


See One, Do One, Forget One: Early Skill Decay After Paracentesis Training

Dana Sall, MD, MEd^{1,2,3} , Eric J. Warm, MD¹, Benjamin Kinnear, MD, MEd^{1,4}, Matthew Kelleher, MD, MEd^{1,4}, Roman Jandarov, Ph.D.⁵, and Jennifer O'Toole, MD, MEd^{1,4}



¹Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, USA; ²Cincinnati Veteran's Affairs Medical Center, Cincinnati, OH, USA; ³HonorHealth Scottsdale Thompson Peak Medical Center, Scottsdale, AZ, United States; ⁴Department of Pediatrics, Cincinnati Children's Hospital Medicine Center, University of Cincinnati College of Medicine, Cincinnati, OH, USA; ⁵Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH, USA.

INTRODUCTION: Internal medicine residents perform paracentesis, but programs lack standard methods for assessing competence or maintenance of competence and instead rely on number of procedures completed. This study describes differences in resident competence in paracentesis over time.

METHODS: From 2016 to 2017, internal medicine residents ($n = 118$) underwent paracentesis simulation training. Competence was assessed using the Paracentesis Competency Assessment Tool (PCAT), which combines a checklist, global scale, and entrustment score. The PCAT also delineates two categorical cut-point scores: the Minimum Passing Standard (MPS) and the Unsupervised Practice Standard (UPS). Residents were randomized to return to the simulation lab at 3 and 6 months (group A, $n = 60$) or only 6 months (group B, $n = 58$). At each session, faculty raters assessed resident performance. Data were analyzed to compare resident performance at each session compared with initial training scores, and performance between groups at 6 months.

RESULTS: After initial training, all residents met the MPS. The number achieving UPS did not differ between groups: group A = 24 (40%), group B = 20 (34.5%), $p = 0.67$. When group A was retested at 3 months, performance on each PCAT component significantly declined, as did the proportion of residents meeting the MPS and UPS. At the 6-month test, residents in group A performed significantly better than residents in group B, with 52 (89.7%) and 20 (34.5%) achieving the MPS and UPS, respectively, in group A compared with 25 (46.3%) and 2 (3.70%) in group B ($p < .001$ for both comparison).

DISCUSSION: Skill in paracentesis declines as early as 3 months after training. However, retraining may help interrupt skill decay. Only a small proportion of residents met the UPS 6 months after training. This suggests using

the PCAT to objectively measure competence would reclassify residents from being permitted to perform paracentesis independently to needing further supervision.

KEY WORDS: assessment; procedures; competency-based medical education; skill decay; paracentesis.

J Gen Intern Med 36(5):1346–51
DOI: 10.1007/s11606-020-06242-x
© Society of General Internal Medicine 2020

INTRODUCTION

Internal medicine residents perform procedures such as paracentesis and central venous catheter insertion^{1–3} yet no universally accepted method for procedural training or determination of competence exists. Many residency programs function under the Halstedian “see one, do one, teach one” method^{4,5} whereby, once a set number of procedures are performed with supervision, the resident can perform the procedure unsupervised in perpetuity. Program directors attest to a resident's competence by relying on the number of logged procedures, but procedural experience alone has not been shown to predict competent performance.^{2,6}

Sawyer et al. have suggested an evidence-based framework for procedural education and assessment of competence called “Learn, see, practice, prove, do, maintain.”⁴ Some programs have shown success in the first phase of the framework by creating robust simulation experiences where residents can *learn* the requisite procedural knowledge, *see* and *practice* in a simulation setting with feedback, and *prove* a minimum level of ability before performing the procedure on live patients^{2,3,7–9}. However, many still struggle to define competence for residents and faculty.¹⁰ Sawyer et al. suggest a tool that combines a skills checklist, global assessment scale, and entrustment ratings,⁴ but there is little published evidence describing use of such a tool.¹¹ Also concerning is that few programs have been able to systematically address *maintenance* of competence.¹² Instead, “once signed off, always signed off” is commonplace. This puts learners and patients in vulnerable positions where procedural opportunities are

Previous Presentations Initial publication of the PCAT and training materials can be found at Sall D, Wigger GW, Kinnear B, Kelleher M, Warm E, and O'Toole JK. Paracentesis Simulation: A Comprehensive Approach to Procedural Education. *MedEdPORTAL*. 2018;14:10747.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11606-020-06242-x>) contains supplementary material, which is available to authorized users.

Received May 21, 2020

Accepted September 11, 2020

Published online September 23, 2020

often sporadic and unpredictable due to factors such as duty hours, rotating resident schedules, and availability of additional procedural services such as interventional radiology and peripherally inserted central catheter teams.^{1, 10, 13, 14} Thus, potentially dyscompetent¹⁵ residents are getting even less competent over time. Thus, two questions remain: (1) How can we best measure procedural competence over time? (2) What is the rate of skill decay for paracentesis?

In this study, we set out to use a novel assessment tool, the Paracentesis Competency Assessment Tool (PCAT)¹¹, to examine paracentesis competence in our internal medicine residency program and to define the time interval at which skill decay occurs and retraining is necessary.

METHODS

Tool Development

We previously developed and gathered validity evidence for the Paracentesis Competency Assessment Tool (PCAT)¹¹. The PCAT contains a skills checklist, global skill assessment scale, and entrustment scale.^{4, 11} Prior standard setting work using the Angoff method¹⁶ sets two categorical composite cut-point scores: the Minimum Passing Standard (MPS) wherein learners must achieve to perform a paracentesis on patients with supervision and the Unsupervised Practice Standard (UPS), reflecting the composite score a learner would achieve if able to perform a paracentesis without supervision. This resulted in the MPS set at 19/26 (73%) on the checklist, combined with an *Advanced Beginner* rating on the global scale (2/4) and an *Able to Perform the Procedure With Direct Supervision* on the entrustment scale (2/5). The UPS was similarly set at 23/26 (88%) on the checklist, combined with an *Able* rating on the global scale (3/4) and an *Able To Perform The Procedure Without Supervision* on the entrustment scale (4/5).¹¹ The tool was piloted to ensure good interrater reliability prior to being used in real time (see [Supplemental Appendix](#)).

Study Design

The study was approved by the University of Cincinnati Institutional Review Board. Our program includes 123 Internal Medicine and Medicine-Pediatrics residents. Residents train at University of Cincinnati Medical Center and the Cincinnati Veterans Affairs Medical Center (VAMC). All residents rotate on services that may require completion of paracentesis, and were thus all eligible for enrollment. This study took place from September 2016 to June 2017.

To participate in the study, residents provided written informed consent. Of the 123 residents eligible, 118 provided consent and enrolled in the study (Fig. 1).

Following Sawyer's framework,⁴ learners were scheduled for individual training simulation sessions. Sessions took place at the simulation laboratory in the VAMC using a Blue

Phantom Paracentesis Ultrasound Training Model (Item # BPPara1301, <https://www.bluephantom.com/>). Prior to individual training simulation sessions, we instructed learners to watch online modules and demonstration videos. On arrival to the simulation center, learners again viewed demonstration videos. Learners practiced completing a therapeutic paracentesis procedure with instructor feedback as many times as needed, until they felt ready to attempt a final performance. This final performance score using the PCAT served as their baseline score.

After initial training, residents were randomized to return to the simulation lab 3 and 6 months later (group A, $n = 60$) or only 6 months later (group B, $n = 58$). There were no significant differences in gender or class year between groups (Table 1).

Residents in group A returned to the simulation center within 3 months of their initial training date. There were no pre-work requirements prior to returning for this session. On arrival to the simulation center, residents were instructed to perform a therapeutic paracentesis from start to finish without any instructor feedback. Faculty raters assessed performance using the PCAT in real time. This served as the "three-month score." After the performance, faculty reviewed the assessment with the learners, and learners were able to practice steps done incorrectly with continued feedback.

Both groups (groups A and B) were scheduled to return to the simulation lab 6 months from their initial training date for a final test. The same approach from the 3-month testing sessions was used for the 6-month sessions.

One resident did not complete the 3-month session and a total of six residents did not complete their 6-month testing session due to factors such as early graduation and scheduling conflicts. All remaining residents ($n = 112$) completed their scheduled sessions in the appropriate time frames (mean time between initial and 3-month session was 92.0 days; mean time between initial and 6-month session was 180.3 days).

General Internal Medicine faculty credentialed to perform paracentesis volunteered ($n = 18$) to rate and help train the residents. Prior to completing any ratings, faculty underwent a 1-h standard rater training session. This involved a discussion of the PCAT, study goals, and rating a video performance using the PCAT followed by a debriefing discussion with the PI (DS). Aside from the PI, faculty raters were blinded to which training group each resident was randomized to at the 6-month test. Raters and residents were staggered such that the same rater did not rate the same resident twice. However, nine unblinded evaluations were performed by the PI due to last minute faculty scheduling conflicts.

Statistical Analysis

Descriptive statistics were used to compare baseline characteristics between training groups. For performance on each PCAT component over time, Student's t test was used. The chi-square test was used to evaluate the proportion of residents

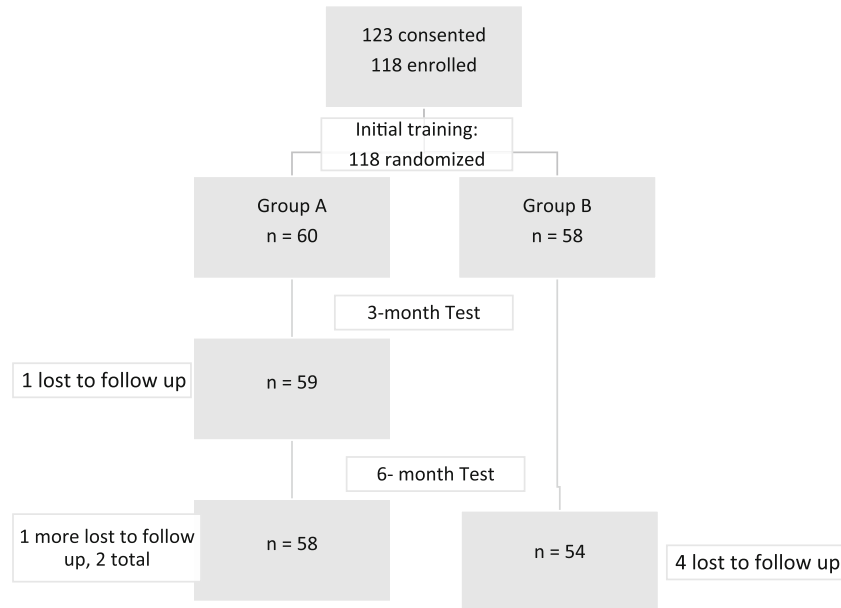


Figure 1 Flow chart of study process.

meeting either the MPS or UPS over time. For all measurements, significance was set at $p < 0.05$. Analysis was completed using R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>).

RESULTS

Following the initial simulation training session, there were no statistically significant differences on average scores on each PCAT component between groups (Table 2). Since a requirement of completing training was to achieve the MPS, all residents in both groups accomplished this after one session. The proportion of residents who met criteria for UPS was also not significantly different between groups (Fig. 2).

At the 3-month session, group A showed significant decline in performance compared with their initial training scores (Table 2; Fig. 2). This was observed for average performance on each PCAT component, as well as for the proportion of residents meeting criteria for either the MPS or UPS. When group A was retested at 6 months, performance on each PCAT component was not significantly different from performance after initial training. The proportion of residents achieving the MPS improved compared with the 3-month test, but was still significantly less than

after initial training ($p = 0.03$). The proportion of residents who achieved the UPS at 6 months, however, was not significantly different compared with after initial training ($p = 0.67$).

At 6 months, group B had significantly lower performance on every PCAT component compared with initial training ($p < .01$ for all comparisons; Table 2). The proportion of residents who met criteria for both the MPS and UPS also significantly declined ($p < .001$ for both comparisons; Fig. 2).

When compared head-to-head at 6 months, group A performed significantly better on all components of the PCAT compared with group B (Table 2; $p < .001$ for each comparison). Of the residents that initially met criteria for either the MPS or UPS after training, the proportion who still met criteria for both categories was higher in group A than in group B (Fig. 2; $p < .001$ for both comparisons). When non-blinded 6-month ratings were removed as part of a sensitivity analysis, scores still significantly declined on each PCAT component between groups, and the proportion of residents meeting MPS and UPS was less in group B as compared with that in group A.

DISCUSSION

Some studies show that procedural skills can be retained after initial training.^{17–21} However, this finding is not uniform. Our study adds to the literature showing decay in other technical skills like surgery,^{22, 23} cardiopulmonary resuscitation,^{24, 25} transthoracic echo,²⁶ and central venous catheter placement.^{27, 28} Our findings suggest that paracentesis performance decays as quickly as 3 months, with mean scores on all three PCAT components declining at 3 months compared with initial training. This difference may be clinically significant given that the proportion of residents who achieved either the MPS or UPS also significantly declined. This is alarming, given the

Table 1 Participant Demographics After Randomization

Demographic		Group A	Group B	<i>p</i> value
Gender	Female	30	26	0.70
	Male	30	32	
Post-graduate year (PGY)	PGY 1	22	23	0.73
	PGY 2	18	18	
	PGY 3	15	15	
	PGY 4	5	2	

Table 2 Performance by Training Group Over Time

Time point	Training group	Checklist score	p value	Global assessment score	p value	Entrustment score	p value
Initial	A	23.12	0.83	2.65	0.87	3.12	0.42
	B	23.05		2.67		2.97	
3 months	A	19.95	< .001	2.32	0.03	2.63	0.01
6 months	A	22.47	< .001	2.74	< .001	3.14	< .001
	B	19.00		2.07		2.33	

Scores by PCAT component for each training group at initial training, the 3-month session (group A only), and the 6-month session. At the initial and 6-month time point, p values compare performance between groups A and B. At the 3-month time point, p values compare performance of group A participants on their initial scores to their 3-month scores

interruption of consecutive procedural opportunities for residents, which often extends beyond 3 or 6 months.

When those that were tested and trained again at 3 months (group A) returned for their final 6-month test, they performed better than those in group B. This is consistent with other studies that show repeated retraining sessions may help interrupt skill decay.^{29, 30} Additionally, while those in group A

initially meeting criteria for MPS showed significant decay compared with initial training, those who met criteria for UPS did not. Others have shown differences in individual learning curves for complex skills³¹ and how this might affect skill decay^{32, 33}, and our findings may suggest that those with higher baseline skills will need less frequent training sessions as time progresses. However, individual differences in skill

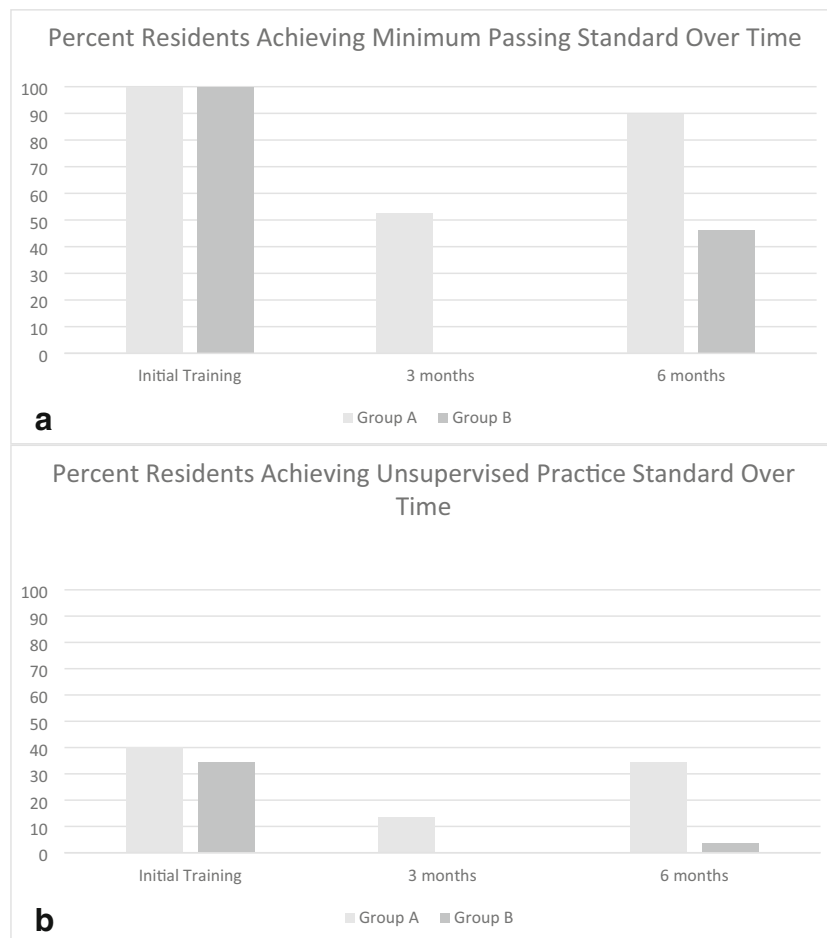


Figure 2 Graphical representation of decline in those achieving either the Minimum Passing Standard (a) or the Unsupervised Practice Standard (b) between groups over time. **a** Comparison between those achieving the Minimum Passing Standard (MPS) between training groups after initial training and at the 3- and 6-month tests. All residents achieved the MPS after initial training. At the 3-month test (group A only), a significantly lower proportion of residents achieved the MPS compared with initial training ($n = 31, 52.5\%, p < .001$). At the 6-month test, a higher proportion of learners in group A retained criteria for MPS ($n = 52, 89.7\%$) than those in group B ($n = 25, 46.3\%; p < .001$). **b** Comparison between those achieving the Unsupervised Practice Standard (UPS) between training groups after initial training and at the 3- and 6-month tests. The proportion of residents achieving UPS after initial training was not different between groups: group A, $n = 24 (40\%)$ compared with group B, $n = 20 (34.5\%), p = 0.67$. At the 3-month test (group A only), a significantly lower proportion of residents achieved the UPS compared with initial training ($n = 8, 13.6\%, p = .002$). At the 6-month test, a higher proportion of learners in group A retained criteria for UPS ($n = 20, 34.5\%$) than those in group B ($n = 2, 3.7\%; p < .001$). In general, the proportion of residents achieving UPS was low at all time points.

decay were not a primary focus of this work, and thus programs may have to create more broad retraining programs for all residents initially.

Another key finding was the small number of residents across class years deemed competent for unsupervised practice. After initial training, when performance should be at its highest, only 37.3% of learners ($n = 44$) achieved UPS. This represented 5 interns (11.1%), 18 PGY-2s (50%), 15 PGY-3s (50%), and 6 PGY-4s (85.7%). Alarming, this further declined to 19.6% of learners ($n = 22$) at 6 months. These 22 learners were comprised of 2 interns (4.8%), 6 PGY-2s (14.1%), 10 PGY-3s (35.7%), and 4 PGY-4s (57.1%). In our residency program's prior process for determining competence (performing three paracenteses on live patients), almost all senior residents and many interns would have been deemed competent without any objective assessment and regardless of time between procedural encounters. This lack of objective competence may help explain other studies that have found residents lack self-confidence in performing procedures or supervising others.^{34, 35} When programs lack more objective measures of competence, and instead rely on procedural numbers as surrogates for competence, unprepared learners are often put in positions where they may lack sufficient skill to perform a procedure safely. When this is compounded with time between procedural opportunities, the risk is even greater.

We hope that use of our tool and our findings will give other programs a framework for training and retraining learners, as well as assessing skill level more objectively. However, retraining learners every 3 months in the simulation setting is resource-intensive, and programs may not have the funds or faculty time to make this feasible. A reasonable alternative may be *just-in-time* training^{36, 37} before procedure heavy rotations like the intensive care unit. However, this approach may fail to capture procedures like paracentesis that can occur sporadically throughout different ward rotations and would need further study.

Limitations

Our study has several limitations. First, all procedure assessments occurred in the simulation setting. Procedures performed on live patients likely introduce different challenges that may affect resident competence over time. Second, assessments from a minimal number of simulated or live procedures are unlikely to be enough to make summative decisions of competence.^{38, 39} Rather, multiple performances on the PCAT in the simulation and live setting will likely be most beneficial to track performance over time and use aggregate assessment data to make such decisions. Future studies are needed to examine how many simulated and live assessments over time are required to make confident summative decisions and for how long one is certified after such decisions before additional assessment should occur. Third, we did not track live patient procedural opportunities between groups over time, and instead chose to examine the data in an intent-to-treat type fashion. Since we now require faculty supervision for all live

procedures, residents have less incentive to log procedures (i.e., logging a certain number will not automatically permit them to perform the procedure independently). As a result, resident logs have been unreliable in our program and we did not want to introduce recall bias into our results. Fourth, nine ratings at the 6-month test were unblinded due to faculty rater scheduling conflicts. However, to compensate for this, we performed a sensitivity analysis that did not show any significant change in the results. Finally, six learners were lost to follow-up during the study due to scheduling conflicts, illness, or early graduation. However, the number was small and relatively evenly distributed between study groups and did not likely affect the results.

CONCLUSIONS

We found that using an evidence-based assessment framework for evaluating paracentesis competence would likely reclassify many learners previously deemed competent to perform paracentesis unsupervised to those who continue to need training and/or supervision. Additionally, significant skill decay occurred as early as 3 months, although further decay was mitigated with retraining. Procedural competence requires frequent retraining, and programs need to determine how best to implement such training with potentially limited resources. Further study is needed to see if these trends persist in the live environment.

Acknowledgments: The authors wish to thank the Cincinnati Veteran's Affairs Medical Center for utilization of the simulation center and resources, without which this project would not have been possible. Specifically, the authors appreciate the support, guidance, and coordination from Dr. Kathleen Chard, Ph.D, LeAnn Schlamb, MSN, RN, Ed.S, and Leroy Mendell, RN.

Corresponding Author: Dana Sall, MD, MEd; HonorHealth Scottsdale Thompson Peak Medical Center, Scottsdale, AZ, United States (e-mail: dana.sall@gmail.com).

Funding This work was supported by resources from and use of facilities at the Cincinnati Veteran's Affairs Medical Center.

Compliance with Ethical Standards:

Conflict of Interest: Dr. Jennifer O'Toole consulted with and received honoraria payment from the I-PASS Patient Safety Institute. She also holds stock options in the I-PASS Patient Safety Institute, a non-publicly traded company.

Ethical Approval: Ethical approval was obtained from the University of Cincinnati IRB, study number 9277-2015.

Disclaimers: This work does not represent the views of the U.S Department of Veteran's Affairs or the United States Government.

REFERENCES

1. Barsuk JH, Feinglass J, Kozmic SE, Hohmann SF, Ganger D, Wayne DB. Specialties performing paracentesis procedures at university hospitals: implications for training and certification. *J Hosp Med*. 2014;9(3):162-168.

2. **Huang GC, Newman LR, Schwartzstein RM, et al.** Procedural competence in internal medicine residents: validity of a central venous catheter insertion assessment instrument. *Acad Med.* 2009;84(8):1127-1134.
3. **Barsuk JH, Cohen ER, Vozenilek JA, O'Connor LM, McGaghie WC, Wayne DB.** Simulation-based education with mastery learning improves paracentesis skills. *J Grad Med Educ.* 2012;4(1):23-27.
4. **Sawyer T, White M, Zaveri P, et al.** Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Acad Med.* 2015;90(8):1025-1033.
5. **Ericsson KA.** Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. *Acad Med.* 2015;90(11):1471-1486.
6. **Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB.** Residents' procedural experience does not ensure competence: a research synthesis. *J Grad Med Educ.* 2017;9(2):201-208.
7. **Barsuk JH, Cohen ER, Caprio T, McGaghie WC, Simuni T, Wayne DB.** Simulation-based education with mastery learning improves residents' lumbar puncture skills. *Neurology.* 2012;79(2):132-137.
8. **McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB.** Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med.* 2011;86(6):706-711.
9. **Wayne DB, Barsuk JH, O'Leary KJ, Fudala MJ, McGaghie WC.** Mastery learning of thoracentesis skills by internal medicine residents using simulation technology and deliberate practice. *J Hosp Med.* 2008;3(1):48-54.
10. **Crocker JT, Hale CP, Vanka A, Ricotta DN, McSparron JI, Huang GC.** Raising the bar for procedural competency among hospitalists. *Ann Intern Med.* 2019.
11. **Sall D, Wigger GW, Kinnear B, Kelleher M, Warm E, O'Toole JK.** Paracentesis simulation: a comprehensive approach to procedural education. *MedEdPORTAL.* 2018;14:10747.
12. **Vaisman A, Cram P.** Procedural competence among faculty in academic health centers: challenges and future directions. *Acad Med.* 2017;92(1):31-34.
13. **Thakkar R, Wright SM, Alguire P, Wigton RS, Boonyasai RT.** Procedures performed by hospitalist and non-hospitalist general internists. *J Gen Intern Med.* 2010;25(5):448-452.
14. **Barsuk JH, Cohen ER, Feinglass J, et al.** Cost savings of performing paracentesis procedures at the bedside after simulation-based education. *Simul Healthc.* 2014;9(5):312-318.
15. **Frank JR, Snell LS, Cate OT, et al.** Competency-based medical education: theory to practice. *Med Teach.* 2010;32(8):638-645.
16. **Norcini JJ.** Setting standards on educational tests. *Med Educ.* 2003;37(5):464-469.
17. **Ahya SN, Barsuk JH, Cohen ER, Tuazon J, McGaghie WC, Wayne DB.** Clinical performance and skill retention after simulation-based education for nephrology fellows. *Semin Dial.* 2012;25(4):470-473.
18. **Barsuk JH, Cohen ER, McGaghie WC, Wayne DB.** Long-term retention of central venous catheter insertion skills after simulation-based mastery learning. *Acad Med.* 2010;85(10 Suppl):S9-12.
19. **Didwania A, McGaghie WC, Cohen ER, et al.** Progress toward improving the quality of cardiac arrest medical team responses at an academic teaching hospital. *J Grad Med Educ.* 2011;3(2):211-216.
20. **Ciurzynski SM, Gottfried J, Pietraszewski J, Zalewski M.** Impact of training frequency on nurses' pediatric resuscitation skills. *J Nurses Prof Dev.* 2017;33(5):E1-e7.
21. **Iserbyt P, Mols L.** Retention of CPR skills and the effect of instructor expertise one year following reciprocal learning. *Acta Anaesthesiol Belg.* 2014;65(1):23-29.
22. **Bekele A, Wondimu S, Firdu N, Taye M, Tadesse A.** Trends in retention and decay of basic surgical skills: evidence from Addis Ababa University, Ethiopia: a prospective case-control cohort study. *World J Surg.* 2019;43(1):9-15.
23. **Bonrath EM, Weber BK, Fritz M, et al.** Laparoscopic simulation training: testing for skill acquisition and retention. *Surgery.* 2012;152(1):12-20.
24. **Braun L, Sawyer T, Smith K, et al.** Retention of pediatric resuscitation performance after a simulation-based mastery learning session: a multicenter randomized trial. *Pediatr Crit Care Med.* 2015;16(2):131-138.
25. **Smith KK, Gilcreast D, Pierce K.** Evaluation of staff's retention of ACLS and BLS skills. *Resuscitation.* 2008;78(1):59-65.
26. **Yamamoto R, Clanton D, Willis RE, Jonas RB, Cestero RF.** Rapid decay of transthoracic echocardiography skills at 1 month: a prospective observational study. *J Surg Educ.* 2018;75(2):503-509.
27. **Thomas SM, Burch W, Kuehnle SE, Flood RG, Scalzo AJ, Gerard JM.** Simulation training for pediatric residents on central venous catheter placement: a pilot study. *Pediatr Crit Care Med.* 2013;14(9):e416-423.
28. **Barsuk JH, Cohen ER, Nguyen D, et al.** Attending physician adherence to a 29-component central venous catheter bundle checklist during simulated procedures. *Crit Care Med.* 2016;44(10):1871-1881.
29. **Cecilio-Fernandes D, Cnossen F, Jaarsma D, Tio RA.** Avoiding surgical skill decay: a systematic review on the spacing of training sessions. *J Surg Educ.* 2018;75(2):471-480.
30. **Mduma E, Ersdal H, Svensen E, Kidanto H, Auestad B, Perlman J.** Frequent brief on-site simulation training and reduction in 24-h neonatal mortality—an educational intervention study. *Resuscitation.* 2015;93:1-7.
31. **Warm EJ, Held JD, Hellmann M, et al.** Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. *Acad Med.* 2016;91(10):1398-1405.
32. **Boutis K, Pecaric M, Carriere B, et al.** The effect of testing and feedback on the forgetting curves for radiograph interpretation skills. *Med Teach.* 2019:1-9.
33. **Crofts JF, Bartlett C, Ellis D, Hunt LP, Fox R, Draycott TJ.** Management of shoulder dystocia: skill retention 6 and 12 months after training. *Obstet Gynecol.* 2007;110(5):1069-1074.
34. **Mourad M, Kohlwes J, Maselli J, Auerbach AD.** Supervising the supervisors—procedural training and supervision in internal medicine residency. *J Gen Intern Med.* 2010;25(4):351-356.
35. **Huang GC, Smith CC, Gordon CE, et al.** Beyond the comfort zone: residents assess their comfort performing inpatient medical procedures. *Am J Med.* 2006;119(1):71.e17-24.
36. **Barsuk JH, Cohen ER, Potts S, et al.** Dissemination of a simulation-based mastery learning intervention reduces central line-associated bloodstream infections. *BMJ Qual Saf.* 2014;23(9):749-756.
37. **Aggarwal R.** Just-in-time simulation-based training. *BMJ Qual Saf England.* 26:2017.
38. **Warm EJ, Mathis BR, Held JD, et al.** Entrustment and mapping of observable practice activities for resident assessment. *J Gen Intern Med.* 2014;29(8):1177-1182.
39. **Schuwirth LW, Southgate L, Page GG, et al.** When enough is enough: a conceptual basis for fair and defensible practice performance assessment. *Med Educ.* 2002;36(10):925-930.

Publisher's Note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.