

ACGME Milestones Within Subspecialty Training Programs: One Institution's Experience

Janae K. Heath, MD

C. Jessica Dine, MD, MSHP

ABSTRACT

Background The Accreditation Council for Graduate Medical Education Milestones were created as a criterion-based framework to promote competency-based education during graduate medical education. Despite widespread implementation across subspecialty programs, extensive validity evidence supporting the use of milestones within fellowship training is lacking.

Objective We assessed the construct and response process validity of milestones in subspecialty fellowship programs in an academic medical center.

Methods From 2014–2016, we performed a single center retrospective cohort analysis of milestone data from fellows across 5 programs. We analyzed summary statistics and performed multivariable linear regression to assess change in milestone ratings by training year and variability in ratings across fellowship programs. Finally, we examined a subset of Professionalism and Interpersonal and Communication Skills subcompetencies from the first 6 months of training to identify the proportion of fellows deemed “ready for independent practice” in these domains.

Results Milestone data were available for 68 fellows, with 75 933 unique subcompetency ratings. Multivariable linear regression, adjusted for subcompetency and subspecialty, revealed an increase of 0.17 (0.16–0.19) in ratings with each postgraduate year level increase ($P < .005$), as well as significant variation in milestone ratings across subspecialties. For the Professionalism and Interpersonal and Communication Skills domains, mean ratings within the first 6 months of training were 3.78 and 3.95, respectively.

Conclusions We noted a minimal upward trend of milestone ratings in subspecialty training programs, and significant variability in implementing milestones across differing subspecialties. This may suggest possible difficulties with the construct validity and response process of the milestone system in certain medical subspecialties.

Introduction

The primary goal of graduate medical education (GME) is to prepare residents and fellows for independent practice. Competency-based medical education was introduced and widely adopted to achieve this goal^{1–4}; however, the paradigm shift also required an innovative assessment system. The Accreditation Council for Graduate Medical Education (ACGME) introduced milestones, a novel criterion-based framework for GME competency assessment.^{5–7} The milestones created an assessment framework addressing the 6 core competencies and provided a trajectory-based metric to demonstrate a trainee's progression toward competence.^{5,7,8}

The milestones were introduced at the residency level in 2013, and subsequently implemented across medical subspecialties the following year, based on collaboration between the ACGME and the American Board of Medical Specialties. Early experiences with the residency milestones provided ample validity evidence in this population.^{9–16} One study supported the construct validity of competency assessment using milestones through the demonstration of a predictable

upward trend of milestone ratings over the course of residency training.¹⁰ While recent data from the ACGME highlighted a similar upward trend in subspecialty training programs nationally,¹⁷ it is not clear if this trend would be observed within research-oriented fellowship training programs. Furthermore, generic milestones may not be sufficient for diverse medical subspecialties, particularly given significant variability in training program structure and possible variability in milestone interpretation. Despite widespread implementation of milestones in subspecialty training programs, there is a lack of extensive validity evidence supporting the use of milestones in the fellowship setting.

We aimed to assess the validity of medical subspecialty milestone ratings within our institution across 2 of Messick's validity domains—construct validity and response process.^{18,19} Our first objective was to evaluate the construct validity of the milestones in a research-intensive fellowship setting by examining the trend of ratings over the duration of training (akin to the validity evidence used in the residency setting). To evaluate the response process of milestones within fellowship, we assessed the variation of ratings between medical subspecialties, testing the hypothesis that appropriate response process

DOI: <http://dx.doi.org/10.4300/JGME-D-18-00308.1>

should result in consistent overall milestone ratings across subspecialties. Finally, with the hypothesis that the majority of fellows would have achieved competency within context-independent Professionalism and Interpersonal and Communication Skills domains prior to initiation of subspecialty training, we evaluated these milestone ratings within the first 6 months of fellowship training as an additional indicator of the response process.

Methods

Setting and Participants

We performed a single center retrospective cohort analysis of milestone data obtained from evaluations of subspecialty fellows within a large academic medical center from 2014 to 2016, using a convenience sample of 5 training programs: cardiology, pulmonology/critical care, endocrinology, hematology/oncology, and rheumatology. Complete milestone assessments of fellows (each consisting of 24 subcompetencies, grouped within the 6 competency domains) were submitted by faculty at the conclusion of each clinical rotation. Milestone assessments were introduced within our institution in 2013, and all faculty were provided written instruction regarding ACGME Milestone ratings. No further rater training was implemented by individual subspecialty programs. All submitted milestone assessments during the 2014–2016 time frame were compiled and deidentified to create the cohort for analysis. Submitted milestone assessments (on a 9-point scale) were converted to a 5-point scale for study purposes, consistent with prior literature analyzing milestones.

The Institutional Review Board at the University of Pennsylvania approved the study as exempt.

Data Analysis

We performed summary statistics across all submitted evaluations. To assess the construct validity of milestone ratings among fellows, we used the surrogate marker of milestone trajectory over the progression of training. This is consistent with previously published research reporting milestones in the residency setting, with the assumption that subspecialty fellows should progress to higher levels of competency throughout fellowship training.¹⁰ We examined the degree that milestone ratings varied by fellows' postgraduate year (PGY) level using linear regression, adjusting for subspecialty and subcompetency topics using forward selection. Our sample size of 75 933 allows identification of a 0.004% difference in milestone ratings, with a statistical power level of 0.80 with an alpha of .05. We also calculated overall effect size using Cohen's f^2 .

What was known and gap

Despite widespread implementation of milestones in fellowship programs, extensive validity evidence supporting their use in the fellowship setting is lacking.

What is new

A single center retrospective cohort analysis of milestone data from fellows in 5 programs.

Limitations

Approach to milestones differs among institutions, limiting generalizability; construct validity could not be assessed.

Bottom line

Milestone ratings improved over time but varied significantly across subspecialties.

We evaluated the variability of milestone ratings across subspecialties as a surrogate assessment of the response process using a multivariable linear regression model. A sensitivity analysis was performed to assess for degree of variation introduced into the model by a trainee, using a mixed effects linear regression model clustered on the individual.

To further evaluate response process, we identified a subgroup of Professionalism and Interpersonal and Communication Skills Milestones submitted within 6 months of beginning fellowship training. These subcompetencies were selected based on the identical nature of descriptions and behavioral anchors to the Internal Medicine Milestones and their presumed context-independent nature. Therefore, a trainee would be expected to have achieved "readiness for independent practice" (correlating to a rating of at least 4 on a 5-point scale) on these domains prior to beginning subspecialty training. We performed descriptive statistics across these 2 competency domains for the first 6 months of fellowship, and then identified the percentage of trainees who were less than the designated "ready for independent practice" target.

All statistical analyses were completed using STATA version 14.3 (StataCorp LLP, College Station, TX).

Results

Complete milestone data were available for 68 fellows from 2014 to 2016, consisting of 75 933 unique subcompetency milestone ratings (BOX). Over 2 academic years, the cohort consisted of evaluations on 3 PGY-3, 36 PGY-4, 49 PGY-5, 20 PGY-6, and 3 PGY-7 fellows. Half of evaluations (54%, 37 of 68) were submitted from the pulmonology/critical care and cardiology fellowships consistent with the larger size of these fellowship programs. The majority of evaluations (85%, 64 452 of 75 933) within our

Box Overview of Subspecialty ACGME Subcompetencies**Patient Care**

- Gathers and synthesizes essential and accurate information to define each patient's clinical problems (PC-1)
- Develops and achieves a comprehensive management plan for each patient (PC-2)
- Manages patients with progressive responsibility and independence (PC-3)
- Demonstrates skill in performing and interpreting invasive procedures (PC-4a)
- Demonstrates skill in performing and interpreting noninvasive procedures and/or testing (PC-4b)
- Requests and provides consultative care (PC-5)

Medical Knowledge

- Possesses clinical knowledge (MK-1)
- Knowledge of diagnostic testing and procedures (MK-2)
- Scholarship (MK-3)

Systems-Based Practice

- Works effectively within an interprofessional team (eg, peers, consultants, nursing, ancillary professionals, and other support personnel; SBP-1)
- Recognizes system error and advocates for system improvement (SBP-2)
- Identifies forces that impact the cost of health care, and advocates for and practices cost-effective care (SBP-3)
- Transitions patients effectively within and across health delivery systems (SBP-4)

Practice-Based Learning and Improvement

- Monitors practice with a goal for improvement (PBLI-1)
- Learns and improves via performance audit (PBLI-2)
- Learns and improves via feedback (PBLI-3)
- Learns and improves at the point of care (PBLI-4)

Professionalism

- Has professional and respectful interactions with patients, caregivers, and members of the interprofessional team (eg, peers, consultants, nursing, ancillary professionals, and support personnel; PROF-1)
- Accepts responsibility and follows through on tasks (PROF-2)
- Responds to each patient's unique characteristics and needs (PROF-3)
- Exhibits integrity and ethical behavior in professional conduct (PROF-4)

Interpersonal and Communication Skills

- Communicates effectively with patients and caregivers (ICS-1)
- Communicates effectively in interprofessional teams (eg, peers, consultants, nursing, ancillary professionals, and other support personnel; ICS-2)
- Appropriate utilization and completion of health records (ICS-3)

TABLE 1

Composition of Milestone Evaluation Data

	Trainees, n (%)	Subcompetency Evaluations, n (%)
Subspecialty		
Pulmonology/critical care	19 (28)	26 147 (34)
Cardiology	18 (26)	23 001 (30)
Hematology/oncology	21 (31)	6560 (9)
Endocrinology	6 (9)	16 025 (21)
Rheumatology	4 (6)	4200 (6)
PGY at time of evaluation ^a		
PGY-3	3 (4)	4369 (6)
PGY-4	27 (40)	35 014 (46)
PGY-5	30 (44)	29 438 (39)
PGY-6	6 (9)	6836 (9)
PGY-7	2 (3)	276 (0.4)
Subcompetency topic		
Interpersonal and Communication Skills	68 (100)	12 630 (17)
Medical Knowledge	68 (100)	11 115 (15)
Practice-Based Learning and Improvement	68 (100)	10 372 (14)
Patient Care	68 (100)	20 107 (26)
Professionalism	68 (100)	14 456 (19)
Systems-Based Practice	68 (100)	7253 (10)

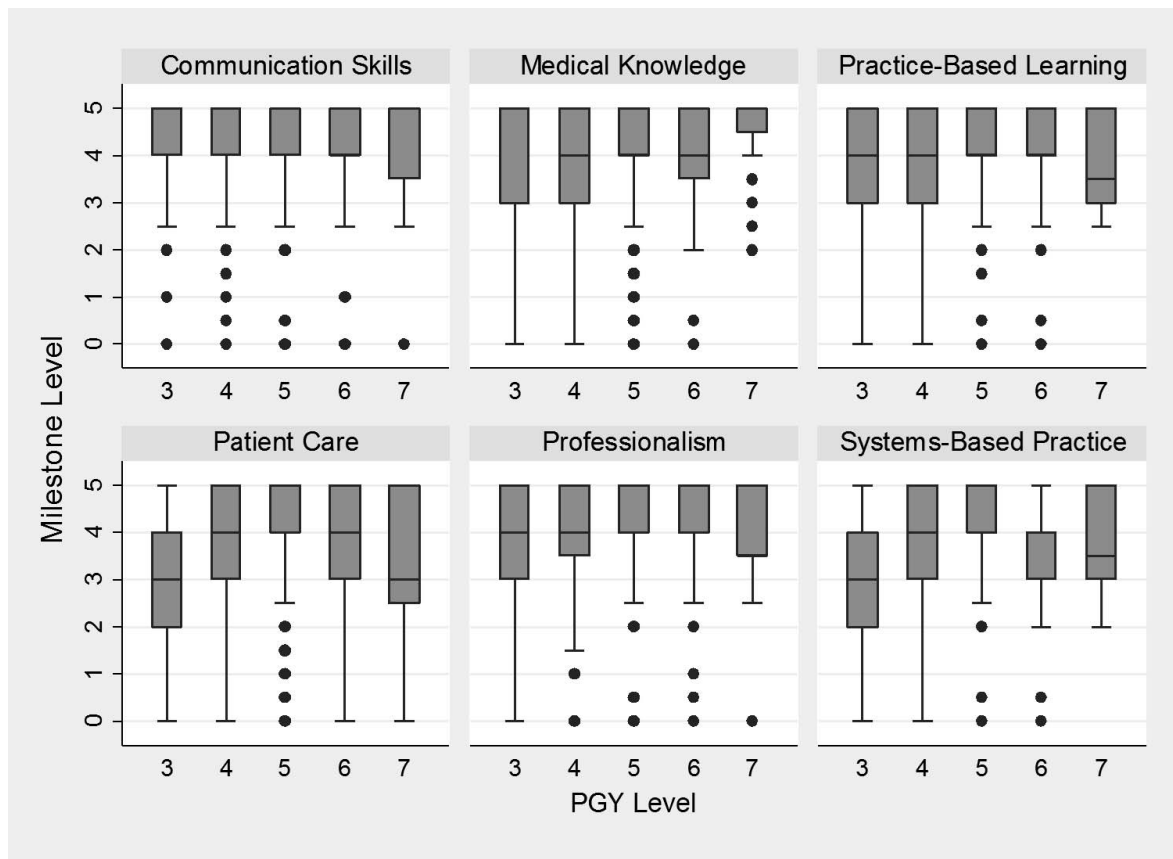
Abbreviation: PGY, postgraduate year.

^a Indicates PGY level at time of first submitted evaluation in the 2-year cohort.

cohort were submitted on PGY-4 and PGY-5 trainees (TABLE 1).

To assess trends in milestone ratings over the progression of training, the FIGURE displays summary data of milestone ratings over the PGY level trajectory across each of the 5 subspecialties. A multivariable linear regression adjusted for subcompetency topic and subspecialty revealed an increase of 0.17 (0.16–0.19) in milestone score with each PGY level increase ($P < .005$). Effect size, calculated using Cohens f^2 , was 0.15, suggesting a medium effect size.

In addition to the progression over PGY level, there was significant variability in mean milestone ratings according to subspecialty (TABLE 2), which persisted after adjusting for subcompetency and PGY level. Specifically, training within 1 subspecialty affected the overall milestone rating by an additional 1.15 points (on a 5-point scale) compared to the reference subspecialty. A sensitivity analysis using a mixed effect linear regression model, clustered on trainee to account for the variability by individual fellow, did not significantly change the findings of the original model.

**FIGURE****Bar Graph of Mean Milestone Ratings Over Postgraduate Level**

Note: As seen in this diagram, the median milestone level for each fellowship year is represented by the horizontal line, bounded by the 25th and 75th rank of milestone ratings (interquartile range). The dots represent outliers, and the bar represents the 95% confidence interval of the findings.

Abbreviation: PGY, postgraduate year.

Finally, we identified 3730 subcompetency ratings within the Professionalism domain and 4753 subcompetency ratings within the Interpersonal and Communication Skills domain submitted on trainees within the first 6 months of fellowship. Of these, 34% (1627 of 4753) of Professionalism subcompetencies and 26% (975 of 3730) of Interpersonal and Communication Skills subcompetencies were less than 4 (mean ratings of 3.78 and 3.95, respectively). Ultimately, the percentage of fellows achieving a value of 4 in Professionalism and Interpersonal and Communication Skills subcompetencies increased over the course of fellowship training (to 78% [663 of 846] and 82% [544 of 663], respectively, in the final year of training).

Discussion

Using the trajectory of milestones as a surrogate for construct validity in subspecialty training programs, we noted an increase in milestones with progression through training. Yet we found significant variability

of milestone ratings across subspecialties, highlighting the differential use of this assessment tool across subspecialties within a single institution. Finally, in assessing the response process using Professionalism and Interpersonal and Communication Skills domains as potential context-independent milestones, we found that 34% of Professionalism subcompetencies and 26% of Interpersonal and Communication Skills subcompetencies were less than the ACGME “graduation target” during the first 6 months of fellowship training.

Although our results show milestone ratings improved with progression through training, this improvement is minimal when compared to the trajectory noted within the residency setting. Specifically, in a study assessing internal medicine residency milestones over a 3-year period, the authors noted an increase from 2.46 to 3.92 over a 36-month period of training across all milestones.²⁰ A similar trend was reported in national Internal Medicine Milestone ratings,⁹ which showed an increase up to 0.77 for each PGY level, highlighting a much greater change

TABLE 2
Multivariable Linear Regression of Milestone Rating^a

Parameter	Coefficient	P Value
PGY level (years in training)	0.18 (0.17–0.19)	< .005
Specialty		
A	Reference	
B	0.23 (0.20–0.27)	< .005
C	0.62 (0.57–0.67)	< .005
D	0.49 (0.45–0.52)	< .005
E	1.15 (1.12–1.19)	< .005
Subcompetency		
Systems-Based Practice	Reference	
Patient Care	0.04 (0.01–0.76)	.016
Medical Knowledge	0.11 (0.07–0.15)	< .005
Practice-Based Learning and Improvement	0.23 (0.20–0.27)	< .005
Interpersonal and Communication Skills	0.32 (0.28–0.36)	< .005
Professionalism	0.37 (0.34–0.41)	< .005

Abbreviation: PGY, postgraduate year.

^a Multivariable linear regression adjusted for PGY level, subspecialty, and subcompetency topic.

per PGY than within our cohort of fellows. While recent data from the ACGME show upward trends of milestone ratings in subspecialty training programs nationwide,¹⁷ this does not account for myriad institutional factors that affect milestone ratings. One potential explanation for our findings may be the structure of subspecialty training within our institution, which consists of intensive clinical training followed by nonclinical activities, such as dedicated research training. Also, fellows may have less dramatic increases in ratings during initial years of training. Regardless, this serves to further highlight unique challenges with this assessment metric in the fellowship setting.

The significant variability noted between subspecialty milestone ratings could indicate problems with the response process of the assessment as currently operationalized. Within our institution, we suspect this difference may reflect differential rater interpretation of milestones rather than differences in skill sets of fellows, perhaps due to the absence of significant rater training across all subspecialties. Further qualitative work to assess potential rater challenges, impact of familiarity bias, or other task influences impacting rater use of the scale will be important future research.

While two-thirds of fellows were rated “ready for independent practice” when starting fellowship training in the domains of Professionalism and Interpersonal and Communication Skills, prior literature suggested that approximately 95% of internal

medicine residency graduates achieve the designated graduation target (milestone rating greater than 4) in these domains.¹⁰ This finding suggests that these subcompetencies are context-dependent in nature, contrary to our original hypothesis, and consistent with the original development of the subspecialty milestones. It is unlikely that a high percentage of residents degraded in professionalism and communication between residency and fellowship, but rather subspecialty faculty may expect more nuanced communication and professionalism skills than expected of an internal medicine resident. However, if the fellowship milestones were entirely context-dependent, incoming fellows’ ratings should primarily be in the range of 2 to 3 (on a 5-point scale). Our results showing initial ratings of 3.78 and 3.95 for the Professionalism and Interpersonal and Communication Skills domains, respectively, are higher than expected, which could be due to variable rater interpretation. Both the context of the training program and the rater interpretation of the subspecialty milestones could have important ramifications on use of the rating scale, and should guide future milestone revisions.

Our study has limitations, with the greatest being its generalizability to other programs, as the approach to milestones within each institution likely differs. The absence of dedicated faculty development and rater training regarding subspecialty milestones within our institution likely affects the interpretation and approach to fellow assessment. Additionally, because the milestones were implemented in 2014, there is a possibility that the novelty of this assessment within subspecialty programs has affected our findings. Finally, although we used a surrogate marker of trend in milestone ratings to imply construct validity, we were unable to truly assess this component of validity. There is potential that a failure to learn and/or a failure to appropriately teach material could result in the similar absence of milestone improvement throughout training (despite appropriate construct validity).

Further multi-institutional evaluation of milestones within fellowship settings is necessary. Evaluation of the rater interpretation of milestone ratings within subspecialties could provide critical information guiding further versions of the milestones. Finally, assessment of the impact of fellowship structure on implementation of milestones is warranted, specifically assessing the construct validity and response process of the use of general, non-specialty-specific milestones within subspecialties.

Conclusion

While we found an increase in milestone ratings during subspecialty training from our analysis of

fellows, there was also significant variability in milestone ratings across subspecialties. A third of fellows received initial milestones ratings below the “ready for independent practice” rating for Professionalism and Interpersonal and Communication Skills subcompetencies, which suggests that fellowship programs are interpreting these competencies in specialty-specific contexts and/or higher expectations for fellow performance.

References

1. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638–645. doi:10.3109/0142159X.2010.501190.
2. Hawkins RE, Welcher CM, Holmboe ES, Kirk LM, Norcini JJ, Simons KB, et al. Implementation of competency-based medical education: are we addressing the concerns and challenges? *Med Educ*. 2015;49(11):1086–1102. doi:10.1111/medu.12831.
3. ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med*. 2007;82(6):542–547. doi:10.1097/ACM.0b013e31805559c7.
4. Frank JR, Mungroo R, Ahmad Y, Wang M, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach*. 2010;32(8):631–637. doi:10.3109/0142159X.2010.500898.
5. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012;366(11):1051–1056. doi:10.1056/NEJMs1200117.
6. Swing SR, Beeson MS, Carraccio C, Coburn M, Iobst W, Selden NR, et al. Educational milestone development in the first 7 specialties to enter the Next Accreditation System. *J Grad Med Educ*. 2013;5(1):98–106. doi:10.4300/JGME-05-01-33.
7. Green ML, Aagaard EM, Caverzagie KJ, Chick DA, Holmboe E, Kane G, et al. Charting the road to competence: developmental milestones for internal medicine residency training. *J Grad Med Educ*. 2009;1(1):5–20. doi:10.4300/01.01.0003.
8. Bartlett KW, Whicker SA, Bookman J, Narayan AP, Staples BB, Hering H, et al. Milestone-based assessments are superior to Likert-type assessments in illustrating trainee progression. *J Grad Med Educ*. 2015;7(1):75–80. doi:10.4300/JGME-D-14-00389.1.
9. Hauer KE, Vandergrift J, Hess B, Lipner RS, Holmboe ES, Hood S, et al. Correlations between ratings on the resident annual evaluation summary and the internal medicine milestones and association with ABIM certification examination scores among US internal medicine residents, 2013–2014. *JAMA*. 2016;316(21):2253–2262. doi:10.1001/jama.2016.17357.
10. Hauer KE, Clauser J, Lipner RS, Holmboe ES, Caverzagie K, Hamstra SJ, et al. The internal medicine reporting milestones: cross-sectional description of initial implementation in US residency programs. *Ann Intern Med*. 2016;165(5):356–362. doi:10.7326/M15-2411.
11. Beeson MS, Carter WA, Christopher TA, Heidt JW, Jones JH, Meyer LE, et al. The development of the emergency medicine milestones. *Acad Emerg Med*. 2013;20(7):724–729. doi:10.1111/acem.12157.
12. Beeson MS. The emergency medicine milestones: with experience comes suggestions to improve. *Acad Emerg Med*. 2016;23(12):1434–1436. doi:10.1111/acem.13055.
13. Beeson MS, Holmboe ES, Korte RC, Nasca TJ, Brigham T, Russ CM, et al. Initial validity analysis of the emergency medicine milestones. *Acad Emerg Med*. 2015;22(7):838–844. doi:10.1111/acem.12697.
14. Aagaard E, Kane GC, Conforti L, Hood S, Caverzagie KJ, Smith C, et al. Early feedback on the use of the internal medicine reporting milestones in assessment of resident performance. *J Grad Med Educ*. 2013;5(3):433–438. doi:10.4300/JGME-D-13-00001.1.
15. Peabody MR, O'Neill TR, Peterson LE. Examining the functioning and reliability of the family medicine milestones. *J Grad Med Educ*. 2017;9(1):46–53. doi:10.4300/JGME-D-16-00172.1.
16. Turner TL, Bhavaraju VL, Luciw-Dubas UA, Hicks PJ, Multerer S, Osta A, et al. Validity evidence from ratings of pediatric interns and subinterns on a subset of pediatric milestones. *Acad Med*. 2017;92(6):809–819. doi:10.1097/ACM.0000000000001622.
17. Accreditation Council for Graduate Medical Education. Milestones Resources. National Reports. 2018 Milestone National Report. <https://www.acgme.org/What-We-Do/Accreditation/Milestones/Resources>. Accessed December 4, 2018.
18. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):166.e7–e16. doi:10.1016/j.amjmed.2005.10.036.
19. Newton PE, Shaw SD. *Validity in Educational & Psychological Assessment*. Thousand Oaks, CA: SAGE Publications; 2014.
20. Warm EJ, Held JD, Hellmann M, Kelleher M, Kinnear B, Lee C, et al. Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. *Acad Med*. 2016;91(10):1398–1405. doi:10.1097/ACM.0000000000001292.



Janae K. Heath, MD, is a Fellow, Pulmonary and Critical Care Medicine, Department of Medicine, and Center for Clinical

Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania; and **C. Jessica Dine, MD, MSHP**, is Associate Professor, Division of Pulmonary, Allergy, and Critical Care, and Associate Program Director, Internal Medicine Residency Program, Department of Medicine, and Leonard Davis Institute of Health Economics, Perelman School of Medicine at the University of Pennsylvania.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

This work was previously presented at the American Thoracic Society International Conference, Washington, DC, May 19–24, 2017.

Corresponding author: Janae K. Heath, MD, Hospital of the University of Pennsylvania, 5045 West Gates Building, 3400 Spruce Street, Philadelphia, PA 19104, 570.594.1857, fax 215.614.0869, janae.heath@uphs.upenn.edu

Received April 20, 2018; revision received October 24, 2018; accepted October 26, 2018.