

Original Publication

 OPEN ACCESS

Milestone-Based Tool for Learner Evaluation of Faculty Clinical Teaching

Karyn Kassis, MD, MPH*, Rebecca Wallihan, MD, Larry Hurtubise, Sara Goode, Margaret Chase, MD, John D. Mahan, MD

*Corresponding author: karyn.kassis@nationwidechildrens.org

Citation: Kassis K, Wallihan R, Hurtubise L, Goode S, Chase M, Mahan JD. Milestone-based tool for learner evaluation of faculty clinical teaching. *MedEdPORTAL*. 2017;13:10626. https://doi.org/10.15766/mep_2374-8265.10626

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

Abstract

Introduction: Traditional normative Likert-type evaluations of faculty teaching have several drawbacks, including lack of granular feedback, potential for inflation, and the halo effect. To provide more meaningful data to faculty on their teaching skills and encourage educator self-reflection and skill development, we designed and implemented a milestone-based faculty clinical teaching evaluation tool. **Methods:** The evaluation tool contains 10 questions that assess clinical teaching skills with descriptive milestone behavior anchors. Nine of these items are based on the Stanford Faculty Development Clinical Teaching Model and annual Accreditation Council for Graduate Medical Education (ACGME) resident survey questions; the tenth was developed to address professionalism at our institution. The tool was developed with input from residency program leaders, residents, and the faculty development committee and piloted with graduate medical education learners before implementation. **Results:** More than 7,200 faculty evaluations by learners and 550 faculty self-evaluations have been collected. Learners found the form easy to use and preferred it to previous Likert-based evaluations. Over the 2 years that faculty self-evaluations have been collected, their scores have been similar to the learner evaluation scores. The feedback provided faculty with more meaningful data on teaching skills and opportunities for reflection and skill improvement and was used in constructing faculty teaching skills programs at the institutional level. **Discussion:** This innovation provides an opportunity to give faculty members more meaningful teaching evaluations and feedback. It should be easy for other institutions and programs to implement. It leverages a familiar milestone construct and incorporates important ACGME annual resident survey information.

Keywords

Faculty Development, Self-Assessment, Faculty Teaching Evaluation, Milestone Assessment

Educational Objectives

By using this evaluation tool, learners will be able to:

1. Assess faculty clinical teaching performance with improved specificity of feedback and alignment of assessments with the Accreditation Council for Graduate Medical Education annual resident survey.
2. Assess faculty self-awareness of their own skill level to guide them in targeted teaching skill acquisition and improvement.
3. Assess institutional faculty teaching competency when self-evaluations and learner evaluation data are combined to guide programming for faculty development.

Introduction

Evaluation of faculty teaching performance plays an important role in academic medical centers, helping faculty identify areas of strength and weakness in their educator skills. These data are often used as a metric for advancement, salary increases, or bonuses. Traditional assessments of faculty teaching rely on evaluations utilizing normative Likert-type scales. While previous studies of normative resident and student evaluations of faculty have shown internal consistency and reliability, the utility of these methods in

Appendices

- A. Faculty Clinical Teaching Milestones.docx
- B. Faculty and Learner Introduction Presentation .ppt
- C. Faculty and Learner Handout.docx
- D. Promotion and Tenure Letter.docx

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

improving faculty performance has been variable.¹ Learners may give inflated ratings of all faculty, making it difficult to distinguish proficient from less skilled teachers, or may be susceptible to the halo effect, assigning similar ratings across dimensions rather than distinguishing between areas of performance.^{2,3}

Faculty members, educational leaders, and faculty developers externally and at our institution have been dissatisfied with the limited relevance and effectiveness of the data collected by traditional normative Likert-type evaluations of faculty teaching.⁴ Out of this dissatisfaction and after positive experiences with milestone-based assessments of resident performance, we explored a method of more detailed evaluation of faculty by learners. Using a well-accepted faculty development construct (the Stanford Faculty Development Clinical Teaching Model) as a framework, we developed milestone-based anchors for 10 separate items relevant to clinical teaching.⁵

In 2013, the Accreditation Council for Graduate Medical Education (ACGME) introduced milestones for assessment of residents and fellows. These milestones utilize behavioral language to provide descriptive anchors and can offer feedback to the learner on areas for improvement, along with skills and behaviors necessary to reach the next level. Milestone-based evaluations of learners also have the potential to improve accuracy of assessments and provide more standardization than Likert-based evaluations.

Although milestones have been used for evaluation of faculty teaching in general surgery with encouraging results, limited data on this method exist in other disciplines.^{6,7} At the outset, we proposed that this tool would provide more granularity for assessment of faculty teaching performance as evaluated by pediatric and medicine-pediatric residents and fellows. We also proposed that these evaluations would align more closely with ACGME annual resident and fellows survey questions around faculty teaching and thereby offer additional insights into these teaching behaviors of faculty. Finally, we proposed that using the tool for self-assessment would stimulate reflective practice and guide both faculty individual development plans and the selection of faculty development topics at our institution.

Methods

Content Development

The Nationwide Children’s Hospital Faculty Clinical Teaching Milestones (NCH-FCTM) are composed of 10 areas of assessment. Six of the areas are based on the Stanford Faculty Development Clinical Teaching Model. Three additional areas are devoted to common important educator themes in clinical medicine identified by ACGME annual resident and fellows surveys, and a final area assesses professionalism (Appendix A; Table 1).

Table 1. Assessment Domains Mapped to Specific Models

Clinical Teaching Assessment Domain	Model From Which Content Is Derived
Milestone 1: Establishes Positive Learning Climate	Stanford
Milestone 2: Maintains Control of Educational Session	Stanford
Milestone 3: Establishes Learning Goals	Stanford/ACGME
Milestone 4: Promotes Understanding and Retention of Knowledge and Skills	Stanford/ACGME
Milestone 5: Provides Formative Feedback	Stanford/ACGME
Milestone 6: Promotes Clinical Reasoning	ACGME
Milestone 7: Promotes Evidence Based Medicine	ACGME
Milestone 8: Promotes Self-Directed Learning in Learners	Stanford
Milestone 9: Balances Supervision and Autonomy	ACGME
Milestone 10: Displays Professionalism	NCH-FCTM

Abbreviations: ACGME, Accreditation Council for Graduate Medical Education annual resident and fellows survey; NCH-FCTM, Nationwide Children’s Hospital Faculty Clinical Teaching Milestones; Stanford, Stanford Faculty Development Clinical Teaching Model.

The Stanford model is organized into seven key topics useful in developing excellence in clinical teaching, including learning climate, control of session, communication of goals, promotion of understanding and retention, evaluation, feedback, and promotion of self-directed learning.⁵ We chose the Stanford model as

it has been previously validated, has been widely used for faculty development of clinical teaching excellence, and has complementary overlap with the ACGME survey.^{8,9} Over 380 faculty members at 155 institutions have been trained facilitators in the Stanford model over the past 30 years.¹⁰ Six of the seven Stanford model topics were included in the NCH-FCTM, with the exception of the evaluation topic. While designing and implementing appropriate evaluation methods are important skills for educators, this topic seemed less appropriate for learners to assess in their clinical teaching faculty. NCH-FCTM milestones 1-5 and 8 are based on the Stanford model.

The ACGME identified resident and faculty surveys as high-value data for program evaluation and improvement.¹¹ Reviewing themes used by the ACGME in its resident surveys helped to identify additional areas for evaluation of faculty clinical teaching. The following domains of the 2016 ACGME survey are included in our model: provide appropriate level of supervision to learners (autonomy and supervision), create an environment of inquiry (learning environment), provide goals and objectives for assignments (learning goals), satisfy with feedback after assignments (giving effective feedback), and provide sufficient instruction (promoting clinical reasoning and teaching using evidence-based medicine). These domains are represented in NCH-FCTM milestones 3-7 and 9. Milestone 10 reflects an NCH institutional priority of assessing faculty professionalism by learners.

The NCH-FCTM was developed by a committee of program directors and faculty development leaders at NCH followed by four cycles of review and revision from other program directors, faculty educators, chief residents, and trainees in the institution. Final revisions occurred after a pilot including 29 resident raters. Faculty evaluations from inpatient rotations were collected using the NCH-FCTM between July and September of 2015 and contrasted with standard inpatient faculty evaluations collected between July and September of 2014 that were organized on a normative Likert-based system. The coefficient of variation and coefficient of dispersion were calculated to compare variability between assessment systems, with a higher coefficient indicating more variability.

Implementation

After design and revisions of the NCH-FCTM evaluation form, it was embedded in our electronic learning management system and distributed to residents and fellows to evaluate faculty teaching during inpatient and outpatient clinical rotations. After the pilot testing, the NCH-FCTM was implemented across residency and fellowship programs in conjunction with the GME Office at NCH. GME training at NCH comprises 66 residency and fellowship training programs. Of those 66 programs, 29 are accredited by ACGME, five are accredited by other external accrediting bodies, 24 are nonaccredited, and eight are joint programs run in cooperation with The Ohio State University. We have approximately 300 trainees split nearly evenly between residents and fellows. NCH is affiliated with 71 outside institutions and serves as a training site for approximately 585 rotating residents and fellows each year. These training programs are supported by approximately 800 physician faculty members. We have not yet used this form with Liaison Committee on Medical Education (LCME) learners.

Additionally, the NCH-FCTM form was sent as an electronic link via email with an introduction and instructions to faculty to fill out as a formative reflective self-evaluation once yearly in 2016 and 2017. There were several goals in using this evaluation tool for the purpose of the faculty self-evaluation. The first goal was to familiarize faculty with the new scale. The second goal was to use the self-evaluation to help faculty reflect on teaching skill strengths and areas for improvement. Finally, we used the data as a needs assessment to help guide programming of faculty development activities at NCH, with plans to focus on the lowest scoring topics.

It was important to consider change management while implementing a new evaluation system of this scale. Since the numerical scale differed from the previous Likert scale and the purpose of the evaluation

was to provide more granular feedback to faculty on their teaching skills, introducing this scale included obtaining buy-in from key constituents. Educational materials were developed to train learners (i.e., residents and fellows) and faculty. Appendix B is a brief PowerPoint presentation used in learner meetings and faculty staff meetings to introduce and explain the milestone-based faculty evaluation system. Appendix C is a one-page handout that was distributed at these meetings and also emailed to faculty and learners to explain the new process.

Because teaching evaluations are used for promotion and tenure, as well as yearly incentive plans, a brief letter to the promotion and tenure committee and administrators was developed to accompany faculty promotion dossiers (Appendix D). This process and form were vetted by our institutional Graduate Medical Education Committee as well as by the College of Medicine’s Dean of Faculty Affairs prior to implementation.

Results

Results from the pilot and the initial implementation, as well as some anecdotal reflections from the faculty self-evaluations, are included. During the pilot period (July-September 2015), 29 resident raters used the NCH-FCTM evaluation. Inpatient faculty evaluations were collected and contrasted with standard Likert-based evaluations collected between July and September of 2014. Resident raters also provided comments about the NCH-FCTM.

The NCH-FCTM had similar variability in overall faculty evaluation scores compared to the previously used normative evaluations. The NCH-FCTM provided lower mean and median values and thus a more complete range to assess higher and lower performance (Table 2).

Table 2. Variability in Faculty Evaluation Scores

Descriptive Statistic	Normative Score (July-September 2014)	NCH-FCTM Score (July-September 2015)
<i>M</i>	4.44	3.63
<i>SD</i>	0.70	0.53
Coefficient of variation	0.16	0.15
<i>Mdn</i>	5	4
Interquartile range	4-5	3.5-4
Coefficient of dispersion	0.20	0.13

Abbreviation: NCH-FCTM, Nationwide Children’s Hospital Faculty Clinical Teaching

The NCH-FCTM form was well regarded by trainees, who felt that the domains better described faculty teaching activities compared to the previous traditional faculty evaluation model. They found the NCH-FCTM method easy to use. Sample pilot participants’ feedback included the following:

- “Descriptions are much more meaningful than the evaluation form we currently use.”
- “This evaluation form is very user-friendly; the lay-out and organization is much better than what we currently use. We should switch to using this form.”

Since implementing the NCH-FCTM, in addition to learner evaluations for each clinical rotation, we have also collected faculty self-evaluations. In 2016 and the first quarter of 2017, we collected over 7,200 learner evaluations of faculty and 550 faculty self-evaluations. The means of the learner evaluations and faculty self-evaluations for each of the clinical teaching categories are shown in Table 3. Faculty self-evaluation means were lower than the learner evaluation means for every category. There was general agreement between learners and faculty regarding which milestones were rated as high and which were rated as low. However, although the top three, the bottom three, and the middle four milestone groupings were the same, there were slight differences in the orders of individual items assigned by learners and faculty.

Table 3. Mean Learner and Self-Evaluation Ratings by Milestone

Clinical Teaching Assessment Domain	Milestone Score ^a	
	Learner Evaluation <i>M</i> (Rank Order ^b)	Self-Evaluation <i>M</i> (Rank Order ^b)
Milestone 1: Establishes Positive Learning Climate	3.60 (3)	3.39 (2)
Milestone 2: Maintains Control of Educational Session	3.57 (5)	3.30 (4)
Milestone 3: Establishes Learning Goals	3.50 (8)	3.04 (10)
Milestone 4: Promotes Understanding and Retention of Knowledge and Skills	3.50 (8)	3.19 (8)
Milestone 5: Provides Formative Feedback	3.48 (10)	3.15 (9)
Milestone 6: Promotes Clinical Reasoning	3.61 (2)	3.38 (3)
Milestone 7: Promotes Evidence Based Medicine	3.55 (7)	3.22 (6)
Milestone 8: Promotes Self-Directed Learning in Learners	3.57 (5)	3.20 (7)
Milestone 9: Balances Supervision and Autonomy	3.59 (4)	3.27 (5)
Milestone 10: Displays Professionalism	3.63 (1)	3.42 (1)

^aScores are based on a 4-point Likert scale (1 = *Inadequate Skills*, 2 = *Variable Skills*, 3 = *Effective Skills*, 4 = *Exemplary Skills/Role Model*).

^bHighest ranking = 1, lowest = 10.

In addition to gathering baseline data and familiarizing faculty with the NCH-FCTM, annual self-assessments using the new tool were implemented to encourage faculty self-reflection on teaching skills and to aid in targeted individual professional development plans.¹² The behavioral descriptors in the NCH-FCTM served as a guided assessment of individual teaching skills and encouraged faculty to think about how they might be assessed by others and themselves. This reflection was stimulated by the introductory email sent to faculty containing the online link to the survey and was demonstrated in some of the faculty comments. Sample comments from the 2017 self-assessment included the following:

- “Good survey. It is somewhat hard in clinic environment to keep all these goals in mind and this is good reminder.”
- “I like seeing them written down—makes me consciously, honestly think about it.”

Finally, the concordance between the faculty self-evaluations and the learner evaluations of faculty suggested three highest-yield areas for faculty development programs. After 1 year of data, faculty development programs were added, including piloting a quality improvement initiative in several divisions that coach faculty to help learners set effective learning goals, adding a half-day workshop designed to create a culture of feedback at our institution, and hosting a book club for clinician educators to discuss the work *Make It Stick*¹³ to promote teaching techniques to increase retention of information.

Discussion

The NCH-FCTM method was developed to provide a more behaviorally defined construct to evaluate the effectiveness of clinical teaching faculty. We found the NCH-FCTM method to be easily understood, to be relevant to clinical faculty in patient care settings, and to focus the evaluator (the learner) on a group of educational activities validated as important by previous research. After our first year, we revised our process of implementation, reflecting feedback from the trainees who desired more training in this method than they had been provided.

The implementation of the NCH-FCTM was straightforward and was well received by learners, who appreciated being able to give faculty more granular and appropriate feedback on issues important to them. In these early results, the NCH-FCTM provided variability in overall faculty scores similar to our previous normative-based Likert-scale method (Table 3). Additionally, both the overall scores and the individual evaluation scores were less negatively skewed using the NCH-FCTM. In its implementation, we found that the NCH-FCTM provided more specific detail in teaching behaviors, and faculty appeared to be motivated and informed to address specific areas when provided with these data.

Faculty self-assessments were consistently lower than learner assessments, likely reflecting more stringent self-evaluations. This effect has been noted previously in comparisons of faculty and learner evaluations.¹⁴ Concordance between the learner evaluations and self-evaluations in the top three and bottom three skill areas ranked suggests general intrarater reliability.

A powerful value of the faculty self-assessment is how it has helped guide faculty development programming directed at specific skill development in certain domains. NCH-FCTM evaluations were used for program planning. The four domains rated the lowest on the self-evaluations were the focus of faculty development grand round presentations over the past 18 months. Topics of the milestones were also the focus of a faculty development journal club and workshops presented to individual departments and divisions. We also used this information to rethink how we presented our faculty development activities to create a lasting impact, such as integrating faculty development in quality improvement, modeling and designing workshops to impact institutional culture, and looking for and reviewing education literature outside of medical education.

The NCH-FCTM data have been impactful for our faculty. We have seen faculty adjust their teaching efforts and concentrate on specific domains based on this feedback, demonstrating that the tool encourages effective self-reflection and planning in faculty. Anytime feedback is used for nonformative reasons, there can be tension between collecting data to improve performance versus focusing on “just getting good scores.” Changing the type of evaluation, especially around data used for promotion purposes, can engender some anxiety for faculty. We sought to address this through education and the letter with the teaching evaluations sent to the promotions committee.

Limitations

This model was only applied and tested in the clinical teaching of residents and fellows. Many of our faculty members also teach medical students and other professional students, each with their own regulations and evaluation requirements. While we believe the skills needed to be an effective clinical educator are similar, this tool has not yet been tested in the LCME or other populations.

Another limitation to this model is that it was tested across several departments in only a single institution (pediatric academic teaching hospital). Additional testing at other types of teaching institutions would strengthen its value.

When implementing a new evaluation system, it is important to consider the limitations of one’s learning management system. Initially, ours was not aligned with the rating scale of the new tool (given the 4-point scale of the NCH-FCTM). This was corrected in year 2 with a transition to a new learning management system.

There are still some faculty who dislike change, especially as they are gathering teaching evaluations for promotion purposes. Continued education and coaching are provided to these faculty through our Center for Faculty Development.

Future Directions

Next steps include conducting more detailed reliability and validity assessments of the tool to enable us to continue to refine it. While the content validity of the tool is good, additional testing for construct validity is needed. We will also be evaluating the impact of this tool by monitoring the ACGME resident and fellow survey questions that correlate with our milestones. Additionally, we plan to assess the alignment and applicability of this model to LCME teaching activities.

We have self-assessment data for 2 years for many faculty, and we plan to continue to track yearly self-assessments and cross-reference them with participation in faculty development activities in order to investigate two questions. Whether faculty attend development sessions about topics that they rate themselves as variable or poor on is something we would like to know. Furthermore, we plan to evaluate change over time in self-assessments to investigate the impact of faculty development programs across the institution.

Karyn Kassis, MD, MPH: Assistant Professor of Pediatrics, Nationwide Children's Hospital; Assistant Professor of Pediatrics, The Ohio State University College of Medicine; Director, Center for Faculty Development, Nationwide Children's Hospital; Director, Center for Faculty Development, The Ohio State University College of Medicine

Rebecca Wallihan, MD: Assistant Professor of Pediatrics, Nationwide Children's Hospital; Assistant Professor of Pediatrics, The Ohio State University College of Medicine; Associate Program Director, Pediatric Residency Program, Nationwide Children's Hospital; Associate Program Director, Pediatric Residency Program, The Ohio State University College of Medicine; Vice Chair for Education, Nationwide Children's Hospital; Vice Chair for Education, The Ohio State University College of Medicine

Larry Hurtubise: Associate Director, Center for Faculty Development, Nationwide Children's Hospital; Adjunct Associate Professor of Biomedical Education and Anatomy, The Ohio State University College of Medicine

Sara Goode: Program Administrator, Office of Graduate Medical Education, Nationwide Children's Hospital

Margaret Chase, MD: Associate Professor of Pediatrics, Nationwide Children's Hospital; Associate Professor of Pediatrics, The Ohio State University College of Medicine; Program Director of the Internal Medicine/Pediatrics Residency, Nationwide Children's Hospital; Program Director of the Internal Medicine/Pediatrics Residency, The Ohio State University College of Medicine

John D. Mahan, MD: Professor of Pediatrics, Nationwide Children's Hospital; Professor of Pediatrics, The Ohio State University College of Medicine; Program Director, Pediatric Residency Program and Pediatric Nephrology Fellowship Program, Nationwide Children's Hospital; Program Director, Pediatric Residency Program and Pediatric Nephrology Fellowship Program, The Ohio State University College of Medicine

Disclosures

None to report.

Funding/Support

None to report.

Prior Presentations

Kassis KL, Hurtubise L, Goode S, Chase M, Wallihan R, Mahan J. A milestone-based tool for faculty evaluations: meaningful data about teaching practice. Poster presented at: AAMC Central Group on Education Affairs Conference; April 6-8, 2016; Ann Arbor, MI.

Ethical Approval

Reported as not applicable.

References

1. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? *J Gen Intern Med.* 2004;19(9):971-977. <https://doi.org/10.1111/j.1525-1497.2004.40066.x>
2. Beckman TJ, Lee MC, Mandrekar JN. A comparison of clinical teaching evaluations by resident and peer physicians. *Med Teach.* 2004;26(4):321-325. <https://doi.org/10.1080/01421590410001678984>
3. Risucci DA, Lutsky L, Rosati RJ, Tortolani AJ. Reliability and accuracy of resident evaluations of surgical faculty. *Eval Health Prof.* 1992;15(3):313-324. <https://doi.org/10.1177/016327879201500304>
4. Mintz M, Southern DA, Ghali WA, Ma IWY. Validation of the 25-item Stanford Faculty Development Program Tool on Clinical Teaching Effectiveness. *Teach Learn Med.* 2015;27(2):174-181. <https://doi.org/10.1080/10401334.2015.1011645>
5. Skeff KM, Stratos GA, Berman J, Bergen MR. Improving clinical teaching: evaluation of a national dissemination program. *Arch Intern Med.* 1992;152(6):1156-1161. <https://doi.org/10.1001/archinte.1992.00400180028004>
6. Shah D, Goettler CE, Torrent DJ, et al. Milestones: the road to faculty development. *J Surg Educ.* 2015;72(6):e226-e235. <https://doi.org/10.1016/j.jsurg.2015.06.017>
7. Mazzaccaro RJ, Rooney K, Donoghue EA. A milestone-based system of pediatric faculty evaluation: they aren't just for residents anymore. Poster presented at: Association for Pediatric Program Directors Annual Spring Meeting; March 30-April 2, 2016; New Orleans, LA. <http://scholarlyworks.lvn.org/cgi/viewcontent.cgi?article=1159&context=pediatrics>. Accessed April 3, 2017.
8. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med.* 1998;73(6):688-695. <https://doi.org/10.1097/00001888-199806000-00016>
9. Litzelman DK, Westmoreland GR, Skeff KM, Stratos GA. Factorial validation of an educational framework using residents' evaluations of clinician-educators. *Acad Med.* 1999;74(10):S25-S27.
10. Participating institutions. Stanford Faculty Development Center for Medical Teachers website. http://sfdc.stanford.edu/participating_institutions.html. Updated January 13, 2017. Accessed August 4, 2017.
11. Accreditation Council for Graduate Medical Education, Department of Field Activities. High-value data suggested for use in program evaluation and improvement. <http://www.acgme.org/Portals/0/PDFs/SelfStudy/HighValueDataSSandAPE.pdf>. Published May 2017. Accessed August 7, 2017.

12. Windish DM, Knight AM, Wright SM. Clinician-teachers' self-assessments versus learners' perceptions. *J Gen Intern Med.* 2004;19(5):554-557. <https://doi.org/10.1111/j.1525-1497.2004.30014.x>
 13. Brown PC, Roediger HL III, McDaniel MA. *Make It Stick: The Science of Successful Learning.* Cambridge, MA: Harvard University Press, 2014.
 14. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach.* 2009;31(8):685-695. <http://dx.doi.org/10.1080/01421590903050374>
-

Received: May 12, 2017 | **Accepted:** August 16, 2017 | **Published:** September 18, 2017