An Investigation of the Objective Examination Questions in a First Year College Physics Course

Dorothy H. Helmick
Drake University

Recommended Citation
Available at: https://scholarworks.uni.edu/pias/vol34/iss1/65
AN INVESTIGATION OF THE OBJECTIVE EXAMINATION QUESTIONS IN A FIRST YEAR COLLEGE PHYSICS COURSE

DOROTHY H. HELMICK

The purpose of this study is an evaluation of the efficiency of the objective-type examination questions used in the College Physics course at Drake University during the first semester 1926-27. Ninety-nine students enrolled in the course, and eighty-one took the final examination.

There were used during the semester 154 questions of the true-false type, 31 questions of the multiple choice (4 or 5 choice) type, 50 analogies (scrambled units and definitions), 5 formulae to complete, and 32 problems.

The true-false, multiple choice, and analogy questions were each graded 1 if correct, the formulae, 2 each, and the problems, 4 each. In order to determine the letter grade of the student, each test was scored in the following manner:

The standard deviation, sigma, and the mean score of the array of raw scores were first determined. The mean raw score was set equal to 50 on an arbitrary “T-score.” The mean raw plus 3 sigma was set equal to 100 on this T-score, and the mean raw score minus 3 sigma was set equal to a T-score of zero. Other T-scores were determined by a graph of a line drawn through the above points. Then the T-score range 0-9 was set equal to F, 10-19 equal to E, 20-29 to D, 30-39 to D plus, and so on to 90-100 equals A plus. According to this basis, in the long run, 0.82% of the students would receive A plus or F, 2.72% A or E, 7.92% B plus or D, 15.92% B or D plus and 22.57% C plus or C.

These percentages may or may not represent a distribution of marks perfect from the standpoint of grading, but at least they furnish us with a working basis by which we may compare the grades made in different tests by eliminating the varying degrees of difficulty from the tests.

1 Cf. with the T-scale of McCall (e.g., Garrett, Henry E., Statistics in Psychology and Education, p. 110. Macmillan). McCall takes 100 on his T-scale equal to the mean raw score plus 5 sigma, and zero as the mean raw score minus 5 sigma. This would give only about 0.003% of the students a T-scale mark of 0.00 to 9.99, or 90.00 to 99.99, and 0.13% from 10.00 to 19.99, or 80.00 to 89.99.
The results of the answers to each question of the test were presented in graphical form in the following manner: The papers from each examination were divided into their respective letter groups on the basis of the T-score grades. The percentage of students in each group answering each individual problem correctly was then computed and graphed. Thus this curve shows the percentage of the A plus or A or B or any other rank of students who answer any single question correctly.

The papers from each examination were also divided into a letter grouping on the basis of the T-score of the student's intelligence quotient (I.Q.). The percentage of students in each I.Q. level which answered any question correctly was also plotted against the I.Q. level, on the chart mentioned above. Thus for each question a chart was made showing: (1) the relation between the percentage of correct answers and the physics grades of the students, and (2), the relation between the percentage of correct answers and the intelligence levels of the students.

In the main, there is good agreement between the physics grade distribution curve, and the I.Q. distribution curve. Throughout the major portion of these curves the variation between the curves averages about 10%. At the extremes of curves, there is sometimes much greater variation between the curves because the points here represent only one or two students (A plus, A, E and F grades).

The fact of this good correlation between I.Q. and physics grade negatives the possibility that any of the questions investigated, which was answered with difficulty by only those students of superior physics grade could also be answered with ease by students of lower mental endowment.

Patterson interprets a good question as one which sharply differentiates between two successive letter groups of students. For example a question upon which all students D and lower fail, but all students D plus and higher pass, would be considered an ideal question.

It appears that few actual questions differentiate so sharply between levels of ability. For our purpose we shall classify a question as good which is answered by few inferior students, but is answered with greater and greater ease as the superiority of the groups increases.

The curves drawn according to the foregoing method could be analyzed into the ten different classes shown in Fig. 1. The per-

---

2 The I.Q. of the students was determined with an Otis Self-Administering Test of Mental Ability Higher Examination—Form A.
3 Preparation and Use of New-Type Examinations, p. 63. World Book Co.
centages after each curve show the proportion of the true-false, choice, analogy, formulae, or problem questions which belong to that particular class, the number of questions which fall into the given class of curve, and finally the number of questions in the particular class expressed as the percentage of the total.

The ideal question of Patterson, or Class I, is approached by true-false questions such as: "When equilibrium is produced in a body, the body has no tendency to rotate about an axis chosen
at any point on the body”; “A body taken down in a mine will weigh less”; and, “The attraction between two one-hundred pound balls placed a few inches apart amounts to about one pound.”

The E and F students failed on these completely, but a major percentage of the students of higher rank answered these correctly. A question of the multiple choice type which falls in this class is: “The motion of the axis of rotation is called (1, gyroscopic; 2, moment of inertia; 3, precession; 4, rotation).” A problem which falls in this same group is the following: “Find the final velocity, the average velocity and the distance traveled by a freely falling body after falling for 4 seconds when its initial velocity was 20 ft. per second.” Students whose physics grade ranged between D plus and F or whose I.Q. ranged between D and F completely failed this question. The question was answered with increasing degrees of ease by students whose physics grades ranged between C and B plus and those students whose physics grades were A and A plus answered it 100% perfect.

True-false questions of Class II are illustrated by: “There is less danger that hammock ropes will break if they are nearly vertical”; “When strips of two different metals are riveted together and heated until they assume a curved form, the strip having the greater coefficient of expansion will be on the inner side of the curve”; and, “A hot-water radiator receives most of its heat by conduction.” Questions of multiple choice are given by “It requires the continued application of a constant force to keep a free body moving at uniform (1, acceleration; 2, momentum; 3, speed; 4, velocity).” Problems involving coefficients of friction, forces to equilibrate bodies held on smooth inclined planes, and the solution of jack-screws also gave this kind of curve. The shape of the curve shows that the questions present the same degrees of difficulty for the inferior student but becomes very much easier of solution for the superior students.

True-false curves of Class III are given by such true-false statements as: “The metric unit of mass is the pound”; “When forces act in the same direction they may be added arithmetically”; “The shorter the handle of the hammer, the greater the torque which can be exerted when pulling a nail.” Many questions of analogy come in this group such as: “U. S. yard and 3600/3937 meter,” “Acceleration and force mass,” “453.6 gm. and 1 lb.”

Curves of Class IV are given by 46% of all the true-false questions.

Curves of the Class I, II, III, and IV may be regarded as
denoting questions of satisfactory form, for they differentiate unmistakably between the various classes of students.

Questions which give a curve of Class V, however, are answered with equal facility by students of all levels of ability and thus contribute nothing to the purpose of the examination which is to differentiate between the various levels of student ability. These curves are given by such questions as: “In a perfect vacuum all bodies fall at the same rate” (96% correct answers); “Every body continues in its state of rest or of uniform motion (1, in a straight line; 2, in a circle; 3, with constant acceleration; 4, in a horizontal direction) unless acted upon by outside forces” (82% correct); “Newton’s second law and \( F = ma \)” (60% correct); “A stone is dropped from a cliff and strikes the ground in 3 sec. Neglecting air resistance, find the speed of the stone as it struck the ground” (80% correct).

There were three curves of type VI, all obtained in the first test given, which were answered better by the low students than by the high students. These questions were as follows: “A foot-pound of work is done when a force of 1 lb. makes a displacement of 1 ft.”; “The higher you aim a rifle, the greater distance you can shoot,” and “It is impossible for a person in an auto travelling on a smooth road to tell when the car accelerates.” It seems possible that for these questions the best students worked so fast that they did not take time to fully analyze them. Undoubtedly the rifle question appeared ambiguous to many students, and this fact may have had some bearing here.

But one curve was found of type VII, where both the high and the low students ranked lower than the average students. The question which produced this peculiar phenomenon was: “Heat can be taken from bodies which are colder than their surroundings.”

Curves of Class VIII were found principally in the true-false questions. Some questions which give this type of curve follow: “If liquid flow is ‘steady,’ each particle in the liquid has the same speed”; “Strain and change of size or shape”; “In the case of a rotating body it is always possible to find a line which is not moving. This line is called the: 1, angular velocity; 2, axis of rotation; 3, coördinate axis; 4, torque axis”; “A body is thrown downwards, initial speed 40 ft. per sec. What will be its speed in 5 sec.?” It appears that the poorest students, knowing their weakness, have specially prepared themselves on these points, and there-
fore obtain results on these questions which are better than the results of the average students.

In the curves of Class IV, the better the student the poorer the grade he obtains. It seems significant that the three questions conforming to this class occurred in the final examination, and concerned matters which had been mentioned in lectures, but which had not been particularly emphasized. An explanation of these curves might be that these poor students appreciating their weaknesses study their lecture notes more carefully than better students in preparation for their final examination, for they appreciate that they have comparatively more to gain from a good final examination than have the good students.

Curves of Class X are useless in differentiating between levels of ability. There were four of this type, all involving the completion of formulae. For this type of question, it appears that such factors of memory and application so complicate conditions that there is no direct relation between correct answers and student ability.

Many of the questions were repeated in the final examination in order to determine what effect this repetition might have. The two sets of results in these cases showed no particular difference.

In conclusion, this study shows that of the 272 questions investigated, 64% clearly differentiated between levels of ability (Classes I to IV), 26% were only for "buffer" questions to place at the beginning of the examination to increase the student's confidence, for they were answered with the same ease by all classes of students, while the remaining 10% did not show a desirable correlation between correctness of answer and level of ability. This study brings out the fact that in preparing objective examination questions, the examiner might avoid the inclusion of considerable dead timber (amounting in this case to 26% of the total number of questions) if he will eliminate those questions, particularly of the true-false, and of the analogy type, where the students' minds need only function at the lowest mental levels.

Drake University,
April 18, 1927.