The Performance and Trajectory of Medical Students With Disabilities: Results From the Pathways Project

Lisa M. Meeks, PhD, MA, Melissa Plegue, MA, Bonnieil K. Svenor, PhD, MPH, Christopher J. Moreland, MD, MPH, Sharad Jain, MD, Christina J. Grabowski, PhD, Marjone Westervelt, PhD, MPH, Benjamin Case, MPH, William H. Eidson, EdD, Rahul Patwari, MD, Nancy R. Angoff, MD, MPH, MEd, Jack LeConche, MS, Bliss M. Temple, MD, Peter Poullos, MD, Mijiza Sanchez-Guzman, EdD, Caitlyn Coates, MS, Christine Low, MSW, Mark C. Henderson, MD, Joel Purkiss, PhD, and Michael H. Kim, MD

Purpose: The number of medical students disclosing disabilities and requesting accommodations is increasing, prompting new questions about performance of students with disabilities (SWD) on United States Medical Licensing Exam (USMLE) examination scores and to match into residency. Historical data on SWD suggest lower performance and extended time to graduation, but little is known about whether category of disability and accommodation on the Step 1 examination impact performance, time to graduation, and match rates. Our aim was to determine performance outcomes and match to residency of medical SWD by category of disability and whether they received accommodations on Step 1.

Methods: A matched-cohort design study of graduated SWD (cases) and nondisabled controls (controls) from 2 cohorts (2018 and 2019) across 11 medical schools. Controls were matched by the final Medical College Admission Test score and self-reported gender.

Results: Primary outcome measures include scores on USMLE Step 1 and Step 2 Clinical Knowledge scores, time to graduation (years), and residency matching on the first attempt (yes/no). Secondary outcomes were leaves of absence (yes/no) and matched specialty. One hundred seventy-one SWD and 341 controls were included in the study, 118 (69%) were categorized as having cognitive/learning disabilities, 48 (28%) physical/sensory disabilities, and 5 were unknown. A higher proportion of SWD took a leave of absence as compared to controls (31.8%; 95% CI: 20.4, 43.3 vs 10.7%; 95% CI: 5.0, 16.4; P < .001). Students with physical/sensory disabilities had similar on-time graduation rates (88.6%; 95% CI: 77.0, 100.0 vs 94.9%; 95% CI: 91.3, 99.9; P = .20), mean Step 1 scores (229.6 vs 233.5; P = .118), and match on first attempt (93.9%; 95% CI: 86.9, 100 vs 94.6%; 95% CI: 91.7, 97.4; P = .84) compared with controls. Students with cognitive/learning disabilities had significantly lower average Step 1 scores than controls (219.4 vs 233.4; P < .001), were less likely to graduate on time (81.2%; 95% CI: 69.2, 93.2 vs 94.9%, 95% CI: 91.3, 99.9; P = .003), and were less likely to match on first attempt (85.3%; 95% CI: 78.0, 92.7; P = .009). When compared with controls, nonaccommodated SWD had average Step 1 scores that were 12.2 points lower (95% CI: −15.9, −8.4; P < .001). Accommodated students’ scores were significantly lower but only by 6 points (B = −6.3; 95% CI: −12.3, −0.29; P = .04). Accommodated SWD had higher average scores than nonaccommodated students by 5.9 points (95% CI: −0.70, 12.5; P = .08). Students with physical/sensory disabilities had higher odds of matching into primary care compared with controls (OR = 2.11; 95% CI: 1.05, 4.26) and students with cognitive/learning disabilities (OR = 2.32; 95% CI: 1.05,5.14).

Conclusions: Barriers remain for medical students with cognitive/learning disabilities, which are partially mitigated by accommodations on board exams.

References
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 was 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.

Preparing Medical Students for Uncertainty in Clinical Practice: Recommendations for Clinical Clerkships

Dimitrios Papanagnou, 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.