

## **Effects of a Faculty Educational Intervention to Improve Quality of Written Evaluations of Pediatrics Clerkship Medical Students**

**Michelle Prong MD, Kunali Gurditta MD, Grace Ng MD, and Margarita Corredor MD**

Introduction: Written evaluations are a cornerstone of clinical assessment in undergraduate medical education; they account for the majority of a student's clerkship grade and, more importantly, students depend on written evaluations to improve their clinical performance. However, improving the quality of faculty-written narrative evaluations remains a challenge. The Narrative Evaluation Quality Instrument (NEQI) is a validated tool to assess quality of medical student narrative evaluations and includes three component arms: breadth of performance domains evaluated, specificity of comments, and usefulness to the trainee. Each component arm has a score range from 0-4; the maximum overall score is 12. The objective of this study was to determine if the quality of faculty-written narrative evaluations of medical students can be improved through trainee-delivered feedback based on NEQI scoring principles.

Methods: All Pediatric Hospital Medicine faculty at the University of Rochester Medical Center were invited to participate in this study. Three Internal Medicine-Pediatrics senior residents trained in NEQI scoring reviewed each participating faculty's three most recent, de-identified narrative evaluations of pediatrics clerkship medical students. Each evaluation was independently scored by two separate resident reviewers. Resident reviewers then completed one 30-minute, in-person feedback session with each faculty participant to review the participant's NEQI scores, current evaluation strengths, and areas for growth. Following these sessions, resident reviewers scored at least two subsequent, de-identified narrative evaluations for each faculty participant. Overall and sub-category scores were compared pre- and post-intervention.

Results: Five out of 17 eligible Pediatric Hospital Medicine faculty participated in the study. Pre-intervention total scores averaged 8.79 out of 12 points (SD=0.96, 95% CI 8.45-9.13). Pre-intervention sub-category scores averaged 3.40 for performance domains addressed, 2.80 for specificity, and 2.60 for usefulness to the trainee. Post-intervention total scores averaged 9.56 out of 12 points (SD=1.89, 95% CI 9.25-9.87). Post-intervention sub-category scores averaged 3.42 for domains addressed, 3.15 for specificity, and 3.00 for usefulness to the trainee. The results demonstrated a p-value of 0.15.

Conclusion and Discussion: There was most improvement in the "specificity of comments" and "usefulness to the trainee" component arms pre- and post-intervention. As medical education tends to encourage trainees to practice individual reflection and continual improvement through formative assessment, the gains these two component arms of evaluations are encouraging. The results are not suggestive of a strong statistical difference in total NEQI scores before and after the intervention, which may represent a type II error. There are undoubtedly factors influencing the quality of written evaluations beyond faculty skills and knowledge, including time constraints and the numerous other administrative tasks of educators.