

Original Publication

 OPEN ACCESS

Evaluating Interprofessional Team Performance: A Faculty Rater Tool

Christopher P. Forest, PA-C*, Désirée A. Lie, MD, MSED, Sae Byul (Sarah) Ma, PharmD

*Corresponding author: cforest@usc.edu

Citation: Forest CP, Lie DA, Ma S. Evaluating Interprofessional Team Performance: A Faculty Rater Tool. *MedEdPORTAL*. 2016;12:10447. https://doi.org/10.15766/mep_2374-8265.10447

Copyright: © 2016 Forest et al. This is an open-access publication distributed under the terms of the Creative Commons Attribution-NonCommercial-Share Alike license.

Abstract

Introduction: Reliable team assessment has become a priority because of growing emphasis on interprofessional education and team-based care. Objective rating scales are needed to evaluate interprofessional student teams and individuals and provide real-time feedback. **Methods:** In response to a need for behavioral rating scales, we modified the McMaster-Ottawa Scale from a 9-point to a 3-point scale and added descriptive behavioral anchors to define three levels of competency (i.e., below, at, and above expected). This modification is intended to provide consistent rating of individuals and teams in patient settings. We then developed a demonstration video using actors representing four professions to demonstrate the three levels of performance within the team. Our faculty rater tool, consisting of the modified scale and video, is designed to provide standardized ratings in interprofessional educational settings that involve patient care. **Results:** We conducted training sessions with 40 faculty members from seven professions (medicine, dentistry, occupational therapy, nursing, pharmacy, physician assistant, and psychology) over a 2-year period. Immediately after each training session, two trained faculty observers rated interprofessional student teams as they conducted history and assessments on standardized patients. Observer scores were compared with one another and with standard expert ratings of the same teams. Trained observer ratings were consistent across the pairs. The observer training can be conducted within 60-90 minutes with the tool. **Discussion:** Results of our implementation of the faculty rater tool confirm that the modified McMaster-Ottawa Scale is feasible to administer in clinical settings and that the demonstration video can be easily adopted for standardizing observer ratings.

Keywords

Interprofessional Education, Rating Scale, Interprofessional Relations, Faculty Training, Team Assessment, McMaster-Ottawa, Behavior Anchors

Appendices

- A. Modified McMaster-Ottawa Scale-Individuals with Behavioral Anchors.docx
- B. Modified McMaster-Ottawa Scale-Teams with Behavioral Anchors.docx
- C. Interprofessional Student Team Video.mp4
- D. Faculty Rater Training Session Slides.pptx
- E. Moderator Guide for Conducting Observer Training.docx

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

Educational Objectives

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

1. Describe six components of an observer rating scale (the modified McMaster-Ottawa Scale) for assessing team and individual performance within an interprofessional clinical team.
2. Review behavioral anchors for the 3-point rating scale for assessing teams and individuals.
3. Standardize scores after viewing and discussing a demonstration video of team-based care.
4. Evaluate interprofessional teams and individuals in practice or in simulated patient settings.

Introduction

Interprofessional education (IPE) is now a strong recommendation or requirement among accrediting bodies of many health professions.¹ IPE is considered a foundation for future team-based care, which has been linked to improved patient and health care quality outcomes.^{2,3} Assessing the performance of teams and the individuals within teams to provide guidance for improved team functioning is an important component of IPE. Although a plethora of measurement tools purport to evaluate team behaviors or performance,⁴ many in fact measure attitudes toward or knowledge about interprofessional care. Most do not rely on objective ratings by observers or provide anchors that define team behaviors. Recent

systematic reviews highlight the paucity of observer rating tools suited for providing feedback to individuals working within teams. Therefore, there is a need for reliable and easily administered observer rating scales designed for clinical settings.^{5,6} This gap can be addressed by identifying and implementing a rating scale to assess team performance based on observable and modifiable behaviors. Such a scale would allow effective feedback to student teams in simulated and actual clinical settings. The McMaster-Ottawa Scale⁷ was designed for this purpose, as it was intended to assess team behaviors around simulated patients using paper cases, videos, and standardized patients.⁸ It allows observers to assess student and interprofessional team performance for six competencies cited as essential to interprofessional teamwork.⁹⁻¹¹ The six competencies are communication, collaboration, roles and responsibilities, collaborative patient-family centered approach, conflict management and resolution, and teamwork/team functioning. There is also a seventh, global score for individual and team performance. There are two versions of the scale with the same six competencies: one version for assessing individuals and one for assessing the team. However, the original scale with a 9-point score range for each competency presents a challenge for standardizing observer ratings, as it can be difficult to distinguish among nine levels of behavior.

We modified the McMaster-Ottawa Scale to enhance its applicability in practice, reducing the score range from 9 to 3 points for each competency. We defined the 3 points as *below*, *at*, and *above* expected (corresponding to levels 1, 2, and 3 respectively), to reflect the achievement of competencies. We then added anchors that described observable behaviors for each level of each competency. Once the scale was modified, we created a demonstration video and tested the modified scale (our faculty rater tool) and video among faculty who attended training sessions and who then rated actual student teams immediately after the training sessions. Our overall goal was to develop and implement a tool for rating teams and individuals that can be used to provide feedback to interprofessional student teams in clinical settings.

Methods

We adopted the plan-do-study-act¹² approach in this project. We first conducted a literature review to identify an existing scale to meet our goal to develop a tool for rating teams and individuals that could be used to provide feedback to interprofessional student teams in clinical settings. Planning included consulting with the selected scale's developers and then developing a demonstration video. For the doing/execution phase, we made modifications and implemented the tool with a group of faculty. We then studied the impact of the tool on faculty ratings of actual interprofessional teams after the training. We next acted on the findings, revised the tool further, and implemented the revised tool with a second group of faculty. Feedback from faculty at the second session enabled us to develop the moderator guide to share with others.

As mentioned above, we first conducted a literature review and contacted authors of potentially useful scales. We found that the McMaster-Ottawa Scale best met the need to assess six core competencies of interprofessional teamwork. Once the McMaster-Ottawa Scale was identified, a core group of expert IPE educators representing the fields of medicine, pharmacy, physician assistant, occupational therapy, nursing, and psychology reviewed the scale and made recommendations. The group followed an iterative process based on literature review, faculty expert opinion, and consensus building. The group modified the scale from 9 to 3 points. The 3 points were defined as *below*, *at*, and *above* expected, corresponding to levels 1, 2 and 3 respectively, reflecting the assessment of the six competencies. Anchors were added to define observable behaviors for each level of each competency. We preserved the two original formats of the scale, creating one for rating individuals and one for rating the team (Appendices A & B, respectively). For these modified scales, the behavioral anchors are included after the scale items within the appendices for easy reference.

We tested feasibility and utility of the modified scale for rating interprofessional students and teams over a period of 2 years.¹³ During this period, we trained 40 faculty from seven professions to apply the modified scale to the rating of student teams. The seven professions represented medicine, pharmacy, physician assistant, dentistry, psychology, nursing, and occupational therapy.

We developed a 14-minute demonstration video (Appendix C) in order to provide faculty with practice assessing a team and to standardize faculty ratings before conducting ratings on actual teams. In this video, a team of students played by actors performed at three different levels (1, 2, and 3) during a huddle followed by a patient encounter. The performance level of the team was planned as level 2 (*at expected*). Students in the video represented four health professions (nursing, medical, pharmacy, and physician assistant). The team interacted with an actor portraying a patient with diabetes. Prior to the patient encounter, the actor team was provided with instructions asking that “the team should first huddle, then take a history and come up with a care plan to present to an attending faculty.” Five minutes were assigned for the huddle and 15 minutes for the actual patient encounter. In the demonstration video, the team spent 3 minutes in a huddle and 11 minutes with the actor patient. The 14-minute video is followed by a 4-minute video of a team functioning at *above expected* (level 3).

We invited 16 faculty observers to participate in a training session in order to test the demonstration video and modified scale (training session 1). PowerPoint slides were used to guide the training sessions (Appendix D). This first training session lasted 45 minutes. After the training session, the faculty went on to evaluate two interprofessional student teams in a simulated patient setting. The students in the teams represented the professions of medicine/physician assistant, nursing, pharmacy, and occupational therapy. The faculty rated each student team in pairs. Each faculty pair was allotted 20 minutes to observe each student team. The paired faculty ratings were then compared.

Based on our findings from training session 1, we modified the tool and conducted an additional training session. We tested the revised tool on two smaller groups of eight faculty and one large group of 16 faculty (training session 2). Training session 2 was 60 minutes in duration, and faculty were provided with the modified scale and video 5 days ahead of the session by email. We did not require faculty to review the materials ahead of time.

Finally, we developed the moderator guide (Appendix E) to allow others to conduct the training session using our tool. The moderator guide was based on faculty participants’ feedback and our experience as trainers in conducting sessions 1 and 2.

Results

A total of 40 faculty experienced our training sessions over a 2-year period. We asked the faculty from sessions 1 ($n = 16$) and 2 ($n = 16$) to complete postsession evaluations. The 32 faculty represented the health professions of medicine, pharmacy, physician assistant, dentistry, psychology, nursing, and occupational therapy. All 32 completed the postsession evaluation (100% response rate). The evaluation consisted of survey questions with 5-point Likert scale responses (from 1 = *strongly agree* to 5 = *strongly disagree*). We also invited faculty to comment on the session and the tool.

For the survey questions, the number and percentage of faculty who *agreed* or *strongly agreed* with the statements below are presented:

- The behavioral descriptions provided on the modified McMaster-Ottawa Scale are helpful for me to rate students and team ($n = 27, 84\%$).
- I am confident about my scores for the team ($n = 26, 81\%$).
- I am confident about my scores for the individual team members ($n = 21, 66\%$).
- I can accurately rate 4 students within the allotted observation time ($n = 32, 100\%$).

Faculty comments included the following:

- “The scale worked well for me.”
- “The scale was easy to use.”
- “The overall flow of the session activities went well.”
- “I appreciate the descriptions of behaviors for scoring.”
- “The duration of training is appropriate.”
- “The anchors for each item were clear and I was able to apply them to rating the student teams.”
- “The training was helpful for standardization purposes.”

- “Reviewing the scale during training and watching the video was very helpful—it enabled me to use the rating scale correctly to assess teams.”

We strived to limit the duration of the training session because it was a logistical challenge for faculty from different professions to take time to attend. However, the training duration of 45 minutes was an issue raised by faculty after session 1. Some faculty commented that the training was “too short,” and four of 16 faculty from session 1 commented in their evaluations that they would benefit from a longer duration for the session. Some faculty also verbally requested that they receive the training materials ahead of time. In response to this session 1 feedback, we increased the duration of training in session 2 from 45 to 60 minutes and emailed faculty the scale and a link to the demonstration video 5 days prior to the session. At session 2, we noted that not all faculty had reviewed the materials sent to them ahead of the session. After session 2, two faculty expressed that they “would like more preparation to increase confidence in rating,” explaining that “the example in the video was easy to score, but with the students in an actual practice scenario it is more difficult.” This feedback suggests that the optimal time for training is likely to be between 60 and 90 minutes to allow additional time for discussion among faculty. Where possible, we also recommend that faculty rate actual student teams in pairs before they rate teams on their own, to increase their confidence as objective raters.

Other suggestions from session 1 that we incorporated successfully in session 2 were “include anchors as part of the rating scale instead of as a separate document” and “we need more time to practice evaluating teams on the video.” As trainers for the sessions, we observed that faculty were in agreement, in both sessions, about the *above*/high-performing (level 3) student in the demonstration video but had a tendency to rate the *below*/low-performing (level 1) student at level 2 for some competencies, a level higher than the student’s behavior reflects. It took some group discussion to reach consensus about the scoring for the level 1 student.

As stated previously, following the training sessions, pairs of faculty from different professions were assigned to observe and rate interprofessional student teams in a simulated patient setting. Their ratings were compared. Analysis of these ratings of students and the team¹³ confirms consistency between faculty in rating both individuals and teams after their training. Overall, we found that the faculty demonstrated greater agreement when rating team performance versus rating individual performance.

Our postsession data from rating of actual student teams affirmed the observation of faculty reluctance to score students at level 1 even when students performed at level 1. Faculty thus were more likely to demonstrate rater error when rating low-performing (vs. high-performing) students, as reported in other studies.^{14,15} We therefore recommend that more time and effort be directed toward assessing level 1 behaviors during faculty training. Since level 1 behaviors occur less frequently in practice and are thus harder to find in practice, the demonstration video we produced has the added value of showing level 1 behaviors in at least one student among the team of four student actors. After the experience of rating two actual student teams, the faculty also reported that the time of 20 minutes we allotted for observing each team was adequate for rating four students and the team. We found no evidence of a halo effect related to faculty observers rating students from their own profession compared with students from other professions, but the number of faculty and students in our project was too small to allow statistical testing.

Discussion

The original McMaster-Ottawa Scale had been demonstrated to be valid and applicable to simulated paper and video cases but had not been tested in clinical settings, and observer training for applying the scale to clinical student teams had not been reported. We selected the scale for its design since the scale allowed rating of individuals and teams on the same six competencies. We adapted the scale to enhance and broaden its application. By modifying the scale from 9 to 3 points for each competency, we were able to develop distinct behavioral anchors for each competency. This, in turn, allowed us to develop a training tool consisting of the modified scale and the demonstration video for faculty to practice assessing individuals and teams. We tested this tool on 40 faculty from seven professions over a period of 2 years. We found the tool to be practical, feasible to implement, and easy to administer. We also found that

training was associated with high agreement between faculty who rated interprofessional student teams immediately after training.

The key advantages of the tool are the following: (1) It assesses team behaviors independent of clinical knowledge and skills, (2) ratings can be standardized with training, (3) feedback can be given based on observable and modifiable behaviors, (4) it can be used to train individual or groups of faculty, (5) it can be adopted for use by any health profession, (6) there is potential to detect change in levels of behaviors for each competency over time or after curricular interventions, and (7) interprofessional competencies recommended by multiple international organizations are addressed.¹⁶

The outstanding challenge for us and, we anticipate, for others who aim to train faculty to rate teams and students is deciding on the optimal duration of training. We therefore designed our PowerPoint slides and the moderator guide to allow flexibility in the planned duration. Because we found that faculty time available for training was limited, we opted to train faculty in shorter time frames of 45 and 60 minutes for sessions 1 and 2, respectively. Our observations as trainers and feedback from faculty suggest that even 60 minutes is inadequate for larger groups of more than five faculty and that greater confidence for rating can be achieved with a longer 90-minute session when training larger groups. Although we shared the training materials with faculty in advance of training session 2, fewer than half of the faculty reported reviewing these materials prior to the session. This low rate of pre-session review likely reflects typical time constraints of faculty, so in our project, we were unable to assess the impact that pre-session review had on training outcomes.

An important challenge faced by IPE educators is that once trained, faculty observers need to find time to observe teams interacting for 20 minutes. As with precepting individual students in clinic, the team observation requires time and can compete with patient care. However, observing and providing feedback to teams can play a large role in the formation of future collaborative practice teams.¹⁷ We anticipate that as IPE gains a higher priority among health care professions and more faculty are trained, the practice of team observation will be integrated into usual clinical education and patient care, and IPE faculty will be given time and opportunity to rate and provide feedback to student teams.

Our project has some limitations. We trained a relatively small number of faculty and were unable to perform statistical reliability tests to ascertain true interobserver reliability. We expect the number of trained faculty to increase as other institutions adopt our tool for use in IPE settings. Multi-institutional projects would be helpful to increase the numbers of faculty in future projects. In our project, we did not focus on the effect of giving feedback to the team. This aspect of team evaluation remains to be studied and is currently a hot topic in IPE evaluation. Our findings have been presented at several national and international meetings and have been published.¹³ Our paper is posted on the Nexus site¹⁸ and disseminated among other IPE measurement tools. We anticipate that the tool consisting of the modified scale with the demonstration video will be useful to many professions seeking to improve evaluation of their IPE curricula.

We plan to offer additional faculty training using this tool in a variety of clinical teaching settings including student-run clinics, geriatric assessment centers, and a mobile clinic where interprofessional student teams function under faculty supervision. With each training session, the plan-do-study-act approach will be applied with each iteration. We plan to develop specific guidelines for giving effective feedback to teams based on their performance.¹⁷

In summary, our project and findings support the concept that observer training is an essential component of improving faculty confidence and accuracy in rating teams and the individual students within the teams. We conclude that when presented with a demonstration video, the modified McMaster-Ottawa Scale is useful and easy to administer for assessing individuals and teams in clinical IPE settings.

Christopher P. Forest, PA-C: Assistant Professor of Clinical Family Medicine, Department of Family Medicine and Director of Research, Division of Physician Assistant Studies, Keck School of Medicine of the University of Southern California

Désirée A. Lie, MD, MSED: Clinical Professor of Family Medicine, Department of Family Medicine, Keck School of Medicine of the University of Southern California

Sae Byul (Sarah) Ma, PharmD: Assistant Professor, Keck School of Medicine and the School of Pharmacy of the University of Southern California

Disclosures

None to report.

Funding/Support

None to report.

Prior Presentations

Findings have been presented at the following conferences: Learn Serve Lead 2015: The AAMC Annual Meeting, November 6-10, 2015; Collaborating Across Borders V (CAB V), September 29-October 2, 2015; 2015 Innovations in Medical Education Conference (IME), February 21-22, 2015; and Physician Assistant Education Association Education Forum, October 15-19, 2014.

Ethical Approval

This publication contains data obtained from human subjects and received ethical approval.

References

1. Zorek J, Raehl C. Interprofessional education accreditation standards in the USA: a comparative analysis. *J Interprof Care*. 2013;27(2):123-130. <http://dx.doi.org/10.3109/13561820.2012.718295>
2. Reeves S, Zwarenstein M, Goldman J, et al. The effectiveness of interprofessional education: key findings from a new systematic review. *J Interprof Care*. 2010;24(3):230-241. <http://dx.doi.org/10.3109/13561820903163405>
3. Zwarenstein M, Goldman J, Reeves S. Interprofessional collaboration: effects of practice-based interventions on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*. 2009;3:CD000072. <http://dx.doi.org/10.1002/14651858.CD000072.pub2>
4. Barr H, Low H. The definition and principles of interprofessional education. Centre for the Advancement of Interprofessional Education (CAIPE) Web site. <http://caipe.org.uk/about-us/the-definition-and-principles-of-interprofessional-education> Published January 2011. Accessed May 26, 2016.
5. Havyer RD, Nelson DR, Wingo MT, et al. Addressing the interprofessional collaboration competencies of the Association of American Medical Colleges: a systematic review of assessment instruments in undergraduate medical education. *Acad Med*. 2016;91(6):865-888. <http://dx.doi.org/10.1097/ACM.0000000000001053>
6. Havyer RD, Wingo MT, Comfere NI, et al. Teamwork assessment in internal medicine: a systematic review of validity evidence and outcomes. *J Gen Intern Med*. 2014;29(6):894-910. <http://dx.doi.org/10.1007/s11606-013-2686-8>
7. McMaster/Ottawa TOSCE (team observed structured clinical encounter) toolkit. McMaster University Web site. <http://fhs.mcmaster.ca/tosce/en>. Published 2010. Accessed May 26, 2016.
8. TOSCE stations. McMaster University Web site. http://fhs.mcmaster.ca/tosce/en/tosce_stations.html. Published 2010. Accessed May 26, 2016.
9. A national interprofessional competency framework. Canadian Interprofessional Health Collaborative Web site. http://www.cihc.ca/files/CIHC_IPCompetencies_Feb1210.pdf. Published February 2010. Accessed May 26, 2016.
10. Health Professions Network Nursing and Midwifery Office. Framework for action on interprofessional education & collaborative practice. World Health Organization Web site. http://www.who.int/hrh/resources/framework_action/en. Published 2010. Accessed May 26, 2016.
11. Interprofessional Education Collaborative Expert Panel. *Core Competencies for Interprofessional Collaborative Practice: Report of an Expert Panel*. Washington, DC: Interprofessional Education Collaborative; 2011.
12. Quality and service improvement tools: plan-do-study-act (PDSA). NHS Institute for Innovation and Improvement Web Site. http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service_improvement_tools/plan_do_study_act.html. Published 2008. Accessed May 26, 2016.
13. Lie D, May W, Richter-Lagha R, Forest C, Banzali Y, Loheny K. Adapting the McMaster-Ottawa Scale and developing behavioral anchors for assessing performance in an interprofessional team observed structured clinical encounter. *Med Educ Online*. 2015;20:26691. <http://dx.doi.org/10.3402/meo.v20.26691>
14. Iramaneerat C, Yudkowsky R. Rater errors in a clinical skills assessment of medical students. *Eval Health Prof*. 2007;30(3):266-283. <http://dx.doi.org/10.1177/0163278707304040>
15. McManus IC, Thompson M, Mollon J. Assessment of examiner leniency and stringency ("hawk-dove effect") in the

MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Med Educ.* 2006;6:42.
<http://dx.doi.org/10.1186/1472-6920-6-42>

16. Thistlethwaite JE, Forman D, Matthews LR, Rogers GD, Steketee C, Yassine T. Competencies and frameworks in interprofessional education: a comparative analysis. *Acad Med.* 2014;89(6):869-875.
<http://dx.doi.org/10.1097/ACM.0000000000000249>
 17. Lie DA, Forest CP, Kysh L, Sinclair L. Interprofessional education and practice guide No. 5: interprofessional teaching for prequalification students in clinical settings. *J Interprof Care.* 2016;30(3):324-330.
<http://dx.doi.org/10.3109/13561820.2016.1141752>
 18. Assessment and evaluation. National Center for Interprofessional Practice and Education Web site.
<https://nexusipe.org/advancing/assessment-evaluation>. Accessed May 26, 2016.
-

Received: May 28, 2016 | **Accepted:** August 4, 2016 | **Published:** August 26, 2016