Use of Key Performance Indicators to Improve Milestone Assessment in Semi-Annual Clinical Competency Committee Meetings

FEI CHEN, PHD
HARENDRA ARORA, MD
SUSAN M. MARTINELLI, MD

INTRODUCTION

The Accreditation Council for Graduate Medical Education's (ACGME) Next Accreditation System (NAS) requires residency programs to semiannually submit composite milestone data on each resident's performance.[1] In the past few years, our Clinical Competency Committee (CCC) members referred to the ACGME milestone assessment rubrics when reviewing resident performance. However, rater judgement in competency-based assessment has been suggested to be variable and fraught with bias.[2,3] Research has shown that faculty often unintentionally generate biased and subjective judgments based on their overall impression of a resident, especially when criteria are not explicit.[4,5] Consistent with these findings in the competency-based assessment literature, we found many of the ACMGE anesthesiology milestone descriptors vague, leaving it to faculty's subjective discretion to translate them into observable clinical practice.[6] This report describes and evaluates post hoc a new assessment review procedure piloted in our departmental CCC semi-annual meeting in June 2016. The new process utilized key performance indicators (KPIs) to link milestone descriptors to clinical practice for the purpose of improving assessment efficiency and reliability.

METHODS

IRB statement

The study obtained exempt status from the Office of Human Research Ethics of the University of North Carolina, Chapel Hill (#17-0448).

Procedure

Identify and finalize KPI. A modified Delphi technique was utilized to develop KPIs aligned with milestone rubrics to assist in linking evaluation data to milestone levels. The KPIs were first identified and developed by the CCC chair and the education specialist, then reviewed by the CCC as a panel. The CCC chair and education specialist reviewed the clinical practice-based criteria and justifications made in past CCC meetings for resident milestone scoring decisions. When reviewing the criteria for each milestone sub-competency, 4 of the 25 sub-competencies were found to be scored based on objective data; for instance, Medical Knowledge (MK) 1 utilized scores from In Training Examinations, United States Medical Licensing Examination Step 3, and the American Board of Anesthesiology Basic Examination; Practice-based Learning and Improvement (PBLI) 1 was based on participation in quality improvement (QI) projects and conference presentations and publications related to QI projects; Professionalism (Prof) 3 took into account

maintenance of case log records, duty hour reporting, and maintenance of Advanced Cardiac Life Support certification; Prof 5 was prescored by the program director and associate program director based on their interactions with the residents as well as reports and feedback from other faculty members on residents with regard to the maintenance of personal well-being. In addition, there were two sub-competencies that had been traditionally scored solely according to performance on specific clinical rotations; for instance, Patient Care (PC) 6 was linked to critical care rotations and PC 7 was linked to pain rotations. These sub-competencies were prescored by the CCC faculty members from these divisions (critical care and pain) based upon evaluations and feedback from these rotations. For the remaining 19 milestone sub-competencies, the CCC chair and the education specialist identified concrete training progression (eg, completion of relevant rotations) and behavior indices (eg, for Professionalism 4, scores increased by 0.5 if the resident consistently sought feedback, was receptive to feedback and showed notable improvement; while scores were decreased by 0.5 if the resident lacked awareness of areas needing improvement after multiple feedback events from faculty or was defensive when receiving feedback), which explicitly communicate criteria

that guide scoring. Depending on KPI-related performance, merit, or deficiency, a resident's milestone placement can be deviated in positive or negative direction from the expected score corresponding to his/her training year by up to 1.0 point under CCC discretion. See Appendix A for KPI examples for 2 milestone sub-competencies subject to CCC group review and 1 milestone that is determined to be prescored and subject to minimal discussion time during the group review.

After the initial review of the milestone evaluation practice, the CCC chair and the education specialist wrote up a change plan for CCC members to review for consensus. All CCC members (N=13) met in April 2016 to review the KPIs for each milestone sub-competency and discuss the milestone sub-competencies suggested to be prescored. After the group review, an anonymous survey was sent to CCC members inquiring about their final votes and thoughts on the sub-competencies that had been largely agreed upon by the committee to be prescored, and the group unanimously allowed the 6 specific milestone sub-competencies linked to a particular set of objective criteria (MK 1, PBLI 1, Prof 3, and Prof 5) or rotations (PC 6 and PC 7) to be prescored prior to the meeting by 1 assigned CCC member. Historically, each resident received scores from 3 CCC faculty members independently for each milestone sub-competency. Although these sub-competencies were still reviewed by the group, they required minimal discussion during the meeting.

Using KPI to explicitly justify scores. We piloted the use of KPI at the CCC meeting in June 2016. While most residents are expected to achieve milestone levels corresponding to their training year and rotations completed, CCC members were instructed to adjust scores according to the KPI for those who had outstanding performance and those who were underperforming. For any scores that deviated from the expected milestone level, the CCC member was required to provide KPI-referenced comments justifying the placement decision. It has previously been shown that the aggregation of written comments helps with accurate clinical

Figure 1. CCC perception of the usefulness of the new milestone review procedure in improving milestone assessment (n=10).
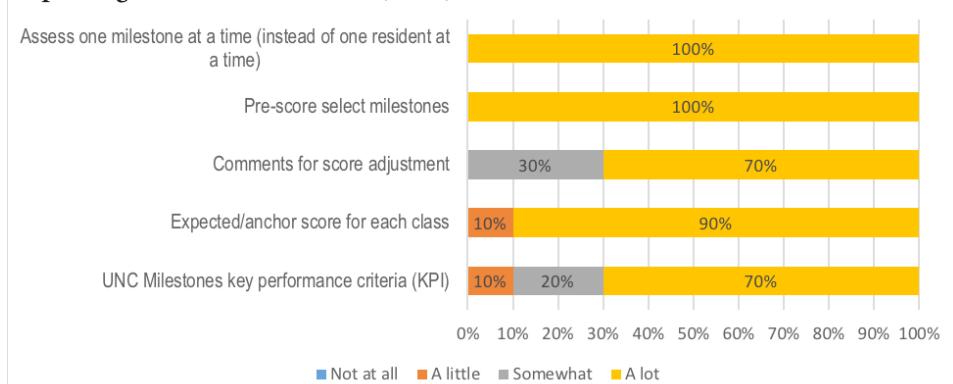
evaluation and provides insight into faculty interpretation of the supporting evidence when level of agreement is low, especially with borderline performance.[2,4,7] The committee reviewed the scores by sub-competency instead of by resident, which reduced the halo effect and emphasized the KPIs for the specific sub-competency under review.[8] Unless a committee member raised specific concerns over the scores, minimal discussion time was utilized for residents whose performance on the specific milestone sub-competency under review was consistently rated the same by all 3 raters. The committee discussion focused primarily on residents whose prescores did not reach consensus. In such circumstances, the CCC utilized the provided comments to determine final placement of the resident.

Data and Analysis

After the June 2016 CCC meeting, we surveyed CCC members on their perception of the new approach. Ten of the 12 CCC members who attended the June 2016 meeting responded to the survey.

We collected milestone placement data of the same cohort of 42 residents (Clinical Anesthesia Years 1–3) to examine the inter-rater reliability of the assessment procedures before and after the changes. Each resident received 3 independently rated scores from 3 CCC faculty members for both the historical and new procedures on all milestone sub-competencies (minus the 6 selected sub-competencies for the new procedure). The CCC consisted of the same 13 faculty members for the 2 meetings and the assignment of the faculty to the residents for scoring were random but

remained consistent across the meetings. Intra-class correlation coefficient (ICC) was used as the measure of the inter-rater reliability. The analysis was performed using SPSS 24.0 (IBM Corp, Armonk, NY).

RESULTS

Implementation of the new process reduced the length of the meeting from the historical 8 hours to 3.5 hours. The 10 CCC survey respondents agreed that the KPI-referenced milestone assessment process improved the efficiency of the CCC meeting. In addition, they all believed that assessing 1 milestone at a time (instead of 1 resident at a time) and prescoring select milestones improved milestone assessment (see Figure 1). Two CCC members commented that this procedure helped to streamline milestone assessment and demonstrated an efficient mechanism without compromising quality of the assessment. Higher inter-rater reliability of the milestone placement was obtained using the implemented KPIs (ICCsingle measure range: before=.53–.94, after=.74–.98). See Table 1.

DISCUSSION

According to Messick, content relevance and representativeness as well as criterion-relatedness are key aspects of construct validity as a unitary concept.[9] The enhanced practice of KPI-referenced assessment helped improve meeting efficiency, increased inter-rater reliability of the milestone scoring, and maintained CCC members' focus on content relevance and representation of resident knowledge and skills. In particular, the CCC found prescoring select milestones and reviewing

continued from previous page

performance by sub-competency instead of by resident helped improve the assessment. The existence of a halo effect in rater judgement in competence-based assessment is well-documented.[2] Research found that faculty raters achieved higher interrater reliability when they were asked to rate 7 dimensions of performance rather than 2 dimensions in competence-based assessment.[8] The results we obtained—higher inter-rater reliability through explicit communication of KPI and emphasis on 1 milestone at a time—are in alignment with the literature. With this new design, were we convinced that improved inter-rater reliability was due to a decrease in cognitive load and halo effect? That is still subject to examination with more rigorous design. However, we do recommend reviewing information by sub-competency, as it was well perceived by our CCC members and helped them focus on content relevance in resident placement on each milestone.

Additionally, the new process helped identify sub-competencies that had lacked supporting evidence for resident performance to provide valid assessments, suggesting a needed change in the evaluation procedure. For instance, we found it difficult to assess Patient Care 4: Management of Peri-anesthetic Complications because it is currently challenging to monitor our residents' postoperative patient visits. Thus, procedural or expectation changes at the departmental level are needed to obtain useful data for milestone placement on this sub-competency.

As next steps, we plan to refine the KPIs and add exemplars to better communicate the criteria of each sub-competency. The aggregated comments used as evidence to support score adjustment provide rich clinical language to specify and exemplify KPIs. Additionally, despite the improved efficiency in integrating the available data in CCC decision-making, the validity of the assessment still relies heavily on the quality of clinical faculty evaluations based on residents' rotation performance and daily interactions. It has been a common challenge to many clinical faculty to accurately translate the observed resident clinical performance into evaluation scores and articulate assessment

processes in feedback.[10,11,12] Therefore, we will provide faculty evaluators with additional training on KPIs and techniques on daily feedback. Doing so allows CCC to have higher quality data for milestone placement.

CONCLUSION

The development of tangible, measurable indicators of performance criteria and an assessment protocol that highlights these indicators improved the CCC's ability to more efficiently and reliably assess milestone performance. Application of such indicators and protocols may also reveal specific sub-competencies that are difficult to assess and thus lead to changes in educational structure or evaluation procedures.

References

1. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system-rationale and benefits. *N Engl J Med.* 2012;366(11):1051-6.

2. Gauthier G, St-Onge C, Tavares W. Rater cognition: Review and integration of research findings. *Med Educ.* 2016;50(5):511-22.

3. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med.* 2014;89(5):721-7.

4. Ginsburg S, McIlroy J, Oulanova O, Eva K, Regehr G. Toward authentic clinical evaluation: Pitfalls in the pursuit of competency. *Acad Med.* 2010;85(5):780-6.

5. Gingerich A, van der Vleuten CPM, Eva KW, Regehr G. More consensus than idiosyncrasy: Categorizing social judgments to examine variability in min-CEX ratings. *Acad Med.* 2014;89(11):1510-9.

6. The anesthesiology milestone project. *J Grad Med Educ.* 2014;6(1 Suppl 1):15-28.

7. Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: Faculty interpretations of narrative evaluation comments. *Med Educ.* 2015;49(3):296-306.

8. Tavares W, Ginsburg S, Eva KW. Selecting and simplifying: Rater performance and behavior when considering multiple competencies. *Teach Learn Med.* 2016;28(1):41-51.

9. Messick S. Validity of psychological assessment. *Am Psychol.* 1995;50(9):741-9.

10. Yeates P, O'Neill P, Mann K, Eva K. Seeing the same thing differently: Mechanisms that contribute to assessor differences in directly-observed performance assessments. *Adv Heal Sci Educ.* 2013;18(3):325-41.

11. 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):1048-60.

12. St-Onge C, Chamberland M, Levesque A, Varpio L. Expectations, observations, and the cognitive processes that bind them: Expert assessment of examinee performance. *Adv Heal Sci Educ.* 2016;21(3):627-42.

Table 1. Intraclass Correlation Coefficient Assessing Milestone Competency Before and After Using the New Process

Competence	Single Measure		Average Measure	
	Before	After	Before	After
PC				
1	.79	.94	.92	.98
2	.79	.95	.92	.98
3	.80	.97	.92	.99
4	.64	.96	.84	.99
5	.64	.96	.84	.99
6	—	—	—	—
7	.94	—	.98	—
8	.84	.98	.94	.99
9	.66	.95	.85	.98
10	.56	.86	.80	.95
MK				
1	.57	—	.80	—
SBP				
1	.82	.95	.93	.98
2	.82	.96	.93	.99
PBLI				
1	.58	—	.81	—
2	.60	.97	.82	.99
3	.60	.95	.82	.98
4	.78	.94	.91	.98
Professionalism				
1	.76	.93	.90	.98
2	.59	.96	.81	.99
3	.56	—	.79	—
4	.53	.96	.77	.99
5	.78	—	.91	—
ICS				
1	.79	.92	.92	.97
2	.71	.93	.88	.98
3	.48	.94	.74	.98

PC = Patient Care; MK = Medical Knowledge; SBP = Systems-based Practice; PBLI = Practiced-based Learning & Improvement; ICS = Interpersonal & Communications Skills.

Six subcompetencies were prescored by the CCC chair after using the new procedure, so no data were available for ICC calculation.

Two-way random effects model where both people effects and measures effects are random.

Intraclass correlation coefficients using an absolute agreement definition.

Appendix A

A1. KPI Example of a Patient Care Milestone Sub-competency Subject to Group Review by the CCC					
Patient Care 1: Pre-anesthetic Patient Evaluation, Assessment, and Preparation					
Has not achieved Level 1	Level 1	Level 2	Level 3	Level 4	Level 5
	Performs general histories and physical examinations. Identifies clinical issues relevant to anesthetic care with direct supervision. Identifies the elements and process of informed consent.	Identifies disease processes and medical issues relevant to anesthetic care. Optimizes preparation of non-complex patients receiving anesthetic care. Obtains informed consent for routine anesthetic care; discusses likely risks, benefits, and alternatives in a straightforward manner; responds appropriately to patient's or surrogate's questions; recognizes when assistance is needed.	Identifies disease processes and medical or surgical issues relevant to subspecialty anesthetic care; may need guidance in identifying unusual clinical problems and their implications for anesthesia care. Optimizes preparation of patients with complex problems or requiring subspecialty anesthesia care with indirect supervision. Obtains appropriate informed consent tailored to subspecialty care or complicated clinical situations with indirect supervision. Must have done all 5 rotations (peds, neuro, OB, vascular, cardiac, thoracic) to reach level 3.0.	Performs assessment of complex or critically ill patients without missing major issues that impact anesthesia care with conditional independence. Optimizes preparation of complex or critically ill patients with conditional independence. Obtains appropriate informed consent tailored to subspecialty care or complicated clinical situations with conditional independence.	Independently performs comprehensive assessment for all patients. Independently serves as a consultant to other members of the health care team regarding optimal preanesthetic preparation. Consistently ensures that informed consent is comprehensive and addresses patient
UNC KPI			Must have done all 5 rotations (peds, neuro, OB, vascular, cardiac, thoracic) to reach level 3.0.		
⇧Move up 0.5 or 1 if show high autonomy dealing with comprehensive assessment beyond expected level. ⇩Move down 0.5 if below expectations. ⇩Move down 1 if severe adverse events happened due to poor preparation and assessment.					
Instructions on scoring: Note that the expected score for an average resident is CA=1.5, CA2=2.5, CA3=3.5 in December and CA=2, CA2=3, CA3=4 in June. Score: Comments:					

The authors are in the Department of Anesthesiology at University of North Carolina School of Medicine in Chapel Hill, NC. Fei Chen is an Education Specialist; Harendra Arora is a Professor of Anesthesiology and the Residency Program Director; Susan M. Martinelli is an Associate Professor of Anesthesiology and the Associate Residency Program Director.

Corresponding author: Fei Chen, PhD, Department of Anesthesiology, N2198 UNC Hospitals CB# 7010, Chapel Hill, NC, 27599-7010. Telephone (919) 445-6676, Fax (984) 974-4873, Email address: fchen@aims.unc.edu

Background: The Accreditation Council for Graduate Medical Education's Next Accreditation System requires residency programs to semiannually submit composite milestone data on each resident's performance. This report describes and evaluates a new assessment review procedure piloted in our departmental Clinical Competency Committee (CCC) semi-annual meeting in June 2016.

Methods: A modified Delphi technique was utilized to develop key performance indicators (KPI) linking milestone descriptors to clinical practice. In addition, the CCC identified six specific milestone sub-competencies that would be prescored with objective data prior to the meeting. Each resident was independently placed on the milestones by 3 different CCC faculty members. Milestone placement data of the same cohort of 42 residents (Clinical Anesthesia Years 1–3) were collected to calculate inter-rater reliability of the assessment procedures before and after the implemented changes. A survey was administrated to collect CCC feedback on the new procedure.

Results: The procedure assisted in reducing meeting time from 8 to 3.5 hours. Survey of the CCC members revealed positive perception of the procedure. Higher inter-rater reliability of the milestone placement was obtained using the implemented KPIs (Intraclass correlation coefficient [ICC] single measure range: before=.53–.94, after=.74–.98).

Conclusion: We found the new assessment procedure beneficial to the efficiency and transparency of the assessment process. Further improvement of the procedure involves refinement of KPIs and additional faculty development on KPIs to allow non-CCC faculty to provide more accurate resident evaluations.

Appendix A *continued*

A2. KPI Example of a Professionalism Milestone Sub-competency Subject to Group Review by the CCC					
Professionalism 4: Receiving and giving feedback					
Has not achieved Level 1	Level 1	Level 2	Level 3	Level 4	Level 5
	Accepts constructive feedback but occasionally demonstrates resistance to feedback while providing patient care.	Provides constructive feedback in a tactful and supportive way to medical students to enhance patient care. Accepts feedback from faculty members and incorporates suggestions into practice.	Consistently seeks feedback, correlates it with self-reflection, and incorporates it into life-long learning to enhance patient care. Seeks out feedback from faculty members and other members of the care team.	Provides constructive feedback in a tactful and supportive way to physician and non-physician members of the patient care team to enhance patient care.	Effectively provides feedback in challenging situations (eg, when there is resistance, there are adverse outcomes, or an experienced practitioner is involved).
UNC KPI					

⇧ Move up 0.5 if consistently seeks feedback, gives great feedback to faculty or trainees, receptive to feedback, and has shown notable improvement.

⇩ Move down 0.5 if lacks awareness of area in need of improvement after multiple feedback from faculty, defensive to feedback, or has shown no change.

⇩ Move down 0.5 if very little myTIPreport feedback (<=5 in 6 months).

⇩ Move down 0.5 if strongly negative 360 evaluations.

Instructions on scoring: Note that the expected score for an average resident is CA=1.5, CA2=2.5, CA3=3.5 in December and CA=2, CA2=3, CA3=4 in June.

Score:

Comments:

A3. KPI Example of a Milestone Sub-competency Prescored and Subject to Minimal Group Review by the CCC					
Professionalism 3: Commitment to institution, department, and colleagues					
Has not achieved Level 1	Level 1	Level 2	Level 3	Level 4	Level 5
	Complies with institutional policies and regulations, including work schedule rules.	Acts as a reliable team member, recognizing the impact of one's own work responsibilities on the institution and on one's colleagues. Volunteers to assist colleagues, when appropriate, to cover illnesses/absences to ensure quality patient care. Completes requested evaluations (eg, faculty member, program, peers, ACGME Resident Survey) in a timely manner.	Serves as a resource and counselor to medical students regarding their professional choices and behaviors.	Serves as a resource and counselor to junior residents regarding their professional choices and behaviors.	Models responsibility and accountability in one's professional choices and behaviors.
UNC KPI					

⇧ Move up 0.5 if has served 6 months or more as chief resident.

⇩ Move down 0.5 if delinquent on logging of duty hours 2–3 x/semester.

⇩ Move down 0.5 if delinquent on logging cases 2–3 x/semester.

⇩ Move down 0.5 if delinquent in G1 responsibilities.

⇩ Move down 0.5 if not currently certified in ACLS.

⇩ Move down 1 if delinquent on logging cases 4 or more x/semester.

⇩ Move down 1 if delinquent on logging of duty hours 4 or more x/semester.

⇩ Move down 0.5 if failed to attend CBA or OSCE; move down 1.0 if failed to do both.

Instructions on scoring: Note that the expected score for an average resident is CA=1.5, CA2=2.5, CA3=3.5 in December and CA=2, CA2=3, CA3=4 in June.

Score:

Comments: