Evaluation of multiple-choice question exam analysis of preclinical subjects

Varanya Srisomsak[1]
Chantacha Sitticharoon[1], Issarawan Keadkraichaiwat[1] and Sunan Meethes[2]
[1] Department of Physiology, Faculty of Medicine Siriraj Hospital, Mahidol University
[2] Education department, Faculty of Medicine Siriraj Hospital, Mahidol University

Abstract

Background: Exam analysis has two main statistics: the difficulty index (p), and the discrimination index (r). Remarkable exam analyses are based on p-value<0.25 and/or r-value<0 (thresholds), but corrections may also be needed when the correct choice's p-value is comparable to other choices.

Summary of Work: The research investigated exam analyses of preclinical subjects (46 subjects, 237 exams) from academic year 2017-2022 to determine characteristics of abnormal exams.

Results: Exam corrections were 34.18% (81/237), with a mean p-value of 0.701 (0-0.780) and a mean r-value of 0.284 ((-0.250)-0.440), resulting from multiple answers (46.91%), wrong answers (40.74%), question removal (4.94%), and awarding points to all choices (7.41%). Excluding question removal (no p-value and r-value available), corrections of p-value≥0.25 accounted for 19.48% of total corrections, 28.95% for multiple answers, and 12.12% for wrong answers. For r-value≥0, corrections were 49.35% of total corrections, 73.68% for multiple answers, and 66.67% for awarding points to all choices, and 18.18% for wrong answers. Regarding multiple answers, the first vs. second correct choices had p-values of 0.215 vs. 0.499, respectively, and r-values of 0.07 for both. For wrong answers, the incorrect vs. correct choices had p-values of 0.101 vs. 0.676, respectively, and r-values of -0.070 vs. 0.201, respectively.

Discussion: The main reasons for exam corrections were multiple answers and wrong answers. Relying on p-value and r-value thresholds may lead to under-detection of abnormal exams. Additional correct choice(s) often had higher p-value than the correct choice.

Conclusion: The use of thresholds resulted in 20-50% underdetection of abnormal exams, making it necessary to include correct choices' p-values comparable to other choices in a new guideline to improve evaluation accuracy.

Take Home Messages: Using only p-value and r-value thresholds can under-detect abnormal exams. Monitoring exams where the correct choice's p-value is equal to or less than other choices might prevent inaccuracies.

References (maximum three)

No references