Proceedings of the Iowa Academy of Science

Volume 56 | Annual Issue Article 35

1949

Confidence in Correctness of Responses on an Objective Test

Martin F. Fritz
Iowa State College

Confidence in Correctness of Responses on an Objective Test

By MARTIN F. FRITZ

A form of objective test has been given at various times in which the student is required to indicate how certain he is of the correctness of his answers. This may be accomplished by having an answer sheet prepared with three columns, each column representing a degree of certainty according to a so-called scale. Variations in certainty, with the number of points to be gained or lost, may be indicated as follows:

Degree of Certainty			*Gain or Loss*
Column One	(G)	Guess	1 point
Column Two	(FC)	Fairly Certain	2 points
Column Three	(AC)	Absolutely Certain	3 points

When a student feels that his response is not much better than a guess, he may place his answer in the first column. If he is wrong, he loses one point but if he is correct he gains one point. An answer placed in the second column indicates he is fairly certain that he has made the correct response and stands to either gain or lose two points. Great confidence in the correctness of responses may be shown by use of the third column. In this case, the reward for each right answer is three points but on the other hand, the penalty for a wrong answer is the loss of three points.

It will be at once apparent that this is a system for rewarding those who are confidently correct in their responses and at the same time a way of severely penalizing those who are confident but wrong. It might well be argued that he who "knows he knows" should receive due credit and that it is also desirable to be more harsh with a person who is certain that he knows but is wrong.

Data on only one class in psychology, Tests and Measurements, involving 43 students are reported in this paper since they are more or less typical of what has been found upon a number of occasions. The test was a final examination of 100 items, 20 true-false and 80 five-part multiple choice questions, over Mursell's textbook, *Psychological Testing.*

One of the questions suggested by such a study is: How well can students identify test-items they do not know? In other words, is there any indication that students are capable of selecting those items on which they are likely to lose credit? The answer to this question is that there is some reason to believe many students can

detect their weak points since 61% of all the answers (for the entire class) placed in the "guess" column were incorrect (177 correct versus 288 incorrect responses). Of the 40 students who placed responses in the "guess" column, 32 lost more credit than they gained. It would seem that when they thought they were guessing, they usually were! The percentage of incorrect answers in the "fairly certain" column dropped to 43½ (439 correct versus 339 incorrect responses). Out of 38 students who placed responses in the "fairly certain" column, 16 lost more credit than they gained. The percentage of incorrect answers in the "absolutely certain" column was only 21 (2409 correct versus 656 incorrect responses). No student lost more credit than he gained on responses placed in the "absolutely certain" column. Although these percentages will vary from one class to another and from one examination to another, there would seem to be little doubt that considerable discrimination among questions can be made by many students before the correct answers are known.

Not all the subjects gained points by moving responses out of the "absolutely certain" category which suggests that there are students who do not know what they know. In other words, more correct than incorrect responses were moved out of the "absolutely certain" category. Eleven failed to gain, 6 in the upper half and 5 in the lower half. To this number might be added 5 more who gained only 1 or 2 points, perhaps not much better than chance. By far, the greater number gained points — 30 gaining as compared to 11 losing points.

Another question of some interest is whether or not the better students are more capable of detecting those questions on which they are likely to lose credit. The combined net gain of all students in the upper half of the class (on this particular examination) as a result of moving responses out of the "absolutely certain" column was 46 points. Those in the lower half of the class made a net gain of 72 points. The chi-square for this difference is statistically significant at the 2% level indicating, if anything, that the poorer students were more aware of the test items which might give them trouble. Also, it would seem that this difference was primarily due to the fact that the poorer students made much greater use of the "guess" column, since both the upper half and the lower half missed almost exactly 50% of all answers not placed in the "absolutely certain" column.

It might well be that personality factors play a part in an examination when it is scored as described in this report. Possibly cau-

tiousness, timidity, willingness to take a chance, or even feelings of inferiority have some influence on the total score. Two students were bold (or rash!) enough to place all their answers in the "absolutely certain" column while at the other extreme was one placing only 31 responses there. Some students placed a rather large number of responses in the "guess" column while others made considerable use of the "fairly certain" category. Perhaps we have here a technique which might be used for certain explorations of personality.

A weighted score of 1, 2, or 3 points, gain or lose, for each test item results in an extremely wide range. On this particular examination, allowing for chance, the scores could vary from a high of 300 to a low of minus 144. Actually, the variation was from a low of 43 to a high of 198 or a range of 155 points. With such a wide scattering of scores, it is possible to draw lines for assigning letter grades with a great deal more ease and confidence. This advantage of greater scatter must be set off against the disadvantage of a scoring system which is both complicated and time consuming. Unweighted scores, simply counting one point for each correct response, would have given a low of 54 and a high of 83 or a range of only 29 points. This method of scoring is simple and rapid but it does result in a relatively compressed range. Since the correlation between weighted and unweighted scores was found to be 0.97, which is about the value that has been found with other classes, it may be safely assumed that the ranking of students by the two methods will be highly similar.

SUMMARY

An objective test was devised in which the student was required to indicate on each item whether he was "guessing," "fairly certain," or "absolutely certain," thereby either gaining or losing 1, 2, or 3 points. There is some reason to believe that many but not all students were able to detect those questions on which they were most likely to lose credit, and that poorer students were more capable than good ones of detecting such questions. The weighted scoring method gives a very wide scatter but the weighted and unweighted methods result in practically the same rank order of the subjects as indicated by a correlation of 0.97. It is suggested that this technique could be used for explorations of personality.

DEPARTMENT OF PSYCHOLOGY AND THE TESTING BUREAU,
IOWA STATE COLLEGE.