

## LETTERS

### Throwing a Curveball: The Impact of Instructor Decisions Regarding Poorly Performing Test Questions

*To the Editor.* Skilled pitchers in the game of baseball often throw a curveball to trick an unsuspecting batter to swing and miss. Much like the batter who is thrown off balance by a tricky pitch, a student may be misled by a tricky question on a multiple-choice test, regardless of whether the item writer intended to “pitch a curveball.” Additionally, decisions regarding what to do with test questions that perform poorly on item analysis can affect scores, and subsequently course grades, in unintended ways. Item analysis of multiple-choice questions is an important tool to ensure test questions fairly assess learning instead of tricking or deceiving students. In addition, students’ test scores can significantly change depending on how we handle poorly performing questions – a situation that can be easily overlooked by course directors. As a result, it is the course director or instructor who is unintentionally deceived.

Although there are some generally accepted guidelines for statistically defining a poorly performing test item, how instructors respond to item analysis information may not follow a consistent pattern throughout a curriculum. At the University of Tennessee Health Science Center College of Pharmacy, we have encountered substantial variability among course directors in dealing with contested and poorly performing test questions. Course directors may choose to score examinations by keeping poorly performing questions, counting multiple answers correct, or by eliminating the entire question from the test, with subsequent readjustment of student scores. General agreement exists among our faculty members that when item analysis reports a negative point biserial, these questions should be reviewed and adjustments made. However, there may be situations when the item analysis indicates a valid question (eg, point biserial value of 0.45), yet the course director at his/her discretion omits the question or counts multiple answers as correct, thereby giving all students credit. We have found that such changes can have a substantial, highly variable, and sometimes unpredictable impact on students’ test scores. Consider the following scenarios that we have encountered where instructor decisions throw curveballs in relation to students’ test scores.

**Example 1: Omitting questions from a test.** An examination has 25 questions total. Item analysis reveals 3 relatively poorly performing questions (40%-60% of

students responded correctly), yet each question has a high point biserial value, indicating that all are strong assessments of student learning. Based on personal preference, the course director removes these 3 questions from the test, reducing the total possible points to 22. Student 1 got 21 of the 25 questions correct including all 3 of the omitted questions. Thus, this student’s initial test score was 21/25, or 84%. His adjusted score, after omitting the 3 questions, is 18/22, or 82%. Student 2 got 21 of 25 questions correct yet missed all 3 of the omitted questions. Her initial score was 21/25, or 84%, while her adjusted score is 21/22, or 95%. In this case, student 1, who selected the correct answers to the omitted questions, experienced a reduction in his score. However, student 2, who selected the incorrect answers, received a final score that was significantly improved after the omission of the questions.

**Example 2: re-keying answers to give credit.** A test has 25 questions. Three questions, all with high point biserials and item discrimination, are contested by students. After reviewing the item analysis, the course director decides to rekey the answers to give all students credit for those 3 questions. Because student 1 got 21 of the 25 questions correct, including all 3 of the questions of concern, her initial and adjusted examination scores are the same 21/25, or 84%. Student 2 got 21 of 25 questions correct, but on initial scoring missed all 3 of the questions of concern. His initial score was 21/25, or 84%, while his adjusted score is now 24/25, or 96%. In this case, only the student who initially selected the incorrect answer receives an improved score.

**Example 3: awarding bonus points.** Using the example above, instead of altering the 3 questions of concern, the questions are kept, and the examination is scored using the students’ current answers. However, the course director decides to give all students 3 bonus points. The initial examination scores for both students 1 and 2 are 21/25, or 84%, while adjusted scores for both with the additional 3 points are 24/25, or 96%. The awarding of 3 bonus points in this scenario provides a greater relative improvement in examination score for students who performed poorly on the examination compared to those who did well. For example, a student’s initial score is 10/25 (40%) but with the 3 bonus points, his score increases to 13/25 (52%, a relative change of 30%). For the student who got 21/25 initially correct, the score increases from 84% to 96%, a relative increase of 14%. Although all students are treated equally in the awarding of bonus points, some students’ performance may not have merited this adjustment. Another approach would be to give bonus questions, where students have extra chances to perform well but are only given credit for those questions answered correctly.

As a result of course directors' variable approaches to what they perceive as poorly performing test questions, student pharmacists' frustration over these inconsistencies to examination grading has become apparent to our college. Unlike baseball where a curveball is intentionally thrown to deceive the player and thereby influence the outcome of the game, the ultimate goal for instructors and course directors should be to fairly and accurately assess student learning in a manner that is equitable to all. Throwing out poorly performing examination questions must be carefully considered, lest we impact student scores in an unintended way, and the curveball is instead

thrown back at us. Although faculty members may be well intentioned in resolving test item performance issues, scoring adjustment is more complex than some of us realize. Colleges should be encouraged to create mechanisms to monitor, guide, and support faculty in this area.

Stephanie J. Phelps, PharmD,  
Sharon K. McDonough, PhD,  
Robert B. Parker, PharmD,  
Shannon W. Finks, PharmD  
College of Pharmacy, The University of Tennessee Health  
Science Center, Memphis, Tennessee