

# Straight Line Scoring by Clinical Competency Committees Using Emergency Medicine Milestones

Michael S. Beeson, MD, MBA  
Stanley J. Hamstra, PhD  
Melissa A. Barton, MD  
Kenji Yamazaki, PhD  
Francis L. Counselman, MD

Philip H. Shayne, MD  
Eric S. Holmboe, MD  
Robert L. Muelleman, MD  
Earl J. Reisdorff, MD

## ABSTRACT

**Background** In 2013, milestone ratings became a reporting requirement for emergency medicine (EM) residency programs. Programs rate each resident in the fall and spring on 23 milestone subcompetencies.

**Objective** This study examined the incidence of straight line scoring (SLS) for EM Milestone ratings, defined as a resident being assessed the same score across the milestone subcompetencies.

**Methods** This descriptive analysis measured the frequencies of SLS for all Accreditation Council for Graduate Medical Education (ACGME)-accredited EM programs during the 2015–2016 academic year. Outcomes were the frequency of SLS in the fall and spring milestone assessments, changes in the number of SLS reports, and reporting trends. Chi-square analysis compared nominal variables.

**Results** There were 6257 residents in the fall and 6588 in the spring. Milestone scores were reported for 6173 EM residents in the fall (99% of 6257) and spring (94% of 6588). In the fall, 93% (5753 residents) did not receive SLS ratings and 420 (7%) did, with no significant difference compared with the spring (5776 [94%] versus 397 [6%]). Subgroup analysis showed higher SLS results for residents' first ratings (183 of 2136 versus 237 of 4220,  $P < .0001$ ) and for their final ratings (200 of 2019 versus 197 of 4354,  $P < .0001$ ). Twenty percent of programs submitted 10% or more SLS ratings, and a small percentage submitted more than 50% of ratings as SLS.

**Conclusions** Most programs did not submit SLS ratings. Because of the statistical improbability of SLS, any SLS ratings reduce the validity assertions of the milestone assessments.

## Introduction

The Accreditation Council for Graduate Medical Education (ACGME) milestones serve as objective competency measures that should be achieved throughout training.<sup>1</sup> For the Emergency Medicine (EM) Milestones, the 6 ACGME competencies were delineated into 23 subcompetencies, with 227 EM Milestones across 5 proficiency levels. Each subcompetency is scored in half increments from 0.5 (not yet achieved level 1) to 5.0 on a 10-point scale. Level 1 reflects the competency expected of a medical school graduate entering residency. Level 4 is the recommended performance for a graduating EM resident, whereas level 5 is expected to be achieved after years of clinical practice.<sup>2</sup> Final approval of a resident's readiness to practice without supervision lies with the program director.

Milestone scores are assessed by each residency program's Clinical Competency Committee (CCC) semiannually, and then entered into the ACGME online reporting system.<sup>3</sup> The CCC is expected to use multiple sources to individualize each resident's

assessment. Milestone scores provide resident feedback by tracking progress through training. Prior research supports validity assertions about using milestones to assess resident performance.<sup>4,5</sup>

Validity of the EM Milestones for resident assessment is affected by the accuracy of the scores. Accurate and independent scoring is important for determining that a resident has acquired the knowledge, skills, and abilities needed to graduate. In addition, because the American Board of Emergency Medicine (ABEM) designs its initial certification examinations on EM knowledge, skills, and abilities, closely linked to the EM Milestones, the accuracy of milestone ratings is important to ABEM to assess the validity of initial certification.

Straight line scoring (SLS) occurs when the same score on the 10-point scale is given to a single resident in each of the 23 EM Milestone subcompetencies. The likelihood of a resident achieving an SLS purely by chance is 1 in  $10^{23}$ . Resident evaluations are certainly not random across a 10-point scale; nonetheless, even if residents were instead rated on an interval scale using only 2 gradations (eg, satisfactory or unsatisfactory), SLS would occur once in every  $2^{22}$  ratings,

DOI: <http://dx.doi.org/10.4300/JGME-D-17-00304.1>

or once for every 4194304 residents. Assuming residents were accurately rated in each of the 23 subcompetencies, SLS would rarely occur.

The primary purpose of this study was to determine the incidence of SLS for EM Milestone ratings. The frequency of SLS for both the program and individual residents was studied.

## Methods

### Study Design

This study was a descriptive analysis of SLS frequency for an academic year. EM Milestone data were collected for fall 2015 and spring 2016 rating sets. The data included subcompetency ratings for residents reported to the ACGME by residencies. The study cohort included 6173 EM residents in the 173 ACGME-accredited categorical EM programs. No categorical programs were excluded.

### Study Protocol

Approximately every 6 months, the ACGME notifies residencies that milestone evaluations are due. The reporting periods reflect resident performance during the previous 6 months. Residencies enter milestone ratings through their unique online ACGME Accreditation Data System accounts.

Program identity was available to 2 investigators (S.J.H. and K.Y.), but was not used for this study. All other investigators viewed the resident and program results as aggregate and deidentified data. Individual resident and program results were not shared with the ACGME Review Committee for EM.

### Methods and Measurements

The EM Milestone data were analyzed for the frequency of SLS. Although the milestones use a 0 to 5 ordinal scale, there are actually 10 discrete ratings (0, 0.5, 1.0, 1.5, 2.0, etc). Within each program, the proportion of residents who received an SLS rating was noted.

Outcomes were defined as the frequency of SLS in the fall 2015 and spring 2016 reports. The fall milestone scores for interns were compared with the fall scores of all other EM residents. The spring milestone scores for residents in their final year of training were compared with the spring scores for all other EM residents. The frequency of programs with no, low, moderate, and high SLS were identified for both scoring sessions. Finally, SLS for EM 1 to 3 programs were compared with EM 1 to 4 programs to determine any difference.

Residencies were categorized as *no SLS*, if 0% of residents received SLS; *low SLS*, if 1% to 9% of

#### What was known and gap

Milestone-based assessments are an important aspect of competency-based education, and the educational community must ensure their validity.

#### What is new

A study examined the incidence of straight line scoring (SLS), defined as a resident being assessed the same score across the milestone subcompetencies, in Emergency Medicine Milestone ratings.

#### Limitations

Observational study that could not explore the reasons for SLS, which may limit generalizability.

#### Bottom line

Twenty percent of emergency medicine programs submitted significant to high percentages of SLS, which reduces the validity assertions for milestone assessments.

residents received SLS; *moderate SLS*, if 10% to 50% received SLS; and *high SLS*, if more than 50% of residents received SLS ratings. Although somewhat arbitrary, percentiles were developed to demonstrate a distinction among resident groups after input from EM faculty with expertise in the milestone rating process and consensus from psychometricians.

The study was approved as exempt research by the Akron General Health System/Akron General Medical Center Institutional Research Review Board.

### Analyses

We calculated descriptive statistics for the general data sample. For the comparison of nominal variables, chi-square analysis was used, and Yate's correction and Fisher's exact test were performed where appropriate. Based on sample sizes and earlier analyses, significance was set a priori as  $P \leq .01$  for analyses involving resident scores; for comparisons involving residency programs, significance was defined as  $P \leq .05$ .

### Results

All 173 ACGME-accredited EM residencies that reported milestone scores for fall 2015 and spring 2016 were included for analysis, including 128 (74%) 3-year (EM 1 to 3) programs and 45 (26%) 4-year (EM 1 to 4) programs. Total resident counts were 6257 residents in the fall and 6588 residents in the spring. Data for 6173 EM residents for whom both the fall and spring scores were submitted were included in the analysis (99% of the total in the fall and 94% of the total in the spring). Among the 6173 residents, there were 4152 (67%) in EM 1 to 3 programs and 2021 (33%) in EM 1 to 4 programs (TABLE 1).

**TABLE 1**  
Distribution of Residents

Program Format and Year of Training	Residents, No. (%) <sup>a</sup>
EM-1	1427 (34)
EM-2	1372 (33)
EM-3	1353 (33)
Subtotal	4152
EM-1	526 (26)
EM-2	519 (26)
EM-3	510 (25)
EM-4	466 (23)
Subtotal	2021
Total	6173

Abbreviation: EM, emergency medicine.

<sup>a</sup> Percentage of residents in same program format.

TABLE 2 shows the number of residents receiving and not receiving SLS in the 2 reporting periods by program format and year. There was no significant difference in the frequency of SLS between the fall and spring scores (chi-square test,  $P = .41$ ). For EM 1 to 3 programs, the difference between the number of residents who received SLS in the fall and spring (280 versus 292) was not statistically significant ( $P = .27$ ). For EM 1 to 4 programs, the difference (140 versus 105) was statistically significant ( $P = .021$ ).

Of the 1819 residents receiving their final milestone ratings in spring 2016, 200 (11%) received SLS, compared with the other 4354 residents, among whom, 197 (5%) received SLS ( $P < .0001$ ). Of the 1953 EM-1 residents receiving their first milestone ratings in the fall, 183 (9%)

received an SLS, which was statistically different from all other residents receiving fall ratings (chi-square test,  $P < .0001$ ).

In the fall, 62% (108 of 173) of programs did not rate any resident using SLS, compared with 59% (102 of 173) in the spring (chi-square test,  $P = .51$ ). Also in the fall reports, 20% (34 of 173) of programs were low SLS, 13% (22 of 173) were moderate SLS, and 5% (9 of 173) were high SLS. For the spring reports, 21% (36 of 173) were low SLS, 16% (28 of 173) were moderate SLS, and 4% (7 of 173) were high SLS. The SLS frequencies by type of program are shown in TABLE 3. Comparing the frequencies of programs in these various categories, there was no difference between the fall and spring ratings (chi-square test,  $P = .75$ ).

Of the 65 programs that submitted any SLS in the fall, 44 (68%) also reported SLS in the spring. Of the 108 programs that did not submit any SLS in the fall, 81 (75%) also did not report SLS in the spring.

Finally, for all findings reported in this study, Yate's correction and Fisher's exact test had no material difference on the results.

## Discussion

In this first study, to our knowledge, to assess the frequency of SLS reported by EM residencies, SLS reporting was found in 20% of programs overall, in EM 1 to 3 and EM 1 to 4 residency formats, and in spring and fall reports. The majority of programs that used SLS in the fall milestone report continued to report SLS in the spring.

The statistical improbability of SLS assumes that each subcompetency was assessed independently

**TABLE 2**  
Resident Scoring Patterns

Program Format and Year	Fall 2015		Spring 2016	
	No SLS	Yes SLS	No SLS	Yes SLS
	No. (%) of Residents	No. (%) of Residents	No. (%) of Residents	No. (%) of Residents
EM 1 to 3, n = 4152				
EM-1	1303 (31)	124 (3)	1356 (33)	71 (2)
EM-2	1296 (31)	76 (2)	1302 (31)	70 (2)
EM-3	1273 (31)	80 (2)	1202 (29)	151 (4)
Subtotal	3872 (93)	280 (7)	3860 (93)	292 (7)
EM 1 to 4, n = 2021				
EM-1	467 (23)	59 (3)	503 (25)	23 (1)
EM-2	488 (24)	31 (2)	500 (25)	19 (1)
EM-3	481 (24)	29 (1)	496 (25)	14 (1)
EM-4	445 (22)	21 (1)	417 (21)	49 (2)
Subtotal	1881 (93)	140 (7)	1916 (95)	105 (5)
Total	5753	420	5776	397

Abbreviations: SLS, straight line scoring; EM, emergency medicine.

TABLE 3

Frequency of Straight Line Scoring (SLS) by Category of Magnitude and Format

Program Format	Range of SLS (%)	Fall 2015	Spring 2016
		No. (%) of Programs	No. (%) of Programs
EM 1 to 3, n = 128	0	78 (61)	70 (55)
	1–9	26 (20)	29 (23)
	10–50	18 (14)	24 (19)
	> 50	6 (5)	5 (4)
EM 1 to 4, n = 45	0	30 (67)	32 (71)
	1–9	8 (18)	7 (16)
	10–50	4 (9)	4 (9)
	> 50	3 (7)	2 (4)

Abbreviation: EM, emergency medicine.

and was based on objective, individualized assessments for each resident. The finding of significant numbers of SLS suggests that such an assessment is not universally occurring. This is particularly troubling in that it suggests either a fundamental misunderstanding of how the CCC should function or the premise of the EM Milestone project itself. The EM programs have wide discretion in how residents are assessed and by whom; variability exists in faculty training on EM Milestone assessments, and CCCs approach assessments differently. The SLS rates raise the question about whether affected residents are being properly assessed and guided.

Residents receiving their first set of milestone ratings had a higher frequency of SLS. This could be due to relative unfamiliarity with a resident's clinical performance. The CCC also might have limited, first-hand clinical experience working with the resident. Under these circumstances, the CCC might generalize ratings based on a few subcompetencies for which the committee felt confident.

The numbers of SLS ratings increased significantly for graduating residents, which may reflect the CCC's opinion that they are ready to practice independently. Although the ACGME does not base accreditation decisions on milestone data other than compliance, there may be concern about program accreditation should residents graduate without achieving level 4 ratings.

A more optimistic explanation for level 4 assessment SLS is the scientific manner in which the milestones were derived from ABEM practice data,<sup>5</sup> with programs designing the resident experience to be well suited toward attaining level 4.

Although most residents did not receive SLS, the use of SLS for a significant portion of residents by 20% of programs raises concerns about the relationship between training and board certification. To date, no other specialty has aligned its

milestone level expected at graduation to that needed to achieve initial board certification. Stakeholders, such as residents, program directors, and the public, must be confident in resident assessment, which decreasing the incidence of SLS could improve.

The SLS rates could be a manifestation of confirmation bias, or halo effect, by the CCC resulting from generalizing performance scores. For 6% of all residents, there are SLS ratings. Straight line scoring is more prevalent in a subset of programs, with approximately 20% of programs reporting 10% or more ratings using SLS, and for a small percentage of programs, more than half of all ratings were SLS. Given the opportunity to identify residents who are lagging in competency acquisition, it is in the program's interest to accurately rate their residents.

This study has limitations. It is descriptive and does not provide any information on why SLS ratings exist. The total number of residents in EM programs for spring 2016 is greater than it was in the fall. It is conceivable that the difference could be from residents who transferred into EM during the academic year, but we cannot determine with any confidence why this difference in the total number of residents exists. In any case, the difference does not appear to have had an impact on the analyses of results. We did not evaluate the effect of a single SLS rating has on resident assessment or set a predetermined threshold for an acceptable degree, if any, of SLS. The results of this study are not generalizable to other specialties. The EM Milestones are unique in the manner in which they were developed and validated.

Future research could analyze the frequency of the same milestone rating for 1 subcompetency across a single EM residency class. It is conceivable that a CCC might expect residents to achieve specific subcompetency milestone ratings based

primarily on their year of training. A qualitative study of programs that used SLS to a moderate-to-high degree would be a reasonable next step to understanding the reasons why CCCs use SLS. Future research also could examine the frequency of SLS ratings across specialties to heighten awareness of CCCs as to the extremely low likelihood that SLS could occur by chance, given a detailed assessment of resident performance. The EM training community also should continue to discuss the impact of SLS for resident assessment.

## Conclusion

The frequency of SLS was low as most residency programs did not submit any SLS ratings but was much higher than statistically probable. SLS ratings were more prevalent for first- and final-year residents.

## References

1. Nasca TJ, Philbert I, Brigham T, et al. The next GME accreditation system—rationale and benefits. *N Eng J Med*. 2012;366(11):1051–1056.
2. Swing SR, Beeson MS, Carraccio C, 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.
3. Accreditation Council for Graduate Medical Education. Frequently asked questions: clinical competency committee and program evaluation committee. [https://www.acgme.org/acgmeweb/Portals/0/PDFs/FAQ/CCC\\_PEC\\_FAQs.pdf](https://www.acgme.org/acgmeweb/Portals/0/PDFs/FAQ/CCC_PEC_FAQs.pdf). Accessed September 5, 2017.
4. Korte RC, Beeson MS, Russ CM, et al. The emergency medicine milestones: a validation study. *Acad Emerg Med*. 2013;20(7):730–736.
5. Beeson MS, Carter WA, Christopher TA, et al. The development of the emergency medicine milestones. *Acad Emerg Med*. 2013;20(7):724–729.



**Michael S. Beeson, MD, MBA**, is Professor, Department of Emergency Medicine, Northeast Ohio Medical University, and Program Director and Medical Director of Simulation Learning Center, Akron General Medical Center; **Stanley J. Hamstra, PhD**, is Vice President, Milestone Research and Evaluation, Accreditation Council for Graduate Medical Education; **Melissa A. Barton, MD**, is Director of Medical Affairs, American Board of Emergency Medicine; **Kenji Yamazaki, PhD**, is Outcome Assessment Project Associate, Accreditation Council for Graduate Medical Education; **Francis L. Counselman, MD**, is Distinguished Professor and Chair, Department of Emergency Medicine, Eastern Virginia Medical School and Emergency Physicians of Tidewater; **Philip H. Shayne, MD**, is Professor and Vice Chair for Education, Department of Emergency Medicine, and Assistant Dean for Graduate Medical Education, Emory University School of Medicine; **Eric S. Holmboe, MD**, is Senior Vice President, Milestone Development and Evaluation, Accreditation Council for Graduate Medical Education; **Robert L. Muelleman, MD**, is Professor and Past-Chair, Department of Emergency Medicine, University of Nebraska; and **Earl J. Reisdorff, MD**, is Executive Director, American Board of Emergency Medicine.

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

Conflict of interest: Drs Beeson, Counselman, and Muelleman are directors for the American Board of Emergency Medicine. Drs Hamstra, Yamazaki, and Holmboe are employed by the Accreditation Council for Graduate Medical Education. Dr Shayne is the Chair of the Residency Review Committee for Emergency Medicine. Drs Barton and Reisdorff are employed by the American Board of Emergency Medicine. None of the investigators received financial compensation for the reported activity.

The authors would like to thank Ms Frances Spring for her assistance with the manuscript.

Corresponding author: Melissa A. Barton, MD, American Board of Emergency Medicine, 3000 Coolidge Road, East Lansing, MI 48823, 517.332.4800, [mbarton@abem.org](mailto:mbarton@abem.org)

Received April 25, 2017; revision received July 25, 2017; accepted August 7, 2017.