

Original Publication

 OPEN ACCESS

Improving Interrater Reliability of Medical Student Assessment by Clinical Supervisors

Scott Moser, MD, Laura Mayans, MD, Nancy Davis, PhD*

*Corresponding author: ndavis5@kumc.edu

Citation: Moser S, Mayans L, Davis N. Improving interrater reliability of medical student assessment by clinical supervisors. *MedEdPORTAL*. 2017;13:10609. https://doi.org/10.15766/mep_2374-8265.10609

Copyright: © 2017 Moser et al. This is an open-access publication distributed under the terms of the Creative Commons Attribution-NonCommercial-Share Alike license.

Abstract

Introduction: Clinical clerkships that rely on community preceptors as faculty and assessors are challenged by the variability in their assessments of student performance. This workshop was designed to improve the interrater reliability of preceptors' assessment of student performance. **Methods:** This workshop uses a series of videos showing standardized student interactions with standardized patients and a preceptor. The videos feature a borderline satisfactory student-patient encounter and a superior student encounter. For each case, there is a video of the student providing an oral presentation of the encounter to a preceptor. The concept of a plus/delta card is also presented. Participants are asked to note positive elements of the interactions and elements needing improvement on the plus/delta card while watching the videos. Next, participants share their assessment of the students with the group using an audience response system. Group discussion allows for consensus and leads to decreased variation in preceptor assessments of each student's performance. This workshop has been delivered to preceptors in our institution as a 1-hour workshop. Additionally, a version has been presented twice at national medical teaching conferences as a faculty development workshop, instructing faculty how to facilitate the workshop. **Results:** Participant evaluations from the workshop show an increase in preceptors' awareness of specific student behaviors to observe as well as increased confidence with assessing more consistently across various student performances. **Discussion:** Using videos with standardized patients, students, and preceptors demonstrating various qualities of performance is effective in improving precision in preceptors' assessment of students in clinical settings.

Keywords

Assessment, Clinical Medicine, Preceptor

Appendices

- A. Plus-Delta Card.docx
- B. Case 1A (Ankle) Student-Preceptor Interaction.mp4
- C. Case 1B (Ankle) Student-Patient Interaction.mp4
- D. Case 2 (Nose) Student-Patient Interaction.mp4
- E. Facilitator's Guide.docx
- F. Clinical Assessment Presentation.pptx
- G. Workshop Evaluation.docx

All appendices are peer reviewed as integral parts of the Original Publication.

Educational Objectives

By the end of this session, learners will be able to:

1. Improve observational skills through reflection with other clinical instructors.
2. Deliver balanced feedback to students.
3. Develop insight through observation of others' perceptions of the same encounter.
4. Generate interrater reliability for clinical performance assessment.
5. Employ effective direct observation in clinical assessment.

Introduction

Nationally, primary care clerkships rely heavily on community volunteer preceptors to deliver much of the clinical education and therefore to assess clinical performance. These preceptor assignments allow students to experience higher continuity and autonomy than other forms of instruction. However, these benefits often come at the cost of greater variability and imprecision of student performance evaluations. Multiple studies have found high variability and low precision both within and between clerkships.^{1,4} In general, the trend appears to be toward grade inflation rather than underestimation.^{2,3} There are several theories about what has led to this trend. Some believe it is due to faculty desires for positive reviews from students. Some believe it is the result of faculty and preceptors' lack of knowledge regarding assessment

methods and standards.³ It is also believed to stem from difficulty recalling and organizing data about students as well as difficulty with separating the roles of mentor/advocate and judge/evaluator.² Kogan, Conforti, Bernabeo, Iobst, and Holmboe found that faculty often graded learners based on how they themselves would handle the clinical scenario, which may not have been in concordance with the clerkship's desired clinical performance.⁵

This is affirmed by our experience at our own institution, where the correlations of preceptor grade versus objective structured clinical examination (OSCE) score (i.e., .133) have been poor. Our OSCEs assess interpersonal communication skills, history taking, physical exam, presentation skills, counseling skills, and basic medical decision-making in the context of standardized patient encounters with associated presentation, note-writing, or EKG-interpretation stations. Furthermore, our family medicine clerkship director noted that preceptor grading for individual students tended to be straight-line satisfactory or superior. We were concerned that preceptors were not taking care in their student performance observation or were not savvy in distinguishing between satisfactory versus high-performing students. While MedEdPORTAL has several publications addressing assessment of students in OSCEs and with entrustable professional activities, we found nothing regarding the precision of preceptor observation and assessment of students in clinical settings.

Methods

The primary content of this 1-hour workshop is a series of standardized patient-student-preceptor videos with facilitated discussion to allow participants to develop more precise skills in their assessment of student clinical performance. The goal for medical educators who train community preceptors is to evaluate this resource for adoption or adaptation in their own preceptor development efforts.

We selected three family medicine faculty to develop standardized cases focused around two important decision points for clinical supervisors, namely, the pass versus fail cut-point and the high-satisfactory versus superior cut-point. The cases were written and edited with the goal of presenting credible standardized patient cases with realistic student performances. The most critical issue for these standardized cases is the pass versus fail cut-point, a decision community volunteer preceptors are loath to make. The high-satisfactory versus superior cut-point is less critical from a competency-based standpoint but is important to students who are trying to distinguish themselves academically and to clerkship directors who are concerned about grade inflation. The cases were developed and vetted by experienced clinical faculty members, and our study showed the cases worked well at generating reflection and discussion among clinical teachers around the intended decision points.

During the workshop, participants are instructed to use a plus/delta card (Appendix A) for formative assessment (i.e., feedback), a tried and true technique for evaluation in a medical setting.⁶ Positive observations are listed on the plus side of the card and things that need improvement on the delta side of the card. For summative assessment, facilitators utilize the clinical assessment form from their own institution in order to familiarize preceptors and encourage reduced variation in the assessments.

A PowerPoint presentation is used to introduce the workshop, topic, and facilitators; to provide learning objectives; and to guide a brief discussion of formative versus summative assessment. The PowerPoint is located in Appendix F. The Facilitator's Guide (Appendix E) provides step-by-step instructions for the workshop, including references to individual slides in the presentation.

The first case is that of a patient presenting for an ankle/midfoot injury and is administered using two videos. The first video (Appendix B) depicts a medical student presenting the patient to her preceptor after she has evaluated the patient. The presentation is cursory and suspect because of gaps in student performance that cannot be confirmed due to missing direct observation. After facilitated discussion of formative and summative assessment of this student, the second video (Appendix C) is shown. This video shows the student's actual encounter with the patient, demonstrating performance at the pass versus fail cut-point. Discussion is resumed about formative and summative assessment of the student. By

presenting the videos in this order, participants are given the opportunity to reevaluate their decisions from the first video, highlighting the importance of direct observation of students with patients.

The second case (Appendix D) involves a patient presenting with nasal congestive symptoms. For this case, only the student-standardized patient encounter in which the student demonstrates skills at the high-satisfactory versus superior cut-point is presented. The facilitator directs the ensuing discussion to address issues of what constitutes superior performance.

The workshop has been presented multiple times to community preceptors in a 1-hour time frame with a brief overview of assessment theory and strategies before introducing the performance observation skills.

Results

This workshop has been delivered four times to family medicine preceptors and other faculty in both local (Kansas: $N = 120$) and national (Society of Teachers of Family Medicine: $N = 35$) settings. Drs. Moser and Mayans have expertise in preceptor training, and Dr. Moser has used the plus/delta card technique extensively over 15 years.

Workshop participants have been very interactive and eager to weigh in with their perceptions of the video cases, including their experiences with similar students in clinical encounters. Workshop evaluation (Appendix G) data are strong, with every metric scoring between 3.4 and 3.8 on a 4-point scale (4 = *Strongly Agree*). When participants were asked what they would do as a result of the workshop, their comments included the following:

- “Differentiate communication skills from clinical skills.”
- “Use plus/delta card.” [This was listed numerous times.]
- “Give feedback as the time progresses and evaluate for improvement based on feedback.”
- “More observation with student/patient in the exam room.”
- “Document intermittently so that I can give appropriate feedback.”

While the workshop evaluation results have been very positive, we have used the feedback to adjust our discussion points. Such improvements included asking audience members for more of their own experiences, as well as clarifying use of the plus/delta card, which many preceptors indicated they would like to implement in their practices.

Discussion

Limitations

While evaluation of the workshop has been favorable, community preceptor participation has been limited. There is some concern that preceptors most in need of the training do not avail themselves of it.

The workshop has been conducted at our own institution and twice at national conferences. There has been interest at the national level in creating a consortium of institutions to share resources and best practices in preceptor assessments of students.

Perceived improved competence on the part of the workshop participants has not been evaluated in practice. One area of future study would be analyzing preceptors’ assessment of students over time to measure impact on variation in reporting.

Reflection

We are encouraged by the end-of-workshop improvement in interrater reliability in a simulated environment, as well as by the anecdotal improved confidence of our preceptor participants. Our goal is to train a critical mass of our primary care preceptors in order to improve the quality of assessment regarding students’ performance in the community clinical setting. We have posted the videos and workshop materials on a private, internal website in order for the workshop to be presented by clerkship directors in the community rather than requiring remote preceptors to come to the medical school or a regional conference for the training.

While the workshop was developed for medical student assessment in primary care, it potentially could be modified to meet the needs of subspecialties or even residents in clinical settings. A key to skill building in this area is for faculty to actually take time to observe learners with patients. As our videos point out, often what the learner reports to the preceptor is not necessarily what occurred in the actual patient encounter.

Future Directions

Following national presentation, we received feedback that others would like more specifics of how we run the intervention. As stated above, we have developed an online tutorial on an internal website that includes the videos of standardized patients, students, and preceptors.

We have refined the workshop content through presentations to a variety of audiences, including family medicine preceptors and full-time faculty and resident physicians in numerous specialties. Most participants have expressed important aha moments as they interact with peers regarding their observations and evaluations of standardized patient-student-preceptor interactions.

Scott Moser, MD: Professor, Department of Family and Community Medicine, University of Kansas School of Medicine–Wichita; Assistant Dean of Curriculum, University of Kansas School of Medicine–Wichita

Laura Mayans, MD: Assistant Professor, Department of Family and Community Medicine, University of Kansas School of Medicine–Wichita; Clerkship Director, Department of Family and Community Medicine, University of Kansas School of Medicine–Wichita

Nancy Davis, PhD: Associate Professor, Department of Family and Community Medicine, University of Kansas School of Medicine; Assistant Dean of Faculty Affairs and Development, University of Kansas School of Medicine

Disclosures

None to report.

Funding/Support

None to report.

Informed Consent

All identifiable persons in this resource have granted their permission.

Ethical Approval

Reported as not applicable.

References

1. Alexander EK, Osman NY, Walling JL, Mitchell VG. Variation and imprecision of clerkship grading in U.S. medical schools. *Acad Med.* 2012;87(8):1070-1076. <https://doi.org/10.1097/ACM.0b013e31825d0a2a>
 2. Ogden PE, Edwards J, Howell M, Via R, Song J. The effect of two different faculty development interventions on third-year clerkship performance evaluations. *Fam Med.* 2008;40(5):333-338.
 3. Weaver CS, Humbert AJ, Besinger BR, Graber JA, Brizendine EJ. A more explicit grading scale decreases grade inflation in a clinical clerkship. *Acad Emerg Med.* 2007;14(3):283-286. <https://doi.org/10.1197/j.aem.2006.09.055>
 4. Plymale MA, French J, Donnelly MB, Iocono J, Pulito AR. Variation in faculty evaluations of clerkship students attributable to surgical service. *J Surg Educ.* 2010;67(3):179-183. <https://doi.org/10.1016/j.jsurg.2010.03.003>
 5. Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: a conceptual model. *Med Educ.* 2011;45(10):1048-1060. <https://doi.org/10.1111/j.1365-2923.2011.04025.x>
 6. Qualters DM. Observing students in a clinical setting. *Fam Med.* 1999;31(7):461-462.
-

Received: February 10, 2017 | **Accepted:** July 3, 2017 | **Published:** July 26, 2017