Can Content Experts Rely on Others to Reliably Score Open-Ended Questions on Summative Exams?

Doreen M. Olvet, PhD, Jeffrey B. Bird, MA, Tracy B. Fulton, PhD, Marieke Kruidering, PhD, Klara K. Papp, PhD, Kelli Qua, PhD, Joanne M. Willey, PhD, and Judith M. Brenner, MD

Purpose: Although the multiple-choice question format is the primary means of assessing medical students’ knowledge, interest in the use of open-ended questions (OEQs) for assessment is growing.1 Time burden and subjectivity of grading OEQs are cited as concerns.2,3 Recruitment of graders without content expertise could reduce the time burden of grading. However, it is unclear if rubrics can be reliably used by individuals who are not content experts. The objective of this study was to evaluate the interrater reliability (IRR) of OEQ scores by faculty, students, and noncontent expert faculty.

Approach/Methods: Content experts at 3 U.S. medical schools created 3 questions testing the knowledge of the pathophysiology of the gastrointestinal system among first-year medical students (MS1s), based on a set of learning objectives that were identified collaboratively. Rubrics consistent with the school’s assessment program were created internally at each school. Schools A and B used a holistic rubric based on a 6-point scale with descriptive anchors. Based on this scale, a categorical determination was made as to whether the student’s response met expectations (5–6 points), was borderline (3–4 points), or did not meet expectations (1–2 points). School C used an analytic rubric that allotted points for each component required for credit. Each component was worth 1 point and the total number of points were summed for each question. Study questions completed by MS1s were scored by content experts, who were considered the gold standard. Each site recruited noncontent experts and fourth-year medical students (MS4s) to independently score the questions. Intraclass correlation coefficient (ICC) was used to determine IRR with absolute score agreement and categorical determination (schools A and B).

Results/Outcomes: Student responses to the 3 exam questions at school A (N = 54), B (N = 50), and C (N = 54) were randomly selected to be scored by noncontent experts and MS4s. Agreement between the content expert and the noncontent expert scorers at school A was in the fair/good range (ICC = 0.40, 0.47) and at school C was in the excellent range (ICC = 0.74, 0.76). Similar patterns were observed when examining each question individually. Agreement between the content expert and the student scorers at school A was in the fair/good range (ICC = 0.55, 0.66) and at school C was in the excellent range (ICC = 0.80, 0.82). At school A, agreement on the categorical determination was generally better than exact agreement for the student scorers (in the upper end of the fair/good range) but did not change substantially for noncontent experts. Data from school B are currently being analyzed.

Discussion: IRR of scoring OEQs was variable among schools but trended toward good/excellent. IRR with content experts was higher when using the analytic rubric compared with the holistic rubric. Furthermore, MS4s were more reliable scorers than noncontent experts. We conclude that reliability among faculty and student scorers is achievable.

Significance: Some of the major concerns associated with OEQ-based assessment can be circumvented by scoring by noncontent experts. Each school must determine what type of rubric works best within their assessment culture. Regardless of what type of rubric is used, reliability among scorers is achievable.

Preparation for Physicians: How best to prepare students for situations of uncertainty in professional practice?

Dimitrios Papapanagou, MD, MPH, Nethra Ankam, MD, David Ebbott, and Deborah Ziring, MD

Introduction: Uncertainty abounds in the clinical environment. Medical students, however, are not explicitly prepared for situations of uncertainty in clinical practice, which can result in students’ anxiety and impact their well-being. To address this gap, we aimed to provide recommendations that would better inform a curriculum that prepares students to acknowledge and navigate uncertainty in clinical practice. We sought to: (1) capture students’ general self-efficacy and intolerance to uncertainty, (2) describe perceived comfort with uncertainty encountered during clerkships, (3) identify curricular elements that best prepare students for these situations.

Methods: This is an observational cross-sectional study of third-year medical students from an urban medical school surveyed at the end of core clerkships.