

Building Emergency Medicine Trainee Competency in Pediatric Musculoskeletal Radiograph Interpretation: A Multicenter Prospective Cohort Study

Michelle Sin Lee, MDCM,¹ Martin Pusic, MD, PhD,² Benoit Carrière, MD, MHPE,³ Andrew Dixon, MD,⁴ Jennifer Stimec, MD,⁵ and Kathy Boutis, MD, MSc¹

ABSTRACT

Objectives: As residency programs transition from time- to performance-based competency standards, validated tools are needed to measure performance-based learning outcomes and studies are required to characterize the learning experience for residents. Since pediatric musculoskeletal (MSK) radiograph interpretation can be challenging for emergency medicine trainees, we introduced Web-based pediatric MSK radiograph learning system with performance endpoints into pediatric emergency medicine (PEM) fellowships and determined the feasibility and effectiveness of implementing this intervention.

Methods: This was a multicenter prospective cohort study conducted over 12 months. The course offered 2,100 pediatric MSK radiographs organized into seven body regions. PEM fellows diagnosed each case and received feedback after each interpretation. Participants completed cases until they achieved a performance benchmark of at least 80% accuracy, sensitivity, and specificity. The main outcome measure was the median number of cases completed by participants to achieve the performance benchmark.

Results: Fifty PEM fellows from nine programs in the US and Canada participated. There were 301 of 350 (86%) modules started and 250 of 350 (71%) completed to the predefined performance benchmark during the study period. The median (interquartile range [IQR]) number of cases to performance benchmark per participant was 78 (60–104; min = 56, max = 1,333). Between modules, the median number of cases to achieve the performance benchmark was different for the ankle versus other modules (ankle 366 vs. other 76; difference = 290, 95% confidence interval [CI] = 245 to 335). The performance benchmark was achieved for 90.7% of participants in all modules except the ankle/foot, where 34.9% achieved this goal (difference = 55.8%, 95% CI = 45.3 to 66.3). The mean (95% CI) change in accuracy, sensitivity, and specificity from baseline to performance benchmark was +14.6% (13.4 to 15.8), +16.5% (14.8 to 18.1), and +12.6% (10.7 to 14.5), respectively. Median (IQR) time on each case was 31.0 (21.0–45.3) seconds.

Conclusions: Most participants completed the modules to the performance benchmark within 1 hour and demonstrated significant skill improvement. Further, there was a large variation in the number of cases completed to achieve the performance endpoint in any given module, and this impacted the feasibility of completing specific modules.

From the ¹Division of Pediatric Emergency Medicine, Department of Pediatrics, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada; the ²Department of Emergency Medicine and Division of Learning Analytics at the NYU School of Medicine, New York, NY; ³Division of Emergency Medicine, CHU Sainte-Justine and Université de Montréal, Montreal, Quebec, Canada; the ⁴Division of Pediatric Emergency Medicine, Department of Pediatrics, Stollery Children's Hospital and University of Alberta, Edmonton, Alberta, Canada; and the ⁵Department of Diagnostic Imaging, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada.

Received October 14, 2018; revision received January 11, 2019; accepted February 4, 2019.

Funded by the Royal College of Physicians and Surgeons: There was no role of the funder in the study design, data collection, analysis, interpretation of the data, writing of the report, or decision to submit the article for publication.

The course implemented in this study operates as a nonprofit under the academic umbrellas of the Hospital for Sick Children and University of Toronto. Although there was a fee for the course (\$150 per participant), none of the authors nor content advisors get financial compensation at any point for their work related to this course. The fee is used to pay for operational expenses for the course as these funds are not available from

Over 20 million diagnostic films of pediatric extremities are done annually in the United States in emergency departments (EDs).¹ Emergency medicine physicians are tasked with real-time interpretations of these images and need to incorporate their diagnostic impressions into immediate decision making at the bedside. An appropriate initial diagnosis decides management, minimizing morbidity and long-term dysfunction. Thus, it is imperative for training programs to ensure that emergency medicine trainees develop competence in the skill of pediatric musculoskeletal (MSK) radiograph interpretation before they make high-stakes decisions with patients in independent practice.^{2,3}

Unfortunately, the interpretation of pediatric MSK radiographs has been shown to be relatively deficient among graduating pediatric and emergency medicine residents who will be working in EDs.⁴⁻⁸ These training results echo what has been noted in clinical practice, where misinterpretation of pediatric MSK radiographs in general and pediatric EDs has been estimated to be up to 19%,^{1,9-11} and pediatric extremity radiographs have been found to be one of the most common causes of discrepant radiograph reports.^{12,13} As many as 86% of these misinterpretations require a change in management,¹⁴ and in the United States, misdiagnoses of pediatric fractures has accounted for the third largest amount of dollars paid out to settle malpractice claims.¹⁵ Research in this area has called for effective education solutions to reduce radiograph interpretation errors at the bedside.^{9,16-18}

To bridge the knowledge–practice gap in pediatric MSK radiograph interpretation among graduating emergency medicine physicians, we developed an online education system that allows active learning of this skill using 2,100 authentic pediatric MSK images that represent seven regions of the pediatric MSK system (<https://imagesim.research.sickkids.ca/demo/msk/enter.php>).¹⁹⁻²⁴ With this platform, the presentation of images is simulated to mirror how clinicians interpret them in the clinical field and 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.^{19,21,22} Specifically, cases are presented with a brief clinical stem, standard images and views, and

juxtaposition of normal and abnormal cases.^{19,23} There are also hundreds of cases to review and after each case the system provides visual and text feedback, which allows for deliberate practice²⁵ and an ongoing measure of performance as part of the instructional strategy.^{26,27} Furthermore, in line with the unfolding competency-based residency framework, which promotes greater accountability and documentation of actual capability,²⁸⁻³¹ this platform also requires learners to complete cases until they reach a predefined performance level.

Despite these aforementioned strengths, the feasibility of embedding this type of learning system into a postgraduate training program needs to be studied rather than assumed. That is, it is equally important to ensure that the course will be completed when it is embedded longitudinally within a training program that has multiple competing learning requirements. Using Bowen's framework of how to conduct feasibility studies,³² we focused on the elements of integration, adaptation, and expansion.

We introduced a Web-based pediatric MSK radiograph learning system with a performance benchmark into pediatric emergency medicine (PEM) fellowships and aimed to establish the feasibility and effectiveness of implementing this intervention. Specifically, we examined the number of cases needed to achieve the performance threshold, the proportion of participants who successfully achieved the performance standard, and skill gains from baseline to performance standard.

METHODS

Study Design and Setting

This was a prospective cohort multicenter study. The education intervention was developed at Hospital for Sick Children (Toronto, Ontario, Canada) and New York University (New York, NY). The education intervention was implemented July 2015 to June 2016 at nine credentialed PEM fellowship programs in Canada ($n = 7$) and United States ($n = 2$).

Selection of Participants

The participants were first-year PEM fellows that completed a pediatrics residency. Sites that embedded the education intervention into their core PEM fellowship

learning goals included the Hospital for Sick Children, New York University, Columbia University, British Columbia Children's Hospital, Stollery Children's Hospital, Alberta Children's Hospital, CHU Sainte Justine, Children's Hospital of Eastern Ontario, and Montreal Children's Hospital. Once the respective program director committed to the education experience, the respective PEM fellows were provided unique username/password access to the learning system. The website outlined that engaging in the education intervention included the possibility that de-identified participant data may be used for research purposes. This study was approved by the institutional review boards of the Hospital for Sick Children and New York University.

Education Intervention

Radiograph Collection. We collected the radiographs by purposefully sampling from a pediatric emergency clinical setting. Specifically, images were identified by reviewing a tertiary care pediatric emergency diagnostic imaging database for pediatric MSK radiographs ordered from the ED from January 1, 2012, to December 31, 2014, in which the indication for the radiograph was exclusion of a fracture/dislocation. Radiographs were then downloaded by research assistants from the institutional picture archiving and communication system in JPEG format along with the respective final attending pediatric radiologist's report. One study team member (KB) reviewed all the images and excluded cases that had markers like casting material or embedded arrows that would suggest a diagnosis, very poor quality films such that radiograph findings were obscured, or those with an incomplete number of views. Cases with an uncertain final diagnosis were additionally reviewed with the collaborating pediatric MSK radiologist (JS) to establish a final diagnosis, and those that remained uncertain were excluded. In a separate diagnosis specific search of the archiving system that included the years 2010 to 2014, rarer but clinically important diagnoses (e.g., Monteggia or scaphoid fractures) were identified to ensure that there were a few examples of these injuries. For each case, a brief clinical history was written based on available clinical data and radiographs were categorized depending on whether there was the presence or absence of a fracture/dislocation. Cases with a fracture/dislocation were further subclassified by diagnosis and the location of the abnormality on the image. This resulted in a pool of 2,985 radiographs from which 885 normal cases were excluded to result in

2,100 radiographs with a case-mix frequency of 50% normal and 50% abnormal cases.^{19,23}

Online Software Application for Presentation of Radiograph Cases.

An online learning platform was previously developed using HTML, PHP, and Flash and a description of this can be found elsewhere.²⁰ In brief, the case experience is organized into seven regions of the pediatric MSK system, each of which contains 200 to 400 case examples: 1) skull, 2) shoulder/clavicle/humerus, 3) elbow, 4) wrist/forearm/hand, 5) pelvis/femur, 6) knee/tibia-fibula, and 7) ankle/foot. The participant reviews all relevant images of a case and commits to a response (fracture/dislocation absent/present) and is provided with immediate text and visual feedback on their interpretation (Figure 1). Once the participant has considered this information, they moved on to the next case and continued doing cases until a performance-based standard of 80% accuracy, sensitivity, and specificity was achieved on the most recently completed 25 cases. In our prior research,²² we found that 25 cases provided stable estimates of learner performance in these metrics. The software tracked participant progress through the cases and recorded responses to a MySQL database.

Performance Benchmark. In the absence of an evidence based performance standard that confirms competency, we considered several factors to select the 80% performance benchmark. The goal was not necessarily to achieve mastery during a fellowship program, which often requires several years of practice with feedback postgraduation. We also considered that beside exposure in the ED, radiology, and orthopedic rotations still had an important role in increasing image interpretation skills for a common clinical problem. Importantly, while diagnostic errors do occur, the vast majority are subtle fractures and are not of high clinical consequences.¹¹ Further, most EDs have radiology review of images available within 24 hours, and as such errors are captured and reported. Finally, we reviewed performance achieved for this participant group in our prior studies to understand what was feasible for most fellows.^{20,22} Based on these principles, our aim was to expose participants to a large number of cases that would likely take years to acquire via bedside emergency medicine practice alone toward the goal of an "acceptable and feasible" performance benchmark in a setting where additional education on this topic was ongoing at the bedside and quality

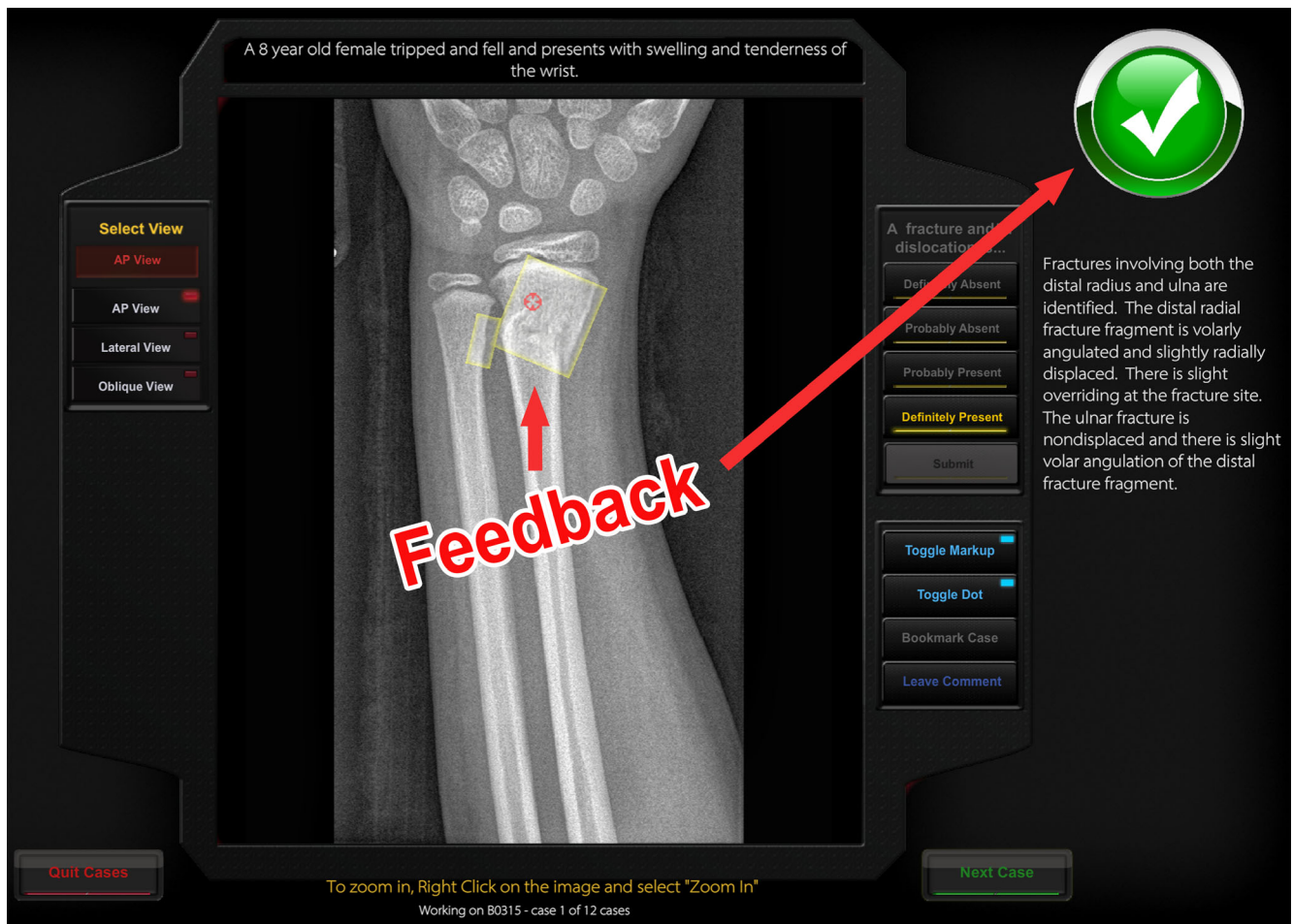


Figure 1. Visual and text feedback after case interpretation.

assurance interventions would capture diagnostic errors.

Course Fees. There was a \$150 Canadian fee per participant for 1 year of access to the cases (www.imagesim.com). The education system operates as a non-profit under the academic umbrellas of the Hospital for Sick Children and University of Toronto. Fees were used to pay for operational expenses. Neither the authors nor the content advisors were paid for their work on this education intervention or research.

Study Protocol

Recruitment. This learning platform was introduced at a national Canadian PEM meeting (Pediatric Emergency Research Canada) in January 2015. Program directors from PEM fellowships interested in participating enrolled their fellow participants in the education intervention.

Participant Engagement. Upon enrollment, PEM fellow participants received a brief 10-slide

presentation on the course reviewing the educational theory and goals of course participation. Fellows participated by using a computer of their choice and having online 24/7 access to the cases. Secure entry was ensured via a participant name and password given to each participant. After the fellow accessed the learning system, each participant was given some general information, which included assurance of confidentiality, the purpose of the exercise, and some information on how to use the system. The following participant demographics were captured: country of participation (Canada or United States) and sex (male or female). Fellows were not provided with any information about the proportion of normal to abnormal cases or types of pathology. No time limitation per case interpretation was imposed and each participant selected the order in which to complete a given module. As described, the participant then completed cases until the performance benchmark of at least 80% accuracy, sensitivity, and specificity was achieved.

Course Integration Into PEM Fellowship.

Prior to course initiation at a given site, each program

director met with the academic director of the pediatric MSK course to review strategies for integrating this course into the curriculum. Program directors were asked to consider if there were specific rotations (e.g., orthopedics or radiology) during which their fellows should be focusing on these modules. Program directors were also encouraged to schedule group sessions at a computer lab every 2 months to provide structure and timelines to complete the modules over the 12-month period. Finally, each PEM fellowship director was provided access to a dashboard that tracked each participant's progress and was encouraged to review the status of completion of modules with the PEM fellows at routine semiannual reviews. To our knowledge, programs did not incentivize fellows to complete the modules nor punish them for lack of completion.

Adaptation and Course Feedback. Feedback was formally solicited from program directors and PEM fellow participants. Every 2 months, the study team contacted the PEM program directors and highlighted their site's participant progress and asked for feedback from the perspective of a program director or PEM fellow participant. When PEM fellow participant challenges were identified, the study team worked with the program director to focus on strategies that may be more successful at their site. Further, the online learning system allowed for each participant to comment on every case they reviewed and provide general comments. Any errors in case details were corrected every 2 months. Participants also had access to technical support 15 hours per day, 7 days per week. Any challenges were typically resolved within 24 hours.

Outcome Measures

The primary outcome was the median number of cases completed to reach the predefined performance benchmark overall and per module. Secondary outcomes included the proportion of participating fellows that started the education intervention and completed at least 25 cases and the proportion who achieved the performance benchmark in a given module. We also measured the median time in seconds required to complete a case and compared this between modules. Further, we determined the mean change in accuracy, sensitivity, and specificity from baseline to performance benchmark overall and independently for each of the seven modules and compared performance gains across different modules. Finally, we reviewed feedback provided by program

directors and participants and used this information to better understand outcomes and consider how we could improve the experience in real time and for future implementations.

Data Analyses

Unit of Analysis. Each case completed by a participant was considered one item. Normal items were scored dichotomously depending on the match between the participant's response and the reference standard diagnosis. Abnormal items were scored correct if the participant had both classified it as abnormal and indicated the correct region of abnormality on at least one of the images of the case. Participant data were included only if they completed a minimum of 25 cases in a given module.

Number of Cases Completed. We calculated the median number of cases to the performance benchmark with respective interquartile ranges (IQRs) for all modules and for each module (primary analysis). We also presented graphic data demonstrating cases completed overall, distinguishing which participants achieved the performance benchmark and which did not. The Kruskal-Wallis test was used to compare for differences between participants.

Participants Who Completed Modules. Proportions were reported with respective 95% confidence intervals (CIs). The Kruskal-Wallis test was used to compare for differences between modules.

Median Time on Case. We calculated the median time on case in seconds for all participants who started the intervention overall and per module. The Kruskal-Wallis test was used to compare for differences between modules.

Performance Increase. Baseline accuracy, sensitivity, and specificity scores were calculated after the initial 25 cases were completed, while final score was determined based on an average of the terminal 25 cases. Comparisons of within-subject performance changes were performed using the paired Student's *t*-test. Analysis of variance was used to compare performance differences between multiple groups and post hoc analyses were performed adjusted using the Bonferroni's test.

Significance was set at $p < 0.05$, but for post hoc analyses was set at <0.01 to adjust for multiple testing.

All analyses were conducted using SPSS software analysis package (version 23).

RESULTS

The program was provided to 59 first-year PEM fellows. Of these, 50 (84.7%) PEM fellows started the assigned modules, 45 (90.0%) of whom were in a Canadian fellowship and 31 (62.0%) were female.

Across all modules, the median (IQR) number of cases required to achieve the performance benchmark (primary outcome) for participants was 78 (60–104), with respective minimum and maximum values of 56 and 1,333. Of the participants, 70.8% completed 100 cases or less to achieve the performance target, while 92.7% achieved this goal in 200 cases or less (Figure 2). Between participants there was a significant variation in number of cases required to achieve the performance benchmark for a given module ($p = 0.03$). Between modules, the median (IQR) number of cases required to achieve the performance benchmark differed between modules ($p < 0.0001$), Table 1. This comparison between the ankle/foot module and other modules demonstrated a difference of 290 cases (95% CI = 245 to 335) completed.

Of the 301 of 350 (86.0%) modules that were started, participants reached the performance benchmark in 249 to 350 (71.1%) modules. Four of the nine (44.4%) sites scheduled sessions at a computer laboratory at regular intervals during the study period. At these sites, 79 of 97 (81.4%) of modules were completed to the performance benchmark, while this was true for 170 of 204 (83.3%) of modules at sites that did not schedule group computer sessions (difference = -1.9% , 95% CI = -6.8 to $+11.8$). The performance benchmark was achieved for 90.7% of participants in all modules except the ankle/foot, where 34.9% achieved this skill level (difference = 55.8 , 95% CI = 45.3 to 66.3). Median (IQR) time on each case was 31.0 (21.0–45.3) seconds. Median time on case differed between modules ($p < 0.0001$; Table 1). Post hoc analyses demonstrated that time on case was greater for the ankle relative to all other modules (difference = $+16.9$, 95% CI = $+10.1$ to $+23.6$).

The overall change in accuracy from baseline to the performance benchmark was $+14.6\%$ (95% CI = 13.4 to 15.8), with the corresponding Cohen's effect size of 1.8 (95% CI = 1.1 to 2.5). The respective overall change in sensitivity and specificity were $+16.5\%$ (95% CI = 14.8 to 18.2) and $+12.6\%$ (95% CI =

10.7 to 14.5); increase in sensitivity relative to specificity was 2.8% (95% CI = 0.4 to 5.2). The per module mean changes in accuracy, sensitivity, and specificity are detailed in Table 2. Overall, the pre-post changes in accuracy and sensitivity were different between modules ($p < 0.0001$) but not for specificity ($p = 0.01$).

DISCUSSION

This study demonstrated that the implementation of an online pediatric MSK image interpretation learning system with a performance benchmark in PEM fellowships resulted in significant increases in PEM fellow interpretation skill. Further, most of these PEM fellow participants completed the seven modules to a preset performance target. However, there was a large variation in the number of cases required to achieve the performance goal between individual participants and between different types of modules, and this impacted module completion for some participants, particularly in the ankle/foot case set.

The finding that the number of cases required to achieve a performance endpoint for a given participant was variable demonstrates that individual learners reach their milestones at varying speeds, and this learning model affords them the flexibility of how they learn while providing guidance on when they have mastered a particular skill.²⁹ Despite this variation, about 70% of participants were still able to achieve performance benchmark in a feasible time frame, about 1 hour. In contrast, some participants needed to complete several hundred cases over several hours before they reached the performance target. This was particularly true for the ankle/foot case set, whereby only about one-third of the participants achieved the 80% performance goal. It is difficult to know why this set posed additional challenges relative to other modules. Feedback identified from comments left on the system stated that many of the fractures were small and difficult to see (e.g., distal fibular fractures, subtle foot fractures), resulting in low performance and the need to do more cases to improve scores. The latter may have led to fatigue and decreased motivation for many participants. Future iterations of this module may consider reducing the number of minor foot/ankle fractures of little clinical significance or incorporating adaptive learning algorithms to refine the case presentation such that the system selectively presents cases most relevant to an individual participant's

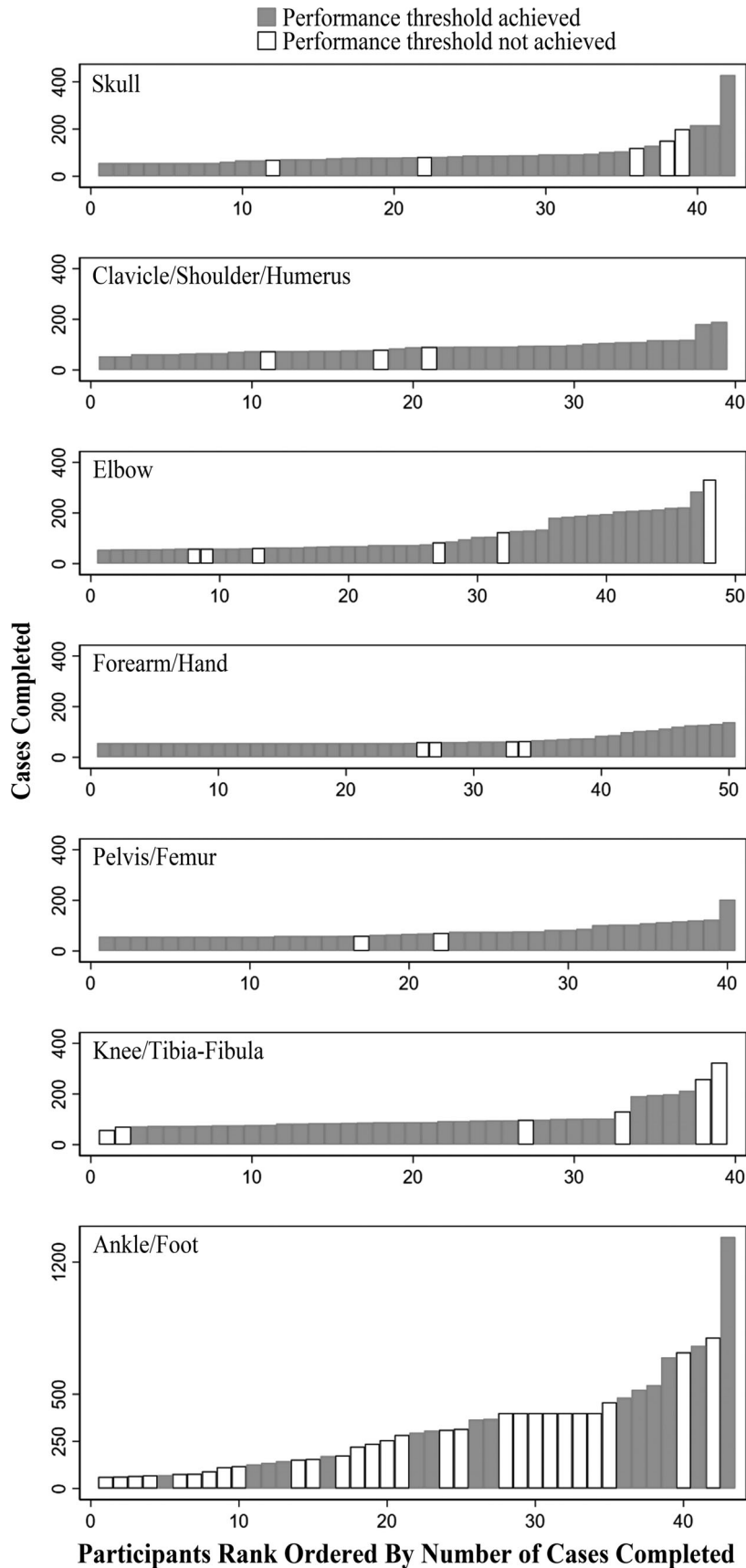


Figure 2. Participants are ranked left to right in ascending order of how many cases they completed. Each bar is one individual. *Gray bars* are participants who successfully achieved the performance threshold. *White bars* are participants who did not achieve the performance threshold. The x-scales differ as not all body regions have the same number of participants. The y-scale is the same for all graphs except the bottom-most where it is adjusted for the substantially greater number of cases completed for the ankle/foot module.

Table 1
Per-module Data for Achieving Performance Threshold

Pediatric MSK Module	Cases Completed to Reach Performance Standard, Median Number* (IQR)	Proportion of Participants Completed Module To Performance Standard,* n/N (%)	Time on Case (Seconds),* Median (Min, Max)
Skull	79 (68–113) n = 37	37/42 (88.1)	36.3 (15.4, 88.6) n = 42
Shoulder/clavicle/humerus	90 (76–119) n = 36	36/39 (92.3)	32.3 (13.4, 58.1) n = 39
Elbow	73 (68–190) n = 42	42/48 (87.5)	28.9 (8.0, 63.2) n = 48
Forearm/hand	72 (56–105) n = 46	46/50 (92.0)	37.4 (10.8, 77.9) n = 50
Pelvis/femur	69 (57–111) n = 38	38/40 (95.0)	36.0 (10.4, 80.0) n = 40
Knee/tibia-fibula	89 (78–103) n = 33	33/39 (84.6)	26.9 (7.6, 79.5) n = 39
Ankle/foot	366 (150–548) n = 15	15/43 (34.9)	41.0 (19.8, 89.5) n = 43

MSK = musculoskeletal.

*Performance standard defined as achieving $\geq 80\%$ accuracy, sensitivity, and specificity.

Table 2
Changes in Accuracy, Sensitivity, and Specificity from Baseline to Performance Standard of at Least 80% Accuracy, Sensitivity, and Specificity

Pediatric MSK Module	Percent Change in Accuracy (95% CI)	Percent Change in Sensitivity (95% CI)	Percent Change in Specificity (95% CI)
Skull (n = 37)	17.6 (15.3–19.9)	18.4 (14.5–22.4)	15.6 (11.3–19.9)
Shoulder/clavicle/humerus (n = 36)	15.9 (12.4–19.4)	13.5 (9.4–17.6)	18.9 (13.6–24.2)
Elbow (n = 42)	15.4 (12.1–18.8)	18.7 (15.5–21.9)	14.2 (9.5–18.9)
Forearm/hand (n = 46)	10.4 (8.2–12.6)	10.6 (7.3–13.9)	10.6 (6.7–14.5)
Pelvis/femur (n = 38)	8.7 (6.1–11.3)	8.7 (6.2–11.2)	7.1 (2.6–11.7)
Knee/tibia-fibula (n = 33)	16.4 (13.2–19.7)	23.8 (19.1–28.5)	7.1 (3.0–11.2)
Ankle/foot (n = 15)	26.1 (20.2–32.0)	32.7 (24.7–40.7)	17.9 (10.3–25.5)

MSK = musculoskeletal.

weaknesses. As emergency medicine transitions to competency-based residency training programs, educators will have to consider how to balance ideal performance thresholds in specific areas that demonstrate a high degree of individual variability with ensuring completion of an entire emergency medicine curriculum in a reasonable time frame for all their trainees.^{30,31} Ultimately, it is likely that while individual participants make take longer in one task, they may achieve competency more quickly in other tasks. Hence the need to compromise on standards for a given task will not likely be necessary.

The differential experience between modules to reach a performance standard highlights the fact that that, from an educational standpoint, the number of

cases to a performance threshold in one module cannot be assumed to be similar for even within the same domain of pediatric MSK images, let alone a different image type entirely (e.g., point-of-care ultrasound images). These findings are consistent with the educational phenomenon of “case specificity,” where the diagnostic performance on one case does not predict performance on another.³³ While the specific reasons for this differential performance were not elucidated in this research, this does map to what has been noted at the bedside for this clinical area. Mounts et al.³⁴ reviewed 220 cases and found that some pediatric MSK radiograph interpretations were more prone to missing fractures than others; the most frequently missed fractures were of the hand phalanges (26.4%) followed by metatarsus (9.5%), distal radius (7.7%),

tibia (7.3%), and phalanges of the foot (5.5%). While some of these identified clinical challenges do map to our findings in the setting of an educational intervention, it is important to note that in this other publication, the study focused only on identifying missed fractures. In contrast, our intervention also identified overcalling of image interpretation, and this may explain some of the differences. Nevertheless, overcalling pathology also carries harm, often in the form of incorrect treatment and unnecessary follow-up. Future research should explore both types of pediatric MSK interpretation errors to allow a more informed view of diagnostic interpretation needs, which will consider data from stand-alone education interventions as well as errors made during clinical practice.

The findings of our study have implications for practice. Our results reinforce the motivations for competency-by-design. A “one-size-fits-all” curriculum applied to postgraduate trainees would result in a variable level of skill as residents prepare to enter the world as independent clinicians. Thus, one of the key benefits of the type of educational platform embedded in this research is that these learning tools allow trainees to automatically get an experience weighted toward evening out case exposure and performance to a specified standard. As such, we encourage the implementation of learning assessment platforms for emergency medicine skills amenable to cognitive simulation and intensive deliberate practice. Specific to the learning intervention described in this study, it is available to any postgraduate trainee or attending-level physician to participate as an individual or as part of a program engagement (<https://imagesim.com/>). Further, at the end of this study period, the duration of access to the course was expanded from 12 to 24 months to be more in keeping with the duration of academic fellowships. Importantly, the pairing of these learning platforms with learning analytics can provide a granular look at individual- and group-level performance.²¹ This will in turn allow educators to identify learners who face difficulties early such that additional learning interventions could be implemented in a timely manner or the development of quality assurance programs in areas identified as common deficiencies. Finally, as the shift toward competency-based medical education progresses, there is a need to explore the new technologies that will help feasibly establish and assess competency.^{35,36} The learning platform presented in this research demonstrated effectiveness for training on clinical tasks with dichotomous outcomes. Where the assessment is more complex, our model

could be adapted to address unique training needs for different types of clinical tasks.

LIMITATIONS

This research has limitations that warrant consideration. Although all participants were PEM fellows, there could have been heterogeneity between the participants (e.g., number of radiology or orthopedic rotations), which may have impacted performance outcomes. Since we did not collect information on these variables, we were not able to control for these potential confounders. Performance on this education platform may not necessarily translate into performance in a clinical setting where real-time patient information and higher resolution monitors may impact interpretation skills. This intervention did not explore skill retention from this education intervention, which is the subject of another study (Boutis et al., manuscript submitted for publication). This education intervention requires a fee for participation and may pose a barrier to participation for some programs. Finally, while we have previously shown that PEM fellows and emergency medicine residents have similar skill acquisition using this system,²⁰ this study enrolled a convenience sample of PEM fellows, and therefore our results may have limited generalizability to the broader group of postgraduate trainees.

CONCLUSIONS

There was significant variation between participants and between modules in the number of cases required to achieve a predefined performance standard. Nevertheless, the performance target was achieved for a high percentage of participants on average in about 1 hour per module, which speaks to the feasibility of implementing similar programs. The learning outcomes for the ankle/foot case set were different than the other case sets, highlighting an area that might require additional training. Overall, there were significant gains made in accuracy, sensitivity, and specificity across all modules. Future studies are required to explore the best ways to establish defensible evidence-based competency thresholds and how similar strategies can be applied to the training of other skill sets.

The authors acknowledge Dr. Martin Pecaric of Contrail Consulting Services for providing technical development and support for the education intervention.

References

- Fleisher G, Ludwig S, McSorely M. Interpretation of pediatric X-ray films by emergency department pediatricians. *Ann Emerg Med* 1983;12:153–8.
- Lewis RE, Chan W, Sheng AY, et al. Research priorities in the utilization and interpretation of diagnostic imaging: education, assessment, and competency. *Acad Emerg Med* 2015;22:1447–54.
- Chew FS. Distributed radiology clerkship for the core clinical year of medical school. *Acad Med* 2002;77:1162–3.
- Dixon AC. Pediatric fractures - an educational needs assessment of Canadian pediatric emergency medicine residents. *Open Access Emerg Med* 2015;7:25–9.
- Ryan LM, DePiero AD, Sadow KB, et al. Recognition and management of pediatric fractures by pediatric residents. *Pediatrics* 2004;114:1530–3.
- Reeder BM, Lyne ED, Patel DR, Cucos DR. Referral patterns to a pediatric orthopedic clinic: implications for education and practice. *Pediatrics* 2004;113(3 Pt 1):e163–7.
- Taras HL, Nader PR. Ten years of graduates evaluate a pediatric residency program. *Am J Dis Child* 1990;144:1102–5.
- Trainor JL, Krug SE. The training of pediatric residents in the care of acutely ill and injured children. *Arch Pediatr Adolesc Med* 2000;154:1154–9.
- Guly HR. Diagnostic errors in an accident and emergency department. *Emerg Med* 2001;18:263–9.
- Klein EJ, Koenig M, Diekema DS, Winters W. Discordant radiograph interpretation between emergency physicians and radiologists in a pediatric emergency department. *Pediatric Emerg Care* 1999;15:245–8.
- Smith JE, Tse S, Barrowman N, Bilal A. Missed fractures on radiographs in a pediatric emergency department. *CJEM* 2016;18(Suppl 1):S119.
- Arora R, Kannikeswaran N. Radiology callbacks to a pediatric emergency department and their clinical impact. *Pediatr Emerg Care* 2018;34:422–5.
- Taves J, Skitch S, Valani R. Determining the clinical significance of errors in pediatric radiograph interpretation between emergency physicians and radiologists. *CJEM* 2017;20:420–4.
- Hallas P, Ellingsen T. Errors in fracture diagnoses in the emergency department—characteristics of patients and diurnal variation. *BMC Emergency Med* 2006;6:4.
- Selbst SM, Friedman MJ, Singh SB. Epidemiology and etiology of malpractice lawsuits involving children in US emergency departments and urgent care centers. *Pediatr Emerg Care* 2005;21:165–9.
- Erhan ER, Kara PH, Oyar O, Unluer EE. Overlooked extremity fractures in the emergency department. *Turk J Trauma Emerg Surg* 2013;19:25–8.
- Freed HA, Shields NN. Most frequently overlooked radiographically apparent fractures in a teaching hospital emergency department. *Ann Emerg Med* 1984;13:900–4.
- Wei CJ, Tsai WC, Tiu CM, Wu HT, Chiou HJ, Chang CY. Systematic analysis of missed extremity fractures in emergency radiology. *Acta Radiol* 2006;47:710–7.
- Boutis K, Cano S, Pecaric M, et al. Interpretation difficulty of normal versus abnormal radiographs using a pediatric example. *CMEJ* 2016;37:e68–77.
- Boutis K, Pecaric M, Pusic M. Using signal detection theory to model changes in serial learning of radiological image interpretation. *Adv Health Sci Educ Theory Pract* 2010;15:647–58.
- 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.
- 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.
- 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.
- Pusic MV, Kessler D, Szyld D, Kalet A, Pecaric M, Boutis K. Experience curves as an organizing framework for deliberate practice in emergency medicine learning. *Acad Emerg Med* 2012;19:1476–80.
- Ericsson KA. Acquisition and maintenance of medical expertise. *Acad Med* 2015;90:1471–86.
- Black P, William D. Assessment and classroom learning. *Assess Educ* 1998;5:7–71.
- Larsen DP, Butler AC, Roediger HL 3rd. Test-enhanced learning in medical education. *Med Educ* 2008;42:959–66.
- Albanese MA, Mejicano G, Anderson WM, Gruppen L. Building a competency-based curriculum: the agony and the ecstasy. *Adv Health Sci Educ Theory Pract* 2010;15:439–54.
- Takahashi SG WA, Kennedy M, Hodges B. Innovations, integration and implementation issues in competency-based education in postgraduate medical education. *The Future of Medical Education in Canada PG Consortium*, 2011.
- Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Med Teach* 2010;32:638–45.
- Iobst WF, Sherbino J, Cate OT, et al. Competency-based medical education in postgraduate medical education. *Med Teach* 2010;32:651–6.
- Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. *Am J Prev Med*. 2009;36:452–7.

33. Wimmers PF, Fung CC. The impact of case specificity and generalisable skills on clinical performance: a correlated traits-correlated methods approach. *Med Educ* 2008;42:580–8.
34. Mounts J, Clingenpeel J, McGuire E, Byers E, Kireeva Y. Most frequently missed fractures in the emergency department. *Clin Pediatr (Phila)* 2011;50:183–6.
35. Nousiainen MT, Caverzagie KJ, Ferguson PC, Frank JR. Implementing competency-based medical education: What changes in curricular structure and processes are needed? *Med Teach* 2017;39:594–8.
36. Caverzagie KJ, Nousiainen MT, Ferguson PC, et al. Overarching challenges to the implementation of competency-based medical education. *Med Teach* 2017;39:588–93.