giver. (6) Philosopher. (7) General. (8) Ruler (King, Emperor, etc.). (9) Builder. (10) Poet.

¹ Special mention should be made of the painstaking assistance of several of my advanced students: Misses Luella Tankersley, Mary Williams, Kathleen Molesworth, and Mr. W. P. Webb. No statement here can give just credit to my colleague, Dr. Truman L. Kelley, for his assistance in the mathematical aspects of the problem. The test material was originally arranged by Miss Hazel Marsh.

Test No. III. Give the historical significance of each of the following: (1) Battle of Tours. (2) Age of Augustus. (3) Battle of Milvian Bridge. (4) The church council of Nicaea. (5) Check of the Saracens before Constantinople. (6) Reign of Alexander the Great. (7) Age of Pericles. (8) Founding of the Hebrew Monarchy. (9) Burning of Carthage. (10) Peloponnesian War.

Test No. IV. Between whom were the following battles fought: (1) Arbela. (2) Marathon. (3) Metaurus. (4) Teutoburg Forest. (5) Chalons. Name the victor in each case.

Test No. V. Give the approximate date of each of the following: (1) Fall of Rome. (2) Battle of Marathon. (3) Crowning of Charlemagne. (4) Establishment of the Saracen Kingdom in Spain. (5) Delian League. (6) The Hegira. (7) Defeat of Saracens by the Germans. (8) Battle of Actium. (9) Defeat of Persians by Alexander. (10) Establishment of the Roman Empire.

Test No. VI. What do you consider the most important contribution to civilization from these peoples? (1) Greeks. (2) Teutons. (3) Phoenicians. (4) Saracens and Arabians. (5) Romans. (6) Hebrews. (7) Persians. (8) Egyptians. (9) Babylonians. (10) Prehistoric Man.

Test No. VII. Mark each of the following peoples as being Hamitic, Semitic, or Aryan: (1) Greeks. (2) Egyptians. (3) Romans. (4) Hebrews. (5) Hindus. (6) Babylonians. (7) Teutons. (8) Assyrians. (9) Phoenicians. (10) Persians.

Test No. VIII. Name and mark the geographical locations on the accompanying map of ten points that you think were most important in ancient history. (Here was inserted an outline map of the Mediterranean World.)

The tests were arranged in a packet with blank spaces for answers and with the brief instructions at the top of each list of ten questions. Four minutes were allowed to each of the eight tests thus bringing the whole series with the preliminary information and the supplementary comments well within the 45-minute period. This time, short as it may seem, was decided upon after repeated trials during which the actual time consumed was observed. Even in regularly giving the tests there seemed to be little disposition to call for more time, or to give evidence of being hurried. Teachers who coöperated in giving the tests

and who anticipated that the time allowed would be inadequate were satisfied when they saw them given. Pupils were asked to write on an attached blank page their comments, criticisms, and complaints. Few complained of the time allowed and most reported that it had given them a fair opportunity to reveal what they knew, or rather that it had revealed to them how little they really remembered.

The nature of the questions combined with the manner in which they were presented test only one particular phase of historical information; viz., the ability to recall definite facts promptly. One of the chief marks of culture along historical lines, however, is to be able, when a lecturer mentions Greek philosophy, to think promptly of one or more exponents; when an ancient nation is mentioned, to be aware at once of some of the chief attributes and contributions of its people; when a battle which represents a turning point in civilization is mentioned, to know the contending forces and the results of the struggle. Thus the tests and the resulting scale measure only a limited and specific, though not an unimportant, kind of historical ability.

On the matter of grading pupils' answers, certain difficulties arose which are not present in arithmetic and spelling tests. The pupils' reactions could not be graded on the absolute standard of right or wrong. In nearly every attempt there were varying degrees of rightness or wrongness. The answer might be a real error or a wide variation from an otherwise correct reaction. To obviate this difficulty as much as possible all grading should have been done by the same person, so that the same standards of grading would be back of the evaluation of the questions. It was found to be impracticable to ask one person to make eighty judgments each on 933 pupils. Two directors of the grading were chosen and they worked over the correct and adequate answers which would be counted as 10 per cent, those that would be considered 5 per cent., and those for certain intermediate points. Sometimes this varied far enough from the method of right and wrong answers to be almost an arbitrary, but it was thought a more equitable, assignment of values. These two persons graded many of the papers and personally supervised and coached the assistants who also took part in the grading.

The following will serve as a sort of rough score card for grading. The ideas represented in the answers are given a weighting

which indicates the value placed upon that item in the answer when that idea is present.

TEST No. I.

1. Hannibal—A Carthaginian general (3) who crossed the Alps (3) and marched against the Romans (4).
2. Khufu or Cheops—An Egyptian ruler (2) who built the pyramids (8).
3. Demosthenes—Famous Athenian (3) orator (5) who roused his people against Philip (2).
4. Darius—Persian ruler (2) who led an army against Greece (8).
5. Solon—An Athenian (3) law-giver (7).
6. Charlemagne—King of France (4) A Holy Roman Emperor (3) and a great educator (3).
7. Attila—Leader of the Huns (4) who sacked Rome (6).
8. Constantine—First Christian (6) Emperor (2) the founder of Constantinople (2).
9. Mithridates—King of Pontus (6) who massacred the Roman inhabitants of Asia (4).
10. Justinian—Emperor of the East (3) who persecuted the Jews and Arians (2) and codified the Roman laws (5).

TEST No. II.

Any man mentioned as such in Myers' General or Ancient Histories.

TEST No. III.

1. Battle of Tours—Decisive victory for Christians (6) and the stopping of the Mohammedan invasion (4).
2. Age of Augustus—Golden age (2) of literature (5) for Rome (3).
3. Battle of Milvian Bridge—Where Christianity was accepted (7) by Constantine (3).
4. The church council of Nicaea—Determination of orthodox Christian doctrines (10).
5. Check of the Saracens before Constantinople—Saved civilization (3) and made Europe Christian instead of Mohammedan (7).
6. Reign of Alexander the Great—Extension of European powers over Orient (6) and the awakening of the civilized world (4).
7. Age of Pericles—Golden age of Athens (5) supremacy of Athenian learning (5).
8. Founding of the Hebrew monarchy—Establishment of Christianity (7) and the separation of the Hebrews from Athens (3).
9. Burning of Carthage—Rome becomes a world power (7) through the downfall of her greatest enemy (3).
10. Peloponnesian War—Gave Sparta supremacy of Greece (7) through the defeat of the Athenians (3).

TEST No. IV.

Graded by right and wrong cases; 5 points for each correct name in first and ten points each for correct answer in second part.

TEST No. V.

Within ten years, 10 points; ten to fifteen years, 9 points; fifteen to twenty years, 8 points, etc., within the same century, 4 points; outside the century, 0.

TEST No. VI. (Values not to be summated.)

1. Greeks—Culture (10), Art (5), Architecture (4), Literature (4), Philosophy and Science (4).
2. Teutons—Basis of civilization (10), Love of personal freedom (5), Respect for women (8).

3. Phoenicians—Alphabet (10), Commercial enterprise (10).
4. Saracens and Arabians—Science originating in Greece (10), Geometry and Algebra (10).
5. Romans—Law (10), Political interests (7), Military tactics (3).
6. Hebrews—Religion (10).
7. Persians—Satrapal form of government (10), Oriental customs (2).
8. Egyptians—Arts (8), Sciences (8) (both 10), Learning (3), Law (3).
9. Babylonians—Architecture (10), Literature and Religion (1), Science (3).
10. Prehistoric man—Fire (10), Language (10), Implements (5), Starting of civilization (2), Domestication of animals (8).

TEST No. VII.

Correct classification, 10 points. (Since chance alone would give $33\frac{1}{3}\%$, these questions have all been eliminated from the calculations.)

TEST No. VIII.

No score card devised and returns not yet employed in scaling.

Returns were secured from the complete surveys of the high schools of two large cities, two small ones, the random sampling of two other large ones and classes from two grades, Freshmen and Sophomores, in the University of Texas. A total of 933 papers were obtained filled out in acceptable form. These returns were graded by the score card and grouped on the basis of school, grade, and sex, though only the totals are employed in the calculations which follow. Grade and sex differences can be reported later.

The method of calculating and scoring was that used by most investigators along this line. The percentage of correctness in answering a question is assumed to be in inverse ratio to its difficulty, just as the percentage of those who spell a word is an inverse measure of its difficulty. This may not really be true in either case, but practically it is true, in that it measures the failure of present school methods to meet the situation, *i. e.*, it is a measure of inefficiency of the present methods of accomplishing the task. For present purposes the two are the same whatever might be the changed relation of difficulties and the new danger zones if teachers should begin to lay stress upon the points where deficiency is now apparent.

It was decided to use the Standard Deviation as the numerical measure for scoring the difficulty and arranging the questions on a linear scale. After calculating this standard deviation of 933 grades for some of the harder, some of the easier, and some near the median of the curve of difficulty and finding small difference, it was assumed that an average of those found would

closely approximate the average of all the sixty questions. This average was found to be 2.8. The probable error is negligible in this case.

TABLE I.

Showing Average Grades, Deviation from the Median, and the Derived Scale Values for Each of the Sixty Questions

Test no.	Av. gr.	M. V.	Sc. Val.	Test no.	Av. gr.	M. V.	Sc. Val.
I 1	3.9	— .32	187	IV 1	2.14	+ .31	250
2	3.01	— .004	219	2	4.28	— .46	173
3	4.9	— .68	151	3	1.66	+ .48	267
4	2.23	+ .27	246	4	2.85	+ .053	224
5	2.9	+ .04	223	5	1.59	+ .5	269
6	3.07	— .025	217	6	2.05	+ .34	253
7	2.91	+ .032	222	7	4.35	— .48	171
8	2.96	+ .014	221	8	1.66	+ .48	267
9	1.4	+ .57	276	9	1.92	+ .38	257
10	4.41	— .5	169	10	1.38	+ .58	277
II 1	7.23	— 1.5	69	V 1	3.4	— .14	205
2	.82	+ .78	297	2	1.76	+ .44	263
3	1.13	+ .67	286	3	2.36	+ .23	242
4	5.59	— .92	127	4	.67	+ .83	302
5	6.17	— 1.13	106	5	.73	+ .81	300
6	4.37	— .49	170	6	1.38	+ .58	277
7	7.83	— 1.72	47	7	.88	+ .75	294
8	7.62	— 1.68	54	8	1.24	+ .63	282
9	3.86	— .3	189	9	.7	+ .82	301
10	4.9	— .68	151	10	1.36	+ .58	277
III 1	2.38	+ .22	241	VI 1	5.73	— .97	122
2	3.46	— .16	203	2	2.24	+ .27	246
3	1.59	+ .5	269	3	4.65	— .59	160
4	1.95	+ .37	256	4	1.66	+ .48	267
5	2.43	+ .2	239	5	2.55	— .91	128
6	2.78	+ .078	227	6	2.59	+ .15	234
7	3.16	— .057	213	7	2.03	+ .34	253
8	.59	+ .86	305	8	2.83	+ .061	225
9	2.82	+ .064	225	9	1.50	+ .54	273
10	1.38	+ .58	277	10	2.77	+ .08	227

By observation of the average grades as found in the first column of Table I it was also found that the median grade of difficulty would fall not far from an average grade of 3 points. This was taken as the true median. The method of scoring was as follows: Question I has an average grade of 3.9 or .9 above the median in grade, or for the sake of clearer English, below the median in difficulty. Dividing this differential by the average standard deviation we have —.32, an index of the distance below the median of difficulty at which question I is to be located. Similar calculations were made on each of the other fifty-nine questions and the results are shown in the second column (M

V) of Table I. These values show the distance of each of the questions above or below the median and what is of greater importance the distance of each from the others. This, however, is still indefinite until some established point of reference is found. An inspection of the rough average grades of 933 subjects on each of the sixty questions shows that to no one was there a universally correct response; *i. e.*, no question had zero difficulty. The nearest approach was Nos. II₇, II₈, and II₁ with averages 7.83, 7.62, and 7.23, respectively. These were then submitted as Nos. B, D, and H in the following list of ten easy questions:

- A. Who were the people who lived in Egypt?
- B. Name a man famed in ancient times as a general.
- C. Was Julius Caesar a great man?
- D. Name a man in ancient history famed as an orator.
- E. Name a man mentioned in ancient history.
- F. Name a city mentioned in ancient history.
- G. On what river is Rome situated?
- H. Name a man in ancient history famed as a ruler.
- I. Was Alexander great?
- J. What is the name given to the large structures near the Nile River in Egypt?

TABLE II.

Ranks of Difficulty Assigned to Questions A to J by Ten Judges

	1	2	3	4	5	6	7	8	9	10	Av.
A . . .	2	8	5	8	6	9	9	8	3	4	6.2
B . . .	7	4	7	7	2	7	7	5	5	6	5.7
C	1	6	2	2	4	5	1	3	1	3	2.8
D	10	9	9	5	10	6	6	6	7	7	7.5
E . . .	5	1	1	1	1	2	3	1	4	1	2.0
F	4	2	3	6	3	3	4	2	8	5	4.0
G	9	10	8	9	8	8	10	10	9	10	9.1
H	8	3	6	4	7	4	8	7	6	8	6.1
I	3	5	4	3	5	1	2	4	2	2	3.1
J	6	7	10	10	9	10	5	9	10	9	8.5

These questions were submitted to ten competent judges who ranked them as to easiness, leading to ten judgments for each question. These rankings are shown in Table II. Standing out prominent in these judgments is the indication that question E is the easiest of the group. This list was also submitted to a class of 27 high school pupils who attempted to answer the question. It was found that question E was answered by all the pupils; *i. e.*, it has zero difficulty for them. While this does

not establish the fact, it is safe to assume that the difficulty of question E does not lie far from zero for any one who has really made a study of ancient history.

It will be remembered that questions B, D, and H in this supplementary list are identical with Nos. 7, 1, and 8 respectively in Test II of the original series. This was arranged so that a zero, if found at all, would be determined in such a way that its relation to the original questions might be established. By calculating the relation existing between the three known questions and the unknowns in the easy series one is able to place the zero question in relation to the original sixty, and thus to give an established zero of reference for all of them. The method for marking this calculation was as follows:

From Table II it is seen that,—

Three of the 10 judges rank D easier than B. (Both known.)

Two of the judges rank B easier than I. (One known).

Four of the ten judges rank I easier than E. (Both unknown.)

Then by using the standard deviation of the judgments on these questions as a measure of their difficulty, and interpolating those values already known the following equations are obtained wherein sigma (σ) signifies standard deviation in judgment of rank.

From D to B= .84σ—.22 units from median
From B to I= .52σ— X " "
From I to E=1.28σ— Y " "

$$\sigma = \frac{.22}{.84}$$

$$.52\sigma = .52 \times \frac{.22}{.84} = .136 = X$$

$$1.28\sigma = 1.28 \times \frac{.22}{84} = .335 = Y$$

Now since question "B" is known to be at -1.72 from the median (See Table I) it follows that question "I" is at the algebraic sum of -1.72 and $-.136 = -1.856$. Likewise, question E is $-.335$ still farther down or at -2.191 . This then establishes a zero of difficulty at a point 2.191 units below the median difficulty. Having this point of reference and correcting all values in Table I (M. V.) to read from zero instead of above and below the median, the values in the second column of Table I (Sc. Val.) are obtained and are expressed with decimal point moved two places to the right to avoid fractions.

TABLE III.

Showing Scale Values, the Intervals Between Established Points on the Scale and Subtotals to Facilitate Finding Desired Points

No.	Sc. Val.	Sc. Int.	Totals	No.	Sc. Val.	Sc. Int.	Totals	No.	Sc. Val.	Sc. Int.	Totals
E	0	0	0	I 2	219	2	3231	V 2	263	6	8005
II 7	47	47	47	I 8	221	2	3452	IV 8	267	4	8272
II 8	54	7	101	I 7	222	1	3674	VI 4	267	0	8539
II 1	69	15	170	I 5	223	1	3897	IV 3	267	0	8806
II 5	106	37	276	IV 4	224	1	4121	IV 5	269	2	9075
VI 1	122	16	398	III 9	225	1	4346	III 3	269	0	9344
II 4	127	5	525	VI 8	225	0	4571	VI 9	273	4	9617
VI 5	128	1	653	III 6	227	2	4798	I 9	276	3	9893
II 10	151	23	804	VI 10	227	0	5025	III 10	277	1	10170
I 3	151	0	955	VI 6	234	7	5259	V 6	277	0	10447
VI 3	160	9	1115	III 5	239	5	5498	V 10	277	0	10724
I 10	169	9	1284	III 1	241	2	5739	IV 10	277	0	11001
II 6	170	1	1454	V 3	242	1	5981	V 8	282	5	11283
IV 7	171	1	1625	VI 2	246	4	6227	II 3	286	4	11569
IV 2	173	2	1798	I 4	246	0	6473	V 7	294	8	11863
I 1	187	14	1985	IV 1	250	4	6723	II 2	297	3	12160
II 9	189	2	2174	VI 7	253	3	6976	V 5	300	3	12360
III 2	203	14	2377	IV 6	253	0	7229	V 9	301	1	12761
V 1	205	2	2582	III 4	256	3	7485	V 4	302	1	13063
III 7	213	8	2795	IV 9	257	1	7742	III 8	305	3	13368
I 6	217	4	3012								

Table III presents these values in the order in which they would appear on the linear scale. Subtotals are also shown to facilitate grading when the scale is used.

The method of using the scale is very simple in plan but somewhat burdensome in detail. Pupils are to be given the tests of the original in whole or in part and their work graded on the basis of 10 per cent. for each question. The score card will serve as a rough guide. The average, or median, for the class is then found for each question. This grade expressed as a fraction of a perfect grade is then multiplied into the scale value for the question. The sum of these products is the class grade in total scale points. For instance, if the scale grade of the class is 3.6 on question 10 of Test II its scale value is $.36 \times 151 = 54$. Similar calculations on each question give the total scale points achieved by the class. Now if the sum of these points is found to be 3200 by reference to the column of subtotals it is found that the class has advanced to a point on the scale between 217 and 219. The exact point may be found at 218.7 if absolute accuracy is desired. This scale point of any individual or class

means that the individual or class can answer perfectly the equivalent of all the questions below this point on the scale. If they really fall short of answering these easy questions perfectly they will answer more or less perfectly enough of the harder questions to compensate for the errors in the easier ones. Thus, the scale becomes a quantitative rather than a rigidly qualitative measure of ability in these particular aspects of ancient history. It can be made more qualitative by scaling each test separately. In its present form, however, it is comparable with many other scales which measure the amount of retention in a certain school subject after a certain amount of training in that field. It still remains to determine the qualitative and quantitative lapsing of historical information as one is further and further from a definite study in the field.