EDUCATION

I. TINKERING WITH A TRUE-FALSE TEST

George Frederick M'ller
Department of Education, University of Oklahoma

The true-false test is one of the oldest and most extensively used and abused objective achievement tests now being developed. Because of its widespread use, refinement and criticism of this type of test are especially needed. The purpose of this study is to point out the importance of a more thoro understanding of the true-false test, and more care in its use.

The following list of 26 statements, all of them except the 24th taken from a published standard examination on a widely used textbook in elementary psychology, was sent to 102 professors of psychology in colleges and universities in this country, with the request that they mark each statement as either true or false.

1. T F Objective observation is observing your own mind work.
2. T F Nerve currents always flow from axon to dedrites.
3. T F The lower sensory centers are located in the spinal cord.
4. T F Muscular fatigue is a response to external stimuli.
5. T F Most native traits are easily eradicated.
6. T F A trait that is universal is sure to be native.
7. T F Acquired traits are independent of native traits.
8. T F There is no instinct of race preservation.
9. T F Complexity of response is the main qualitative differences between instincts and reflexes.
10. T F Thirst is strictly speaking an emotion because its stimulus is internal.
11. T F Most social behavior is based on the instinct of gregariousness.
12. T F Any instinct has "play value", but some have also "survival value."
13. T F We do not learn fear, but learn what to fear.
14. T F Sensations are native mental responses.
15. T F The extreme periphery of the retina is color-blind.
16. T F Red-green color-blindness is the most common form.

17. T F Two colors are complementary if they neutralize each other.

18. T F The intensity color series runs from white to black.

19. T F Attention is fundamentally an acquired form of behavior.

20. T F Eye movement affords a picture of sustained attention.

21. T F The field of attention is co-extensive with the field of consciousness.

22. T F It is practically impossible to do everything attentively.

23. T F The IQ tells the intellectual level at any given time.

24. T F Psychoanalysis is a pseudo-science.

25. T F Persons who score high in intelligence tests are very apt to have leadership ability.

26. T F There is a high correlation between the size of the brain and the degree of intelligence.

The directions and information included with the above statements were:

1. Mark every statement as either true or false, whether they all seem clear and definite enough for such treatment or not. You may make explanations of any of them if you care to do so.

2. If the statement is true, encircle (T); if false, encircle (F).

3. What is your title (e. g. Professor of Psychology)?

4. Do you teach a general elementary course in psychology? Or have you ever taught such a course?

5. The statements that you are asked to mark, are, with one exception, part of a standard test in psychology. Have you used or are you familiar with those identical statements?

The direction to make all of the statements as either true or false whether they seemed clear and definite enough for such treatment or not, needs some explanation. It was given that way in order to make the conditions under which it was marked the same in that respect as they are for students who are examined by the test from which the statements were taken. One of the professors who marked the list said that he would be willing to change his marks on about two-thirds of the statements; which means that on that proportion of the statements his marking was a matter of chance. Another questioned the sense or use of making statements by chance. These fact sare mentioned because they need to be taken into consideration in interpreting the data. The element of chance is always present to some

degree in marking a true-false test, and of course had to be encountered in this case.

The 52 replies received are summarized in Table I.

Table I.

Summary of the results obtained from 52 professors of psychology marking 26 statements taken from a standard test in psychology.

Explanations

1. The "standard key" shows the marking that the author of the standard test gave the statements.

2. The "majority key" shows the way the majority of the 52 professors marked the statements.

3. "T" means true, "F" means false, and "X" means that no mark was given.

No. of Question	Total T's	Total F's	Total X's	"Standard"	"Majority"
1	4	46	2	F	F
2	14	37	1	T	F
3	19	29	4	F	F
4	8	43	1	T	F
5	0	52	0	F	F
6	15	35	2	F	F
7	2	50	0	F	F
8	31	18	3	T	T
9	29	23	0	F	T
10	0	52	0	F	F
11	13	36	3	T	F
12	31	19	2	T	T
13	44	7	1	T	T
14	36	14	2	T	T
15	43	9	0	T	T
16	52	0	0	T	T
17	50	2	0	F	T
18	31	21	0	T	T
19	7	43	2	T	F
20	15	35	2	T	F
21	22	30	0	F	F
22	44	5	3	T	T
23	25	25	2	F	T
24	23	25	4	?	F
25	32	19	1	R	T
26	11	41	0	F	F

Since the significance of Table I depends partly on the kind

of persons represented in it, an analysis of the personnel on the basis of institutions where located and academic rank comes next.

Table II.

Institutions and the number of professors of psychology in each from whom replies to the questions were received.

	Institutions	No. Sent	No. R'c'd
1	Agricultural College of Utah	1	1
2.	Arizona, University of	1	1
3.	Bowdoin College	1	1
4.	Brigham Young University	1	1
5	Buffalo, University of	1	0
6.	California, University of	4	2
7.	College of William and Mary	1	0
8.	Columbia University	10	4
9.	Coe College	1	1
10.	Cornell University	4	3
11.	Delaware, University of	1	1
12.	Florida, University of	1	0
13.	Hamlin University	1	1
14.	Harvard University	2	2
15.	Haverford College	1	0
16.	Idaho, University of	1	1
17.	Illinois, University of	1	0
18.	Indiana, University of	4	2
19.	Iowa State College of Agriculture	3	1
20.	Iowa, University of	2	1
21.	Johns Hopkins University	1	1
22.	Kentucky, University of	1	1
23.	Leland Stanford Jr. University	3	0
24.	Louisiana, University of	1	1
25.	Maine, University of	1	1
26.	Miami University	1	0
27.	Michigan, University of	2	0
28.	Nevada, University of	1	1
29.	New Mexico, University of	1	0
30.	New York City College	1	1
31.	New York University	1	0
32.	North Carolina, University of	1	1
33.	North Dakota, University of	1	0
34.	North Western University	2	1
35.	Ohio State University	9	7

36.	Oregon, University of	4	3
37.	Pennsylvania State College	1	1
38.	Pennsylvania, University of	2	0
39.	Pittsburgh, University of	5	1
40.	Princeton University	3	1
41.	Smith College	2	1
42.	Tulane University of Louisiana	2	1
43.	Utah, University of	1	0
44.	Vassar College	1	0
45.	Vermont, University of	2	1
46.	Virginia, University of	1	0
47.	Washington State College	3	1
48.	Washington University	2	2
49.	Washington, University of	2	1
50.	Wisconsin, University of	2	1
	Totals	102	52

A summary of further information about the individuals who marked the statements is given in the following table.

Table III.

The academic ranks and the experience in teaching elementary psychology of the 52 who marked the questionnaire.

Rank	a	b
Professor	22	8
Associate Professor	4	2
Assistant Professor	11	1
Instructor	2	0

a represents number now teaching elementary psychology.

b represents number who have taught elementary psychology but are not now teaching that subject.

Explanations

1. The title of all of the professors except five is "Professor of Psychology." The five exceptions are: a. "Professor of Philosophy and Social Science." b. "Professor of Experimental Psychology." c. "Professor of Educational Psychology." d. "Professor of Education and Psychology." e. "Professor of Education."

2. The titles of the others are "Associate Professor of Psychology," or "Assistant Professor of Psychology," or "Instructor in Psychology"; except one "Assistant Professor of Educational Psychology."

3. Only 12 of the 14 assistant professors among the 52 who replied to the questionnaire appear in the above tabulation. This is because one of the assistant professors had never taught ele-

mentary psychology, and another failed to state his experience.

Probably the most noticeable fact in Table I is the disagreement in the markings by the different individuals on only three of the statements, 5, 10, and 16, is there complete agreement. On the 23rd opinions are equally divided.

One question that arises is, How is this disagreement to be explained? Perhaps it is due to faulty wording of some of the statements, causing a lack of clearness in the meaning. Ambiguity as in No. 2, comes under this head. Some of the disagreement might be due to the question being a controverted one among psychologists, or to the fact that the point has not yet been scientifically determined or discovered so that no one knows it definitely. Some other possible reasons for the disagreement are: carelessness in reading and marking the questions; the meaning of the sentence might have been missed, because the individual was not familiar with the context or setting from which it was taken; or even simply not knowing the psychology involved.

Just which of these causes were effective in the results obtained, and to what extent each was potent can't be answered from the data at hand. There is evidence, however, that not understanding just what was meant, in some cases caused disagreement. This is indicated by the correlation between the number of comments made on the statements and the disagreements on them, which was r equals .76 plus or minus .06. This means that as a rule, the statements that aroused most comment were the ones on which the marking was most evenly divided. If we assume that the psychologists commented most frequently on those about which they felt most dissatisfied and uncertain as to the meaning, a lack of clearness (whether due to poor wording of the sentence or to its being taken out of its context) was one of the sources of disagreement. A feeling of uncertainty about the meaning of a statement, of course, is not the only possible factor back of commenting, but it is probably the principal one. Table IV will help to explain how the above correlation was derived.

TABLE IV.

A quantitative statement of agreement or disagreement of the 52 individuals in marking the 26 statements, and the number of comments that each individual made on each statement—r equals .76 plus or minus .06.

a—number of the statements.

b—total number of individuals marking each.

c—"Minority" markings.

d—Percent "minority" is of total.
e—number of comments on each.

a	b	c	d	e
1	50	4	8	5
2	51	14	27	12
3	48	19	39	12
4	51	8	16	12
5	52	0	0	1
6	50	15	30	9
7	52	2	4	5
8	49	18	37	11
9	52	23	44	8
10	52	0	0	3
11	49	13	26	9
12	50	19	38	9
13	51	7	14	9
14	50	14	28	10
15	52	9	17	9
16	52	0	0	2
17	52	2	4	7
18	52	21	40	11
19	50	7	14	6
20	50	15	30	9
21	52	22	42	12
22	49	5	10	5
23	50	25	50	17
24	48	23	48	17
25	51	19	37	10
26	52	11	21	16

Note: To obtain the correlation the rankings for comments were made by counting the number of individua's who commented on each statement. For example, only one person wrote a comment on No. 5, whereas 17 of the 52 answering commented on No. 23. This is shown in the table above. To find a quantitative statement of agreement the percent that the "minority marking" was of the "total marking" of each question was taken. The reason that the "total marking" is often less than 52, is because some individuals failed to mark some of the questions either true or false.

Another pertinent question is, Are these 26 statements good to use in a true-false test for students in elementary psychology? Several of the psychologists anticipated a similar question, and suggested answers. One opinion was that the more the indivi-

duals know about psychology, the more they will disagree in such a test, and that since beginners are sometimes taught in a dogmatic way, such questions might be all right for them. There is a slight indication in the data of this study that the more a person knows about psychology the poorer mark he will make on this particular test. The average score of the 30 professors marking the test was 4.26. The average for the 14 assistant professors was 6.30. The recency of experience in teaching elementary psychology, however, favored the assistant professors, since only 2 of the 14 were not teaching it, whereas 8 of the 30 professors were not teaching it (Table III). Probably the chief advantage of the assistant professors over the professors (if the small number of cases does not render all comparison insignificant) is that nearly all of the former are in the larger universities, whereas a large proportion of the latter are in small colleges where the requirements are usually not so high. In the matter of familiarity with the test, the professors and the assistant professors were practically equal. The 6 associate professors and the 2 instructors, both numbers too small to be of much significance, made average scores of 7.16 and 7.00 respectively.

Another opinion was that the type of test was good, but that most of these particular questions were poor. The data themselves, especially Table V, will throw some light on this point.

TABLE V.

Scores that the 52 individuals made on the 25 statements of the standard test, marked according to the "standard key"; percentage marks based on the scores; and whether or not each individual was familiar with the standard test from which the 25 statements were taken.

Explanations: 1. The scores were obtained by subtracting the "wrong" answers from the "right". Statements that were not marked were first subtracted from the total number, 25. The formula followed is the usual one. (25—unmarked questions) minus (2 x number wrong).

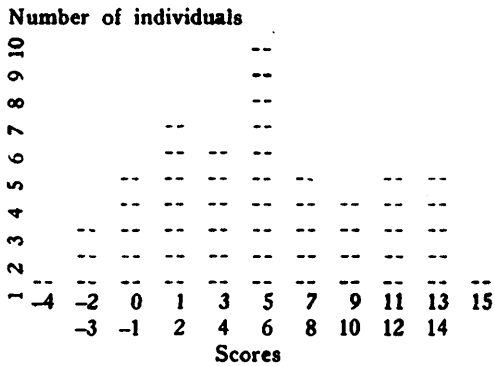
2. In this table 25 statements are used instead of 26 as in Table I, because one of the 26 (No. 24) was not in the original standard test, and could not be scored by the "standard key."

3. The "standard key" means the marking that the author of the standard test has specified (Table I).

Individuals	Scores	Percentage marks	Familiar with test or not
1	-2	-8	Yes
2	5	20	No
3	-1	-4	No
4	0	0	?
5	-1	-4	?
6	-2	-8	No
7	11	44	No
8	3	12	?
9	3	12	No
10	5	20	No
11	0	0	?
12	2	8	Yes
13	11	44	Yes
14	5	20	No
15	8	32	No
16	3	12	No
17	9	36	Yes
18	5	20	No
19	11	44	Yes
20	1	4	No
21	-4	-16	No
22	7	28	No
23	9	36	Yes
24	13	52	?
25	10	40	No
26	1	4	Yes
27	12	48	No
28	5	20	No
29	5	20	No
30	5	20	No
31	3	12	No
32	13	52	No
33	13	52	Yes
34	5	20	No
35	3	12	Yes
36	15	60	Yes
37	-1	-4	No
38	-2	-8	No
39	7	28	No
40	5	20	No
41	13	52	No

42	1	4	?
43	1	4	No
44	2	8	No
45	11	44	?
46	1	4	No
47	3	12	No
48	8	32	?
49	13	52	Yes
50	9	36	No
51	5	20	No
52	7	28	?

Graph for Table V



It will be noticed that the scores are low, compared with the highest possible score of 25. The highest score made was only 15 and only 11 persons made more than 10. If these scores are translated into percents, counting the perfect score of 25 equal to 100%, the highest mark is only 60, and 46 of the 52 psychologists marking the test made a mark of less than 50%. If they had been a class in elementary psychology, and the passing mark had been 70%, all would have failed. If the passing mark had been 60%, only one would have passed. The author of the textbook on which the standard test is based, himself made only 52%. In the light of such results, and when it is remembered that among the individuals represented in Table V are a number of the most eminent psychologists of the times, it requires a study faith to hope that this particular test might be suitable for elementary students of the subject.

Table V shows that those who stated that they were familiar with this identical standard test made a much higher score on

the average than the others. Those familiar with it averaged 7.4, those not familiar with it, 4.6, and those who did not state whether they were or not averaged 3.7. It is fair to assume that most of those familiar with the test were acquainted with the textbook upon which it was based, and so would have as much of the setting or context as would the students who use that text. At least the author of the text would have the necessary background. While this background was evidently a help in most cases, it did not enable those who had it to make a very good score. There seems to be something wrong in addition to lacking the setting.

The frequency distribution graph for Table V does not support the theory advanced by some that the marking would show to what school of psychology individuals belonged. It represents about the usual distribution for a group of that size.

The foregoing facts support the following conclusions:

1. This particular true-false test is as a whole poor.

2. Statements 5, 10, and 16; and probably 1, 7, 17 and 22 are sufficiently clear and well established to be good for a test of this type. This study, however, throws no light on the importance or difficulty of those statements—considerations that are very essential in making up a good true-false test.

3. It would be best to change the marking of No. 17 from F, as the "standard key" has it, to T, which is the majority marking.

4. All of the other statements except the seven mentioned above should be modified or discarded.

5. This study is no evidence against the value of true-false tests in general, but it does indicate that the preparation of such tests should not be considered a simple easy matter. Considerable investigation and precaution should be exercised in their construction and use.