

# The Variable Journey in Learning to Interpret Pediatric Point-of-care Ultrasound Images: A Multicenter Prospective Cohort Study

Charisse Kwan, MD<sup>1</sup>, Martin Pusic, MD, PhD<sup>2</sup>, Martin Pecaric, PhD<sup>3</sup>, Kirstin Weerdenburg, MD<sup>4</sup>, Mark Tessaro, MD<sup>1</sup> , and Kathy Boutis, MD, MSc<sup>1</sup> 

## ABSTRACT

**Objectives:** To complement bedside learning of point-of-care ultrasound (POCUS), we developed an online learning assessment platform for the visual interpretation component of this skill. This study examined the amount and rate of skill acquisition in POCUS image interpretation in a cohort of pediatric emergency medicine (PEM) physician learners.

**Methods:** This was a multicenter prospective cohort study. PEM physicians learned POCUS using a computer-based image repository and learning assessment system that allowed participants to deliberately practice image interpretation of 400 images from four pediatric POCUS applications (soft tissue, lung, cardiac, and focused assessment sonography for trauma [FAST]). Participants completed at least one application (100 cases) over a 4-week period.

**Results:** We enrolled 172 PEM physicians (114 attendings, 65 fellows). The increase in accuracy from the initial to final 25 cases was 11.6%, 9.8%, 7.4%, and 8.6% for soft tissue, lung, cardiac, and FAST, respectively. For all applications, the average learners (50th percentile) required 0 to 45, 25 to 97, 66 to 175, and 141 to 290 cases to reach 80, 85, 90, and 95% accuracy, respectively. The least efficient (95th percentile) learners required 60 to 288, 109 to 456, 160 to 666, and 243 to 1040 cases to reach these same accuracy benchmarks. Generally, the soft tissue application required participants to complete the least number of cases to reach a given proficiency level, while the cardiac application required the most.

**Conclusions:** Deliberate practice of pediatric POCUS image cases using an online learning and assessment platform may lead to skill improvement in POCUS image interpretation. Importantly, there was a highly variable rate of achievement across learners and applications. These data inform our understanding of POCUS image interpretation skill development and could complement bedside learning and performance assessments.

From the <sup>1</sup>Division of Pediatric Emergency Medicine, Department of Pediatrics, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada; the <sup>2</sup>Department of Emergency Medicine and Division of Learning Analytics, NYU School of Medicine, New York, NY; <sup>3</sup>Contrail Consulting Services, Toronto, Ontario, Canada; and the <sup>4</sup>Department of Emergency Medicine, IWK Health Centre and Dalhousie University, Halifax, Nova Scotia, Canada.

Received May 17, 2019; revision received June 17, 2019; accepted June 20, 2019.

This study was funded by the Ambulatory Pediatrics Association. The funders had no role in the execution, results, or manuscript.

Product disclosure: After completion of this research, we included this POCUS case-experience into our ImageSim Continued Professional Development and Training platform, which is now available for purchase. ImageSim operates as nonprofit course at the Hospital for Sick Children and the University of Toronto ([www.imagesim.com](http://www.imagesim.com)). CK, KW, MPu, MT, and KB do not receive funds for their participation in development or management of this education intervention that is now part of ImageSim. Dr. Martin Pecaric was paid for 10% of the IT development work from the APA grant received for this research, and the other 90% of the effort was provided free of charge. Dr. Pecaric provides IT support to the ImageSim POCUS CPD and Training course and continues to be paid for 10% of his contributions under a contract with the Hospital for Sick Children. Dr. Boutis is married to Dr. Pecaric and this relationship has been disclosed to the Hospital for Sick Children and the University of Toronto.

Supervising Editor: Michael Gottlieb, MD, RDMS.

Address for correspondence and reprints: Kathy Boutis, MD MSc; e-mail [kathy.boutis@sickkids.ca](mailto:kathy.boutis@sickkids.ca).

AEM EDUCATION AND TRAINING 2020;4:111–122.

The use of emergency point-of-care ultrasound (POCUS) in pediatric emergency medicine (PEM) can improve patient outcomes and expedite patient disposition.<sup>1</sup> As such, learning POCUS has become an increasing priority in PEM, with the learning experience typically including introductory courses and skill acquisition that relies on case-by-case clinical exposure.<sup>2-5</sup> However, this approach may lead to challenges in achieving POCUS proficiency across a wide range of learners.<sup>6</sup> Opportunities for bedside feedback may be limited by the number of POCUS-trained attendings available at a given site.<sup>2</sup> This deprives the learner of immediate feedback, one of the most powerful methods of skill acquisition.<sup>7</sup> Further, since the baseline pathology rate in pediatrics is relatively low, relying on case-by-case exposure to achieve sufficient skill may take years with clinical practice alone.

Gaining POCUS expertise is multifaceted and complex since the technique includes image acquisition, image interpretation, and integration of interpretation into clinical decision making. To facilitate learning of complex tasks, instructional design models recommend intentionally alternating part-task with whole-task training.<sup>8</sup> In this light, e-learning provides an opportunity to expose learners to an image interpretation learning experience (i.e., part-task) that could complement the resource intensive face-to-face teaching and learning at the bedside that addresses all facets of POCUS (i.e., whole-task).<sup>2</sup> However, to date, most POCUS simulation and online education focuses on image acquisition,<sup>3,5</sup> specific conditions (e.g., cardiac tamponade),<sup>9</sup> or adult applications.<sup>10-12</sup> As such, existing teaching tools are limited in being able to significantly improve learner exposure to the part-task of pediatric POCUS image interpretation. However, Web-based learning and assessment image banks that provide intentional sequencing and targeted analytic feedback on hundreds of cases have had demonstrated success for increasing electrocardiogram and musculoskeletal radiograph image interpretation skill.<sup>13-15</sup> Using these types of learning platforms, the presentation of images is simulated to mirror how clinicians interpret them at the bedside. Specifically, cases are presented with a brief clinical stem, standard images and views, and juxtaposition of normal and abnormal cases.<sup>16,17</sup> The cases are also presented in large numbers so that learners can learn similarities and differences between diagnoses, identify weaknesses, and build up a global representation of possible diagnoses.<sup>15,16,18</sup> After each case the system provides visual

and text feedback, which allows for deliberate practice<sup>19</sup> and an ongoing measure of performance as part of the instructional strategy.<sup>20,21</sup> This type of learning assessment platform could be applied to the image interpretation component of POCUS and potentially improve our understanding of POCUS image interpretation skill development.

We developed a POCUS image interpretation learning and assessment system that included four common pediatric POCUS applications (100 cases/application): soft tissue, lung, cardiac, and focused assessment sonography for trauma (FAST).<sup>3</sup> We sought to determine PEM physician performance metrics and the number of cases and time within which most participants could achieve specific performance benchmarks.

## METHODS

### Education Intervention

We used previously established methods to develop the education intervention.<sup>15-17,22</sup> Deidentified images in the four POCUS applications acquired using the departmental ultrasound machine (Zonare Medical Systems) were exported from the local POCUS archive (Q-Path, Telexy Healthcare) in JPEG (still images) or MPEG-4 (cine-clips) formats. From this pool, two PEM POCUS fellowship-trained physicians selected a consecutive sample of 100 cases per application (400 total) that demonstrated acceptable image quality and a spectrum of pathology and normal anatomy. Any images with embedded technical clues that pointed to a diagnosis were excluded. Further, each application contained 50 cases with and 50 cases without pathology.<sup>17</sup> For lung and cardiac applications, both video and still images were used because recognition of movement is key to interpretation. For soft tissue and FAST (except cardiac view) still images alone were used since these were more likely to adequately capture the subtleties distinguishing abnormal and normal cases and were more cognitively efficient for learners.<sup>23</sup>

As the skill of assigning clinical significance and a specific diagnosis to POCUS findings (e.g., pneumonia) is best assessed at the bedside, the main goal of this education intervention was to develop the skill of distinguishing normal from any abnormal POCUS findings on video/still images, alongside forcing the learner to visually locate any abnormality for cases allocated as “abnormal.” The key initial educational outcome of accurately distinguishing normal from abnormal cases (vs. confirming a specific diagnosis) is also in keeping with

the user goals pediatric POCUS at the bedside since identifying an abnormal POCUS image should then prompt the physician to consider additional tests or consultations as needed to confirm or refine the diagnosis.<sup>24</sup> Nevertheless, learning the possible specific diagnoses from POCUS imaging findings is also important. As a result, this information is presented with every case for the participant to consider in the text feedback. This approach satisfies the essential learning goals for pediatric POCUS, while providing additional information to participants to also learn specific diagnostic interpretations for each case.

All images were reviewed jointly by two POCUS experts and classified as POCUS images as normal versus abnormal, where abnormal images had changes that could be considered pathological (Table 1). A third POCUS expert then independently provided normal/abnormal classifications, and discrepancies between the classifications were resolved by consensus ( $\kappa = 0.86$ , 95% confidence interval [CI] 0.81 to 0.91). Two POCUS experts then subsequently marked up the abnormal images using graphics to highlight pathology, creating clickable areas over the pathology with 2 to 3 mm of allowance just outside the abnormal area. These images were then embedded into a template generated using a Flash-integrated development environment.

**Table 1**  
Pathology Presented by the Learning and Assessment System for Each POCUS Application

Pathology ( <i>n</i> = 50) <sup>†</sup>	<i>n</i> (%)
<b>Soft tissue</b>	
Cellulitis and abscess	29 (58)
Cellulitis	19 (38)
Foreign body	2 (4)
<b>Lung</b>	
Pneumonia, with and without pleural effusion	25 (50)
Inflammation/bronchiolitis	20 (40)
Pneumothorax	3 (6)
Pulmonary edema	2 (4)
<b>Cardiac</b>	
Pericardial effusion	32 (64)
Poor function	18 (36)
<b>FAST</b>	
Multiple areas of free fluid	25 (50)
Pelvic free fluid	8 (16)
Right or left upper quadrant free fluid	12 (24)
Pericardial effusion	5 (10)

FAST = focused assessment sonography in trauma.

<sup>†</sup>50 cases with pathology per application; 50 cases without pathology to make up 100 total.

A website was developed using HTML, PHP, and a MySQL database.<sup>15</sup> In brief, once a participant was provided unique access, they were taken to the online system and presented with 100 cases per application. For each case, they considered a brief clinical description stem and an unmarked POCUS image (still  $\pm$  video). Cases were presented in random order unique to each participant. After review of the case, the user declared the case as definitely normal, probably normal, probably abnormal, or definitely abnormal. The definitely/probably assignments permitted the user to express the certainty of their response. If their answer was in the “abnormal” category, the user was then required to designate one area of abnormality using an interactive system. When ready, the participant submitted their response, after which they received immediate visual and written feedback on the correctness of their response, diagnosis of the case, and normal anatomy (demonstration at <https://imagesim.com/demo/> or Figure 1). Prior to launch, the system was pilot tested on five participants (one POCUS nonexpert, four POCUS experts) for technical and content issues and revised in consequence.

## Study Design and Setting

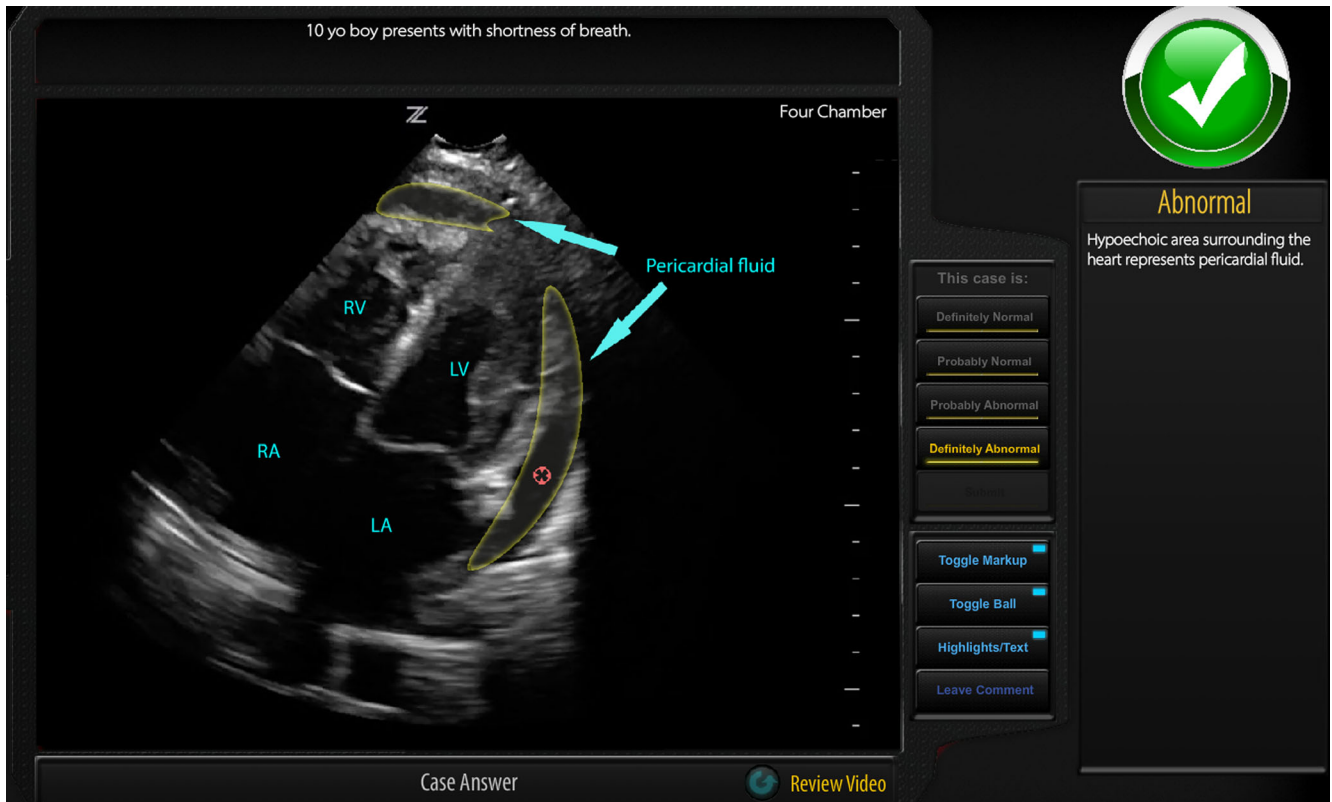
The education tool was developed in collaboration with two tertiary care pediatric centers and an industry partner. This research was undertaken using a multicenter prospective cohort design. We recruited a convenience sample of PEM physicians in the United States and Canada from September to November 2018.

## Selection of Participants

An e-mail was sent to PEM division heads, PEM fellowship program directors, and P2 Network (an international collaborative of pediatric POCUS physicians, <https://p2network.com/>) site leads asking them to forward the e-mail to their respective fellows and attending physicians. Interested participants contacted the study coordinator. We also recruited five “expert” PEM POCUS physicians (separate from study team) who had each completed a PEM POCUS fellowship and performed at least 1,500 bedside scans.<sup>25</sup> This study was approved by the institutional review boards at the study institutions.

## Measurements

Secure entry was ensured via unique participant login credentials and access to the system was available 24 hours per day, 7 days per week. We collected information on type of learner (fellow vs. attending), geographic



**Figure 1.** Visual feedback includes labeling of normal anatomy as well as any abnormalities (if present), and the red circle is the token placed by the user within the mask designating the abnormal region. Text feedback details if the case is normal or abnormal and the specific abnormalities where present. LA = left atrium; LV = left ventricle; RA = right atrium; RV = right ventricle.

location, and self-reported POCUS scans completed (none, <50, 51–100, 101–500, 501–1,000). Participants were asked to complete an introductory tutorial and at least one of the four 100-case POCUS applications. The system automatically time stamped the time a case was started to the time a case response was submitted. Participants were given a time limit of 4 weeks to minimize decay of learning that may confound results.<sup>26</sup> At 2 weeks, participants who had not completed at least one application were sent an e-mail reminder.

## Outcomes

By surveying 150 members of the P2 network, we determined that the performance benchmarks of 80, 85, 90, and 95% accuracy could be considered educationally meaningful for a variety of contexts. Thus, we provided data that demonstrated learning curves of the participants and the median number of cases that participants need to complete to achieve the latter performance benchmarks (primary outcome). Since we anticipated that there would be a variation between participants to achieve these performance benchmarks, we provided this data separately for the average (50th percentile) and least efficient learner (95th percentile) of nonexpert participants.

We also examined the change in accuracy per application between the first and last 25 cases.<sup>26</sup> Further, we measured the effect size of learning gains, changes in sensitivity and specificity per application, and differences in learning gains between fellows versus attendings and made comparisons between applications. We examined if there was any association between demographic variables and the odds of achieving expert-level performance. Expert-level data were also used to examine relations-to-other variables validity (construct),<sup>27</sup> where we would expect that expert performance would be significantly higher than those that are relative non-experts. Finally, to examine feasibility, we calculated average amount of time spent by participants per case and per 100-case module.

## Data Analyses

**Sample Size.** From our previous work, the educationally important difference between initial and final scores was approximately 10% and the proportion of discordant pairs is about 12%.<sup>26</sup> We also assumed that  $\alpha = 0.05$  and  $\beta = 0.80$ . Using a McNemar test power analysis (PASS 11), we calculated a minimal sample size of 95 of PEM POCUS nonexpert participants per application.

**Scoring.** Participants were scored only on the broad category selections of “normal” or “abnormal” and not the subassignments of “probably or definitely” normal/abnormal. Specifically, normal items were scored dichotomously, while abnormal items were scored correct if the participant classified the case as abnormal and indicated one correct region of abnormality.

**Number of Cases to a Performance Benchmark.** We modeled the learning curves of each individual participant with a random coefficients hierarchical logistic regression model.<sup>28</sup> Using these learning curves, we predicted the median number of cases required for a participant to attain a performance benchmark by solving the individual regression equation for the number of cases required to achieve the log odds (logit) that would correspond to the selected performance benchmark. From these data, we also plotted the proportion of participants that would reach 95% accuracy for a given number of cases to determine if these histograms were normally distributed by testing for both skewness and kurtosis of the distributions.<sup>29</sup>

**Learning Outcomes.** Using the initial (pre) and final (post) 25 cases, we calculated change in accuracy, sensitivity, specificity, and Cohen’s d-effect sizes for each application, with respective 95% CI. We analyzed for the association of achieving POCUS expert accuracy performance (dependent variable) and a priori selected independent variables using a logistic regression model. The independent variables included years in practice ( $\leq 5$  years vs.  $> 5$  years), academic setting (university affiliated children’s hospital vs. other),

POCUS training during fellowship (yes vs. no), number of any type of scan completed prior to participation ( $\geq 100$  vs.  $< 100$ ), and accuracy score on initial 20 cases ( $\geq 80\%$  vs.  $< 80\%$ ).

**Comparisons Between Applications.** Independent and dependent normally distributed parametric data were compared with a Student’s t-test and paired Student’s t-test, respectively. Analysis of variance (ANOVA) testing was used to compare three or more means from independent proportions, and the Bonferonni test was used to perform post hoc analyses.

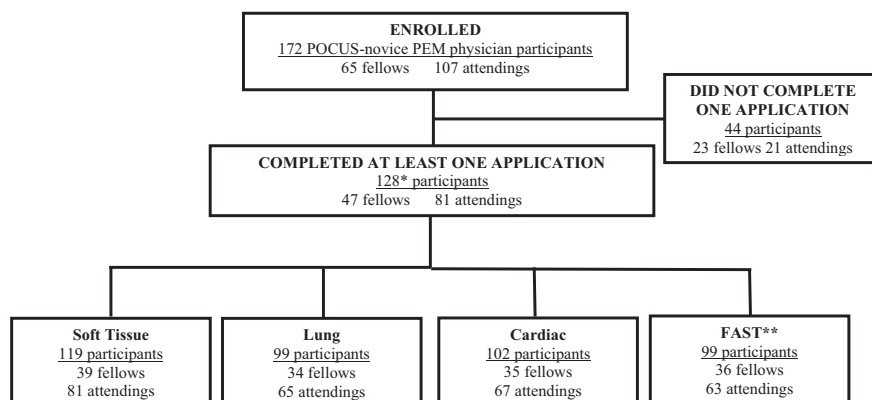
**Time Commitment.** We determined the median time it took to complete each 100-case application with respective interquartile range. We compared the time commitment between applications using the Kruskal-Wallis test.

Significance was set at  $p < 0.01$  to account for multiple testing. All analyses except the regression analyses were carried out using SPSS (Version 23, IBM 2015). The regression models were performed using Stata (Version 14, StataCorp LLC).

## RESULTS

### Study Participants

We enrolled 177 PEM physicians. Of these, 172 were PEM POCUS learner physicians (fellows [ $n = 65$ ] and attendings [ $n = 107$ ]) and five were PEM POCUS experts (Figure 2). Participants represented 28 (56%) of the 50 US states and five (50%) of the 10 Canadian provinces. There were significantly more fellows than attendings that received POCUS training as part



**Figure 2.** POCUS PEM physician participation. \*Sum total of all specific application participants was greater than 128 since most participants completed more than one application. FAST = focused assessment sonography for trauma; PEM = pediatric emergency medicine; POCUS = point-of-care ultrasound.

of their fellowship experience (82.0% vs. 30.8%; difference = 51.2% [95% CI = 36.2–61.7]; Table 2).

At least one application was completed by 128 (74.4%) participants (Figure 2), 88 (68.8%) of whom completed all four applications, while 11 (8.6%) completed three, seven (5.5%) completed two, and 22 (17.2%) completed one application.

### Number of Cases to Performance Benchmarks

A qualitative review of the modeled learning curves demonstrates significant variation between participants in rates of achieving higher performance (Figure 3). For the average (50th percentile) learners, the predicted median number of cases needed across our four applications was 0 to 45 for 80% accuracy, 25 to 97 for 85% accuracy, 66 to 175 for 90% accuracy, and 141 to 290 for 95% accuracy (Figure 4A). The least

efficient (95th percentile) of learners would have to complete 60 to 288 cases to achieve 80% accuracy, 109 to 456 cases for 85% accuracy, 160 to 666 cases for 90% accuracy, and 243 to 1,040 cases for 95% accuracy. Participants required the highest number of cases for the cardiac application and the lowest number for the soft tissue application to reach a specified benchmark (Figure 4B).

The distribution of the number of participants that would achieve the expert benchmark for a given number of cases was skewed for all applications ( $p < 0.0001$ ). The lung and cardiac distributions also demonstrated kurtosis ( $p < 0.0001$ ; Figure 5).

### Performance Outcomes

The pre/post change in accuracy for each application is detailed in Table 3. The Cohen's d-effect sizes for each application were moderate to large and ranged

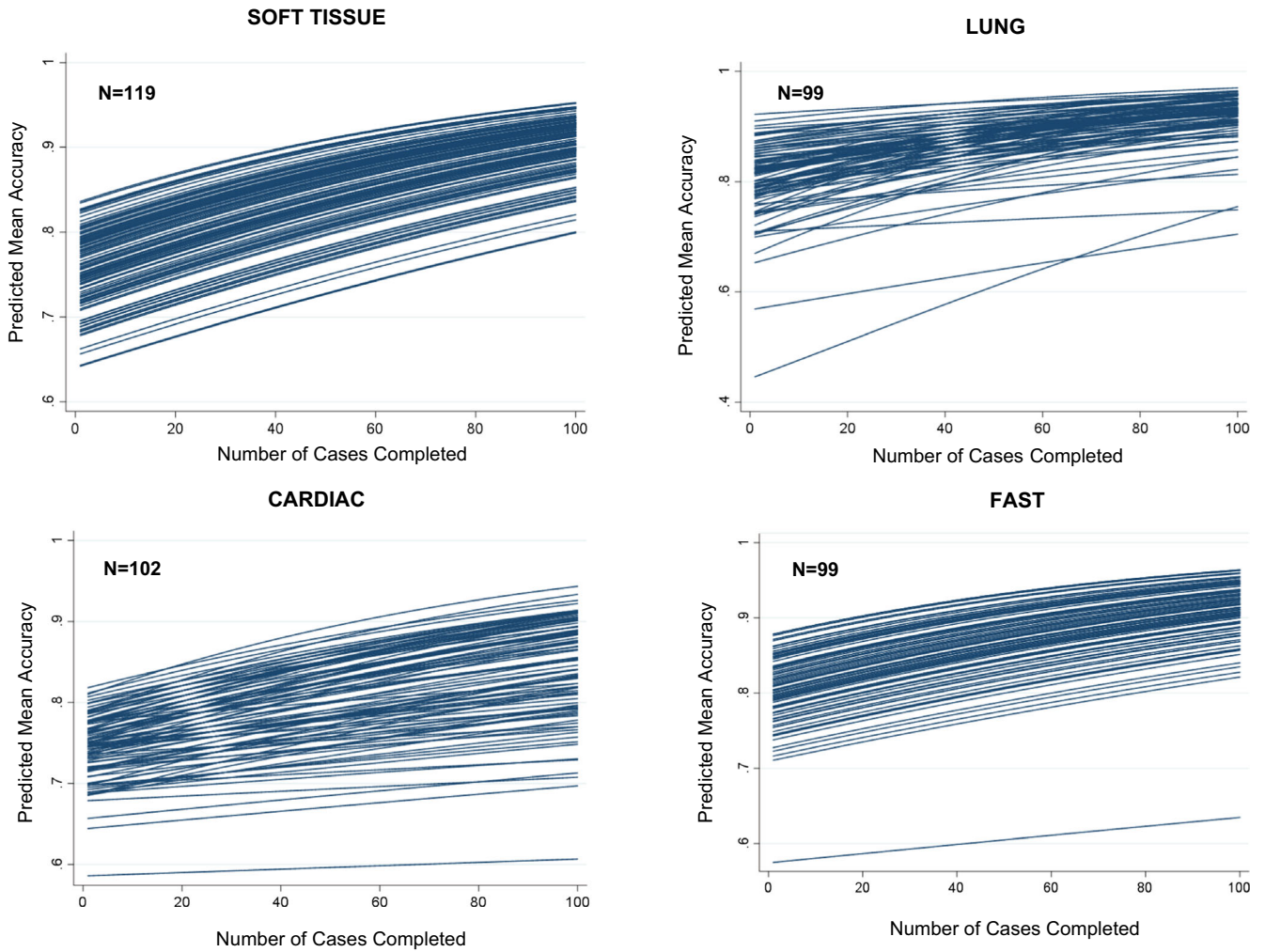
**Table 2**  
Participant Demographics

Demographic Variable	PEM Fellow N=65	PEM Attending N=107	p-value
Participant from the United States, no. (%)	52 (80.0)	67 (62.6)	0.02
Years since completing postgraduate training, no. (%)			
< 6 years	NA <sup>†</sup>	44 (41.1)	NA <sup>†</sup>
6-10 years		20 (18.7)	
11-15 years		13 (12.1)	
16-20 years		10 (9.3)	
> 20 years		20 (18.7)	
Female sex, no. (%)	44 (72.1)	72 (63.2)	0.23
Employment Setting, no. (%)			
University affiliated general/community hospital	3 (4.6)	6 (5.6)	0.53 <sup>§</sup>
University affiliated children's hospital	61 (93.8)	99 (92.5)	
Non-university affiliated general/community hospital	2 (3.1)	1 (0.9)	
Non-university affiliated children's hospital	0	1 (0.9)	
Type of Point-of-Care Ultrasound Training, no. (%) <sup>‡</sup>			
None	2 (3.1)	13 (12.1)	<0.0001 <sup>§</sup>
Integrated into emergency medicine residency training	3 (4.6)	9 (8.4)	
Integrated into pediatric residency training	3 (4.6)	1 (0.9)	
Integrated into PEM fellowship training	53 (82.0)	33 (30.8)	
Workshops/ Courses	7 (10.8)	75 (70.70)	
Institutional faculty training	NA	61 (57.0)	
Self-directed learning	8 (12.3)	27 (25.2)	
Number of educational/clinical point-of-care ultrasound examinations completed, no. (%)			
None	2 (3.1)	5 (4.4)	0.12
<50	30 (45.2)	43 (40.4)	
51-100	17 (26.2)	20 (18.4)	
101-500	15 (23.1)	28 (26.3)	
501-1000	1 (0.2)	15 (14.0)	

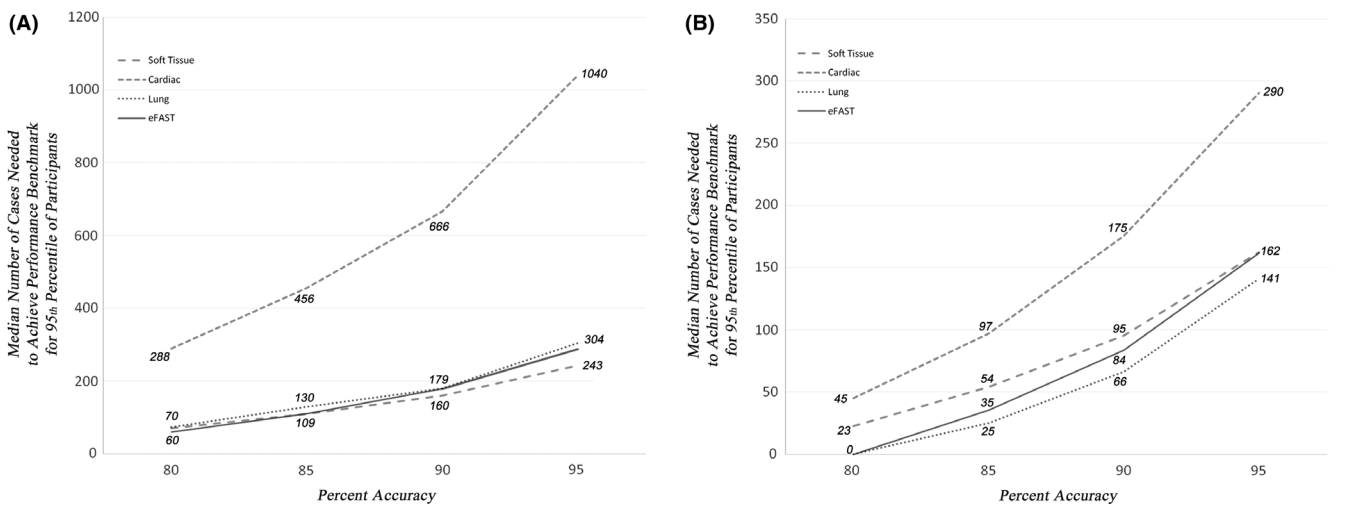
†NA - Not applicable

‡Participants could choose all that apply

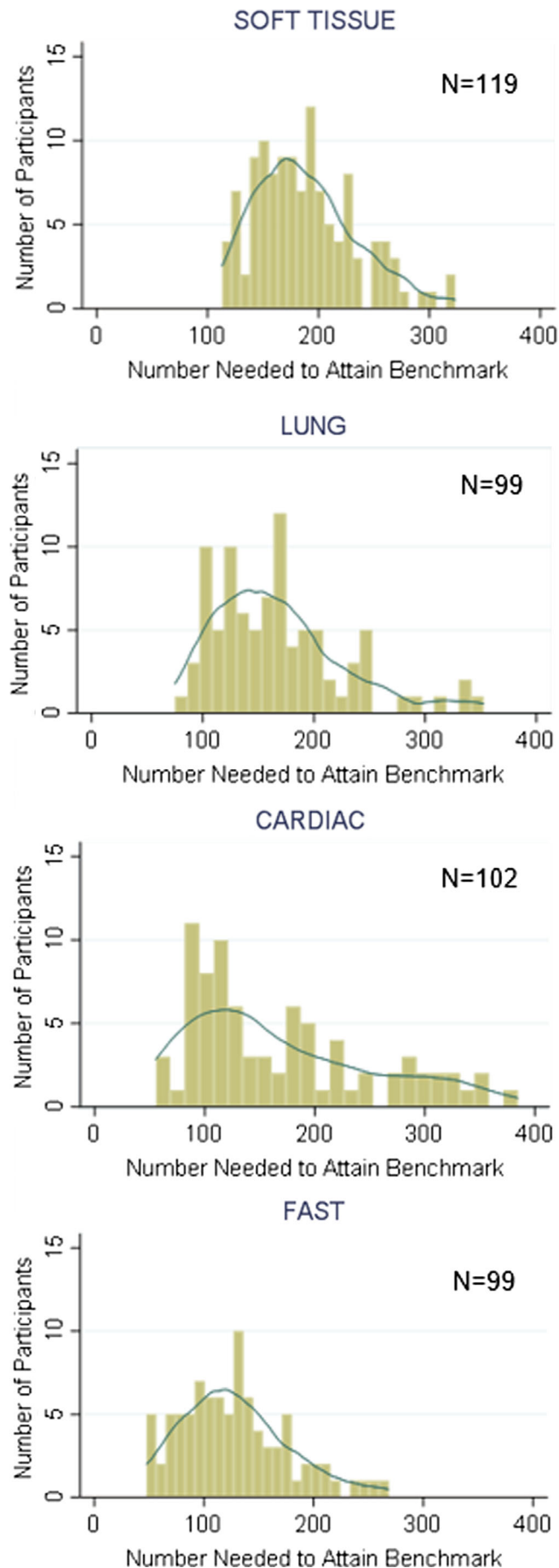
§p-value comparing distributions in each learner group



**Figure 3.** Predicted learning curves for soft tissue, lung, cardiac, and FAST. FAST = focused assessment sonography for trauma.



**Figure 4.** (A) The x-axis represents the benchmark level of proficiency as chosen by the educator. The y-axis represents the predicted number of cases required to achieve that benchmark based on the logistic regression model described in the text. The two panels represent two different educational contexts. (A) The median learner is represented—50% of learners required fewer cases to achieve the benchmark; 50% required more. (B) A higher ambition is presented: the number of cases required of the marginal learner so that 95% of the learners will have achieved the given benchmark (i.e., 95% of learners would need to do fewer; 5% would need to do even more). FAST = focused assessment sonography for trauma.



**Figure 5.** The predicted number of cases needed to achieve 95% accuracy, based on the modeled learning curve for each individual. The overlaid curves are smoothed kernel density plots, which demonstrate that all four distributions are significantly skewed (i.e., not normally distributed). FAST = focused assessment sonography for trauma.

from 0.6 to 1.0.<sup>30</sup> There were no differences in fellow versus attending final accuracy, sensitivity, or specificity scores for soft tissue, lung, or cardiac applications. For FAST, however, attendings had higher final accuracy, (+6.0% difference; 95% CI = 1.8, to 10.2) and sensitivity (+7.5% difference; 95% CI = 0.6 to 14.4). There was no association of PEM POCUS physician learners achieving expert performance with any of the baseline variables: >100 POCUS scans experience versus <100 (odds ratio [OR] = 1.2, 95% CI = 0.5 to 2.6), initial accuracy >80% versus <80% (OR = 2.1, 95% CI = 0.8 to 5.6), <5 versus >5 years in practice (OR = 1.4, 95% CI = 0.5 to 3.5), children's hospital versus other setting (OR = 1.6, 95% CI = 0.3 to 8.8), POCUS training in fellowship versus none (OR = 2.1, 95% CI = 0.9 to 4.5). Per application, PEM POCUS expert mean (95% CI) accuracy scores were soft tissue 96.0% (92.3% to 99.7%), lung 96.0% (93.9% to 98.1%), cardiac 90.0% (81.8% to 98.2%), and FAST 93.0% (88.0% to 98.0%). Expert final scores were significantly higher than those of nonexpert participant scores, with an accuracy difference of 7.3% (4.4% to 10.4%).

### Comparisons Between Applications

The final accuracy sensitivity and specificity (Table 3) differed between applications (ANOVA  $p < 0.001$ ), with post hoc Bonferroni analyses showing that the cardiac application had lower final accuracy ( $p < 0.01$ ), sensitivity ( $p < 0.01$ ), and specificity ( $p < 0.01$ ) relative to each of the other applications. For the soft tissue application, there were greater learning gains for normal cases (specificity = +18.9%) versus abnormal cases (sensitivity = +4.9%), with a difference of 14.0% (95% CI = 9.8% to 18.2%). There were no differences in learning gains for abnormal versus normal cases in the other applications.

### Time Commitment

The initial (first 25 case) mean time per case was 31.7 seconds, which decreased to 21.2 seconds on the final 25 cases (difference = -10.5 seconds, 95% CI = -8.6 to -12.4). The median time (interquartile range) in



**Table 3**  
Performance Metrics for Point-of-care Ultrasound Pediatric Emergency Medicine Physicians

Application	Accuracy (95% CI)			Sensitivity (95% CI)			Specificity (95% CI)			Cohen's effect size, d (95% CI)
	Pre	Post	Difference	Pre	Post	Difference	Pre	Post	Difference	
<b>Soft Tissue N=119</b>	74.8	86.4	11.6 (9.6, 13.6)	86.4	89.5	4.9 (2.4, 7.5)	64.8	83.7	18.9 (15.6, 22.2)	1.0 (0.9, 1.2)
<b>Lung N=99</b>	79.8	89.6	9.8 (7.2, 12.5)	79.4	87.2	7.8 (3.9, 11.7)	80.6	91.8	11.2 (7.6, 14.8)	0.8 (0.6, 1.0)
<b>Cardiac N=102</b>	74.2	81.6	7.4 (4.9, 10.0)	70.8	80.6	9.8 (5.5, 14.0)	76.3	82.5	6.3 (2.6, 9.9)	0.6 (0.4, 0.8)
<b>FAST N=99</b>	79.3	88.0	8.6 (6.2, 11.1)	78.3	88.0	9.7 (6.0, 13.4)	80.2	87.5	7.3 (4.0, 10.6)	0.7 (0.5, 0.9)

Pre – performance on the initial 25 cases

Post – performance on the final 25 cases

N – Number of participants

minutes it took to complete each 100-case application was as follows: soft tissue (still image) 29.5 (20.9 to 40.7), lung (video + still image) 52.3 (39.7 to 72.7), cardiac (video + still image) 52.0 (39.7 to 72.7), and FAST (video + four still images) 63.6 (49.1 to 76.6) minutes ( $p < 0.0001$ ).

### Missing Data

There were no differences in the demographics of the 44 of 172 (25.6%) who dropped out versus the 128 who completed at least one 100-case application. Further, there were no differences in initial accuracy of those participants who completed the 100-case application versus those who did not (mean initial accuracy difference = 0.14%, 95% CI = -4.2 to 3.9).

## DISCUSSION

We demonstrated that the case numbers required to reach the performance benchmarks ranged considerably for both the average and least efficient learners. Further, we noted that the deliberate practice of POCUS image interpretation led to skill improvement within a 100-case online learning experience, and most participants needed only about 2 to 3 hours to achieve the highest performance benchmark accuracy of 95% for all applications except cardiac, which required 3 to 10 hours. These data could inform education strategies and potentially add to the skill development of POCUS image interpretation.

Most organizations are faced with decisions about ideal credentialing POCUS standards to allow their physicians to safely practice at the bedside. This complex discussion often includes consensus building, stakeholder engagement, and the use of standardized competency-setting methods.<sup>31,32</sup> Our results can inform these organizational discussions on the variable journey learners take to a given performance benchmark, while the broader POCUS community

considers which performance benchmarks are most appropriate. Our data also demonstrate that the numbers of cases required to achieve a performance benchmark were not normally distributed, with skewed distribution tails, indicating that a minority of learners required considerably more cases to attain each benchmark. This has policy implications for education guidelines based on an “average” learner. Specifically, our data suggest that a shift away from the current standard of recommending specific numbers of cases for POCUS proficiency (e.g., 25–50 cases<sup>2,33</sup>) in favor of learners achieving a performance-based competency benchmark.<sup>11</sup> This approach is also in keeping with evolving performance-based competency frameworks in residency training, which promote greater accountability and documentation of actual capability.<sup>34</sup>

Separating out the POCUS image interpretation task and then integrating knowledge gained into whole task activities can reduce cognitive load during face-to-face bedside POCUS teaching.<sup>8</sup> This may result in more efficient and effective learning and greater learning satisfaction as demonstrated in similar forms of blended learning.<sup>35</sup> Alternatively, one could argue that POCUS image interpretation may be inherently easier to learn at the bedside, where one can pursue visual hints, make comparisons to unaffected areas, or better integrate the clinical context. However, these bedside advantages are challenged by the learner also needing to simultaneously acquire the images and conditions that lead to suboptimal image acquisition may limit the learning of image interpretation. Further, given the relatively low frequency of pathology in pediatrics, bedside learning may not efficiently offset the fact that many cases may be needed for most learners to achieve practice-ready standards, whereby a 50% or higher abnormal rate has previously been shown to be optimal for achieving an acceptable user sensitivity.<sup>17</sup> Given the pace of case exposure, bedside education may also not be very effective at identifying specific areas of individual participant weaknesses or

applications that are more difficult to learn. For example, our results provided evidence that cardiac and normal soft tissue cases were more difficult than other POCUS applications. Future research should explore the interaction between learning the image interpretation skill via online deliberate practice and the real-time application of POCUS at the bedside.<sup>36,37</sup>

The education intervention demonstrated a large effect size for the soft tissue and lung applications and a moderate effect size for the cardiac and FAST applications.<sup>30</sup> One possible explanation for these disparities is that cardiac and FAST applications were more difficult to interpret, due to increased complexity of anatomy (cardiac, FAST), differences in number of views (FAST), or low a priori skill due to low rates of pathology of these applications at the bedside. Strategies that may enhance diagnostic performance outcomes using the education intervention in this study include scaffolding (e.g., embedding electronic hints or coaching),<sup>38,39</sup> more deliberate review of incorrect cases (including referring to supplemental resources), and/or repeating cases as many times as needed to reach a desired performance outcome.

None of the participant baseline variables that we tested predicted for achieving expert-level performance. While it is not possible to be certain of the reasons for this from our data, we can consider some possible explanations. With respect to the variable of number of scans (100 vs. >100), the number alone may not be sufficient to predict for achieving expert level if a participant did not routinely receive feedback on images acquired at the bedside, which may limit a participant's ability to learn from each scan performed.<sup>7</sup> POCUS training during PEM fellowship versus no training during fellowship may not have alone impacted participant ability to achieve expert scores since many of our study attending-level participants also engaged in institutional and other POCUS workshops/courses. Further, since almost all our participants worked at children's hospitals, we were likely underpowered to examine the impact of practice setting. Finally, since scores are weighted for the 25 most recent cases and case interpretation difficulty varied over the 100-case experience, initial accuracy scores may not have predicted for subsequent and final scores.

## LIMITATIONS

---

Image interpretations were based on the expert opinion of three POCUS experts and the interobserver agreement between these opinions was high; however,

expert opinions may be subject to error.<sup>40</sup> Participants were able to select the applications and about one-quarter of our enrolled participants did not complete the minimum study intervention; therefore, our results may be biased by increased engagement for the selected applications and/or toward more POCUS-motivated PEM physicians. Since this is a part-task education intervention removed from the bedside and utilizes a higher proportion of pathological cases than is typically experienced at the bedside, it is uncertain how knowledge gained via this tool will translate to patient-level skill and outcomes. Some participants required considerably more than the 100 cases prescribed within the study design. Like other models of educational outcome distributions, we have the most information about persons at the mean and the extreme predictions run the risk of spectrum bias.

Over the 4-week study period, other factors may have contributed to the study's reported learning outcomes. However, our data demonstrates that most participants completed the cases in one sitting so it is unlikely additional POCUS exposure influenced our study results significantly. The changes in participant performance were reported using a pre-post design, which may be subject to confounders that impact the validity of the results. Finally, this study included PEM physician participants and thus may not be generalizable to other categories of physician learners.

## CONCLUSIONS

---

Deliberate practice of pediatric point-of-care ultrasound image cases using an online learning and assessment platform may lead to skill improvement in point-of-care ultrasound image interpretation. Further, this method allows an efficient review of a larger number of cases than would be typically available with bedside practice alone. Our results also demonstrated that the rate of learning the image interpretation task of point-of-care ultrasound can be highly variable across learners. These data could inform education strategies and potentially add to our understanding of how the skill of point-of-care ultrasound image interpretation is acquired among pediatric emergency medicine physicians.

We acknowledge the efforts of Ms. Kelly Sobie who facilitated participant recruitment and providing access information to the study participants. We also thank the pediatric emergency medicine physicians who participated in this research.

## References

- Marin JR, Lewiss RE. Point-of-care ultrasonography by pediatric emergency medicine physicians. *Pediatrics* 2015;135:e1113–22.
- Abo AM, Alade KH, Rempell RG, et al. Credentialing pediatric emergency medicine faculty in point-of-care ultrasound: expert guidelines. *Pediatr Emerg Care* 2019. <https://doi.org/10.1097/PEC.0000000000001677>
- Cohen JS, Teach SJ, Chapman JI. Bedside ultrasound education in pediatric emergency medicine fellowship programs in the United States. *Pediatr Emerg Care* 2012;28:845–50.
- Marin JR, Zuckerbraun NS, Kahn JM. Use of emergency ultrasound in United States pediatric emergency medicine fellowship programs in 2011. *J Ultrasound Med* 2012;31:1357–63.
- Vieira RL, Hsu D, Nagler J, et al. Pediatric emergency medicine fellow training in ultrasound: consensus educational guidelines. *Acad Emerg Med* 2013;20:300–6.
- Gold DL, Marin JR, Haritos D, et al. Pediatric emergency medicine physicians' use of point-of-care ultrasound and barriers to implementation: a regional pilot study. *Acad Emerg Med Educ Train* 2017;1:325–33.
- Hattie J, Timperley H. The power of feedback. *Rev Educ Res* 2007;77:81.
- van Merriënboer JJ, Kirschner PA, Kester L. Taking the load off a learner's mind: instructional design for complex learning. *Educ Psychol* 2010;38:5–13.
- Augenstein JA, Deen J, Thomas A, et al. Pediatric emergency medicine simulation curriculum: cardiac tamponade. *MedEdPORTAL* 2018;14:10758.
- Jensen JK, Dyre L, Jørgensen ME, Andreasen LA, Tolsgaard MG. Collecting validity evidence for simulation-based assessment of point-of-care ultrasound skills. *J Ultrasound Med* 2017;36:2475–83.
- Jensen JK, Dyre L, Jørgensen ME, Andreasen LA, Tolsgaard MG. Simulation-based point-of-care ultrasound training: a matter of competency rather than volume. *Acta Anaesthesiol Scand* 2018;62:811–9.
- Madsen ME, Konge L, Nørgaard LN, et al. Assessment of performance measures and learning curves for use of a virtual-reality ultrasound simulator in transvaginal ultrasound examination. *Ultrasound Obstet Gynecol* 2014;44:693–9.
- Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;79:S70–81.
- Hatala R, Gutman J, Lineberry M, Triola M, Pusic M. How well is each learner learning? Validity investigation of a learning curve-based assessment approach for ECG interpretation. *Adv Health Sci Educ Theory Pract* 2019;24:45–63.
- Pusic M, Pecaric M, Boutis K. How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. *Acad Med* 2011;86:731–6.
- Boutis K, Cano S, Pecaric M, et al. Interpretation difficulty of normal versus abnormal radiographs using a pediatric example. *CMEJ* 2016;7:e68–77.
- Pusic MV, Andrews JS, Kessler DO, et al. Determining the optimal case mix of abnormal to normals for learning radiograph interpretation: a randomized control trial. *Med Educ* 2012;46:289–98.
- Pecaric MR, Boutis K, Beckstead J, Pusic MV. A big data and learning analytics approach to process-level feedback in cognitive simulations. *Acad Med* 2017;92:175–84.
- Ericsson KA. Acquisition and maintenance of medical expertise. *Acad Med* 2015;90:1471–86.
- Black P, William D. Assessment and classroom learning. *Assess Eval High Educ* 1998;5:7–71.
- Larsen DP, Butler AC, Roediger HL 3rd. Test-enhanced learning in medical education. *Med Educ* 2008;42:959–66.
- Lee MS, Pusic M, Carriere B, Dixon A, Stimec J, Boutis K. Building emergency medicine trainee competency in pediatric musculoskeletal radiograph interpretation: a multicenter prospective cohort study. *AEM Educ Train* 2019;3:1–11.
- Garg AX, Norman G, Sperotable L. How medical students learn spatial anatomy. *Lancet* 2001;357:363–4.
- Marin JR, Lewiss RE; American Academy of Pediatrics, Committee on Pediatric Emergency Medicine, 2013-2014, et al. Point-of-care ultrasonography by pediatric emergency physicians. Policy statement. *Ann Emerg Med* 2015;65:472–8.
- Emergency Ultrasound Fellowship Guidelines: An Information Paper 2011. Available at: [https://urldefense.proofpoint.com/v2/url?u=https-3A\\_\\_www.acep.org\\_globalassets\\_uploaded\\_uploaded-2Dfiles\\_acep\\_membership\\_sections-2Dof2Dmembership\\_ultra\\_eus-5Ffellowship-5Fip-5Ffinal-5F070111.pdf&d=DwIFaQ&c=Sj806OTFwmuG2UO1EEDr-2uZRzm2EPz39TfVBG2Km-o&r=nfDsffzV5rncAztCfqsZ7m-NHGa vxHNQ5FMO-bdw2jI&m=XnC1X3SyZ0H8WWhrpetpmOmOugPqMzuDZYdaKkdovfQ&s=LjC5anT0fwK2nIOlA0CKw8jw9rVBKZxmTFhTKmcCx8o&e=](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.acep.org_globalassets_uploaded_uploaded-2Dfiles_acep_membership_sections-2Dof2Dmembership_ultra_eus-5Ffellowship-5Fip-5Ffinal-5F070111.pdf&d=DwIFaQ&c=Sj806OTFwmuG2UO1EEDr-2uZRzm2EPz39TfVBG2Km-o&r=nfDsffzV5rncAztCfqsZ7m-NHGa vxHNQ5FMO-bdw2jI&m=XnC1X3SyZ0H8WWhrpetpmOmOugPqMzuDZYdaKkdovfQ&s=LjC5anT0fwK2nIOlA0CKw8jw9rVBKZxmTFhTKmcCx8o&e=). Accessed December 12, 2018.
- 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;2:756–64.
- Cook D, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)* 2016;1:31.
- Mason WM, Wong GY. The hierarchical logistic regression model for multilevel analysis. *J Am Stat Assoc* 1985;80:513–24.
- D'Agostino RB, Belanger AJ, D'Agostino RB Jr. A suggestion for using powerful and informative tests of normality. *Am Stat* 1990;44:316–21.

30. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, 1988.
31. Lundsgaard KS, Tolsgaard MG, Mortensen OS, Mylopoulos M, Østergaard D. Embracing multiple stakeholder perspectives in defining trainee competence. *Acad Med* 2019;94:838–46. <https://doi.org/10.1097/ACM.0000000000002642>
32. Yarris LM. Defining trainee competence value is in the eye of the stakeholder. *Acad Med* 2019;94:760–62. <https://doi.org/10.1097/ACM.0000000000002643>
33. *Ultrasound Guidelines: Emergency, Point-of-care, and Clinical Ultrasound Guidelines in Medicine*. 2016. Available at: [https://www.acep.org/globalassets/uploads/uploaded-files/acep/membership/sections-of-membership/ultra/ultrasound-policy-2016-complete\\_updatedlinks\\_2018.pdf](https://www.acep.org/globalassets/uploads/uploaded-files/acep/membership/sections-of-membership/ultra/ultrasound-policy-2016-complete_updatedlinks_2018.pdf). Accessed December 12, 2018.
34. Iobst WF, Sherbino J, Cate OT, et al. Competency-based medical education in postgraduate medical education. *Med Teach* 2010;32:651–6.
35. Garrison DR, Kanuka H. Blended learning: uncovering its transformative potential in higher education”. *Internet High Educ* 2004;7:95–105.
36. Lesgold A, Rubinson H, Feltovich P, Glaser R, Klopfer D, Wang Y. Expertise in a complex skill: diagnosing x-ray pictures. In: Chi MT, Glaser R, Farr M, editors. *The Nature of Expertise*. Hillsdale, NJ: Erlbaum, 1989. p. 311–42.
37. Lesgold AM. Human skill in a computerized society: complex skills and their acquisition. *Behav Res Methods Instrum Comput* 1984;16:79–87.
38. McConnaughey S, Freeman R, Kim S, Sheehan F. Integrating scaffolding and deliberate practice into focused cardiac ultrasound training: a simulator curriculum. *MedEdPORTAL* 2018;14:10671.
39. Pea RD. The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *J Learn Sci* 2009;13:423–51.
40. Bankier AA, Levine D, Halpern EF, Kressel HY. Consensus interpretation in imaging research: is there a better way? *Radiology* 2010;257:14–7.