Alternate Multiple-Choice Grading Method

Kyle Nazarchuk

Abstract

Exam scores are often used as a measure of a student's competency and therefore mastery of content. However, when using multiple-choice tests, there is no ability to gauge partial knowledge of a concept when using the standard method (also referred to as the number of correct answers), where there is only one correct answer to receive full credit. If a student understands a concept partially and can eliminate one or two distractor answers but chooses the last distractor, they are evaluated the same as a student who completely guessed their response. A method to gauge partial knowledge is to have a secondary answer that is similar to the correct answer conceptually but flips a component that would be common for students to choose. This response has the potential to demonstrate that the student understands the concept partially and would benefit from receiving partial credit to reflect their true level of understanding.

Keywords: Testing, Assessment, Multiple-Choice

An Inquiry into an Alternate Multiple-Choice Grading Method with a Secondary Answer

Research that focuses on how assessments are graded is limited. When looking at testing, the way instructors inform students about how they are grading their exams can be just be as important as the grading method. Dr. Yoella Bereby-Meyer, professor of psychology at Ben-Gurion University, investigated this by having two groups of students take an exam; one group was told they would be awarded credit for getting correct answers and the other was told they would be punished for being wrong. Rather than looking at grades, the team looked at how students answered a set of questions that had no correct answer. What they found is that how the instructions were provided did change student behavior, with students who were told they would be awarded for getting a question correct answering these questions 72% of the time, compared to 43% of those who were told they would be punished for being wrong (Bereby-Meyer et al., 2013).

Research on how to award partial credit for multiple choice questions has been going on for decades, with most requiring the student to select multiple answers. Dr. David DiBattista, a professor of psychology at Brock University, examined a way to assess partial knowledge by allowing students to rank the potential answers. If the

correct answer was their first selection, they were awarded four points, if the correct answer was their second, they were awarded three points, and so on. However, this method did not improve the validity or reliability of test scores (DiBattista et al., 2009).

One method described by two professors from London South Bank University calls for students to not select any answer they are confident is incorrect, but to mark any that they are unsure of (Otoyo & Bush, 2018). This would give students the opportunity to say, "I know these aren't true, but I cannot tell between these X answers." Assuming the correct answer is included in the answers selected by the student, the student would be awarded full credit for the question if they only selected one answer, half credit for two, one third for three, and one quarter if all four were selected. In looking at the average test scores of three tests that used this system, there was almost no change, or even a decrease in scores, when using this subset selection method (Otoyo &Bush, 2018).

Methodology

In the Fall 2022 semester, three non-cumulative examinations were administered in a science course designed for non-science majors. Each exam had 35 multiple choice questions, where students were to select one of four answers, one correct worth full credit, a secondary answer worth half credit, and two incorrect answers worth no credit. Answers were recorded on a bubble sheet and graded automatically with Remark, a test grading software by Gravic Inc. Tests were completed in a normal 50-minute class, with students receiving one of two versions of the test, with one having the questions reversed compared to the other, and no other differences. Tests were graded with just

correct answers, the standard grade, and with both correct and secondary answers, a secondary grade. A paired-t test was performed to evaluate the statistical significance of the grade departure with the secondary method compared to that of the standard method.

Results

All three exams had a statistically significant non-zero improvement to class averages, with the lowest average improvement being 6.96%. Exam one had the largest improvement of more than a letter grade. Those who already performed well on the exams in the normal grading scheme had little to no improvement with the secondary system, which is to be expected as these students would have the fewest questions where the secondary answer could be selected, and manifest in the student's secondary grade. Conversely, the students who received the highest improvements from standard to secondary grading methods were those who performed relatively poorly, 50% and below, and had more questions where they could select the secondary response.

Conclusion

The objective of this study was to investigate a multiple-choice testing method with four potential answers, one correct worth full credit, one secondary worth half credit, and two incorrect worth no credit. These three non-cumulative exams were conducted in a 100-level general education science course. The average of all improvements across the exams were 6.96 ±.44%, which was in line with other attempts to award partial credit for partial knowledge. To continue expanding on this research, future studies could better control for variables such as study habits and comfort with exam content.

References

Bereby-Meyer, Y., Meyer, J., & Budescu, D. V. (2003). Decision making under internal uncertainty: The case of multiple-choice tests with different scoring rules. *Acta Psychologica*, *112*(2), 207–220. doi:10.1016/s0001-6918(02)00085-9

DiBattista, D., Gosse, L., Sinnige-Egger, J.-A., Candale, B., & Sargeson, K. (2009). Grading scheme, test difficulty, and the immediate feedback assessment technique. *Journal of Experimental Education*, *77*(4), 311–338. doi:10.3200/jexe.77.4.311-338

Frary, R. B. (1989). Partial-credit scoring methods for multiple-choice tests. *Applied Measurement in Education*, *2*(1), 79–96. doi:10.1207/s15324818ame0201_5

Otoyo, L. & Bush, M. (2018). Addressing the Shortcomings of Traditional Multiple-Choice Tests: Subset Selection Without Mark Deductions. Practical Assessment, Research & Evaluation.. *Practical Assessment, Research & Evaluation*, vol. 23, no. 1