JB&JS
OPEN ACCESS

AMERICAN
ORTHOPAEDIC
ASSOCIATION

AOA Critical Issues in Education

A Data-Driven Methodology to Comprehensively Assess Bone Drilling Using Radar Plots

Aman Nigam, PhD, James F. Kellam, MD, Catherine G. Ambrose, PhD, and Bruce L. Tai, PhD

Background: The study aims to develop a data-driven methodology to assess bone drilling in preparation for future clinical trials in residency training. The existing assessment methods are either subjective or do not consider the interdependence among individual skill factors, such as time and accuracy. This study uses quantitative data and radar plots to visualize the balance of the selected skill factors.

Methods: In the experiment, straight vertical drilling was assessed across 3 skill levels: expert surgeons (N = 10), intermediate residents (postgraduate year-2-5, N = 5), and novice residents (postgraduate year-1, N = 10). Motion and force were measured for each drilling trial, and data from multiple trials were then converted into 5 performance indicators, including overshoot, drilling time, overshoot consistency, time consistency, and force fluctuation. Each indicator was then scored between 0 and 10, with 10 being the best, and plotted into a radar plot.

Results: Statistical difference ($p < 0.05$) was confirmed among 3 skill levels in force, time, and overshoot data. The radar plots revealed that the novice group exhibited the most distorted pentagons compared with the well-formed pentagons observed in the case of expert participants. The intermediate group showed slight distortion that was between the expert and novice groups.

Conclusion/Clinical Relevance: This research shows the utility of radar plots in drilling assessment in a comprehensive manner and lays the groundwork for a data-driven training scheme to prepare novice residents for clinical practice.

Introduction

Bone drilling is a crucial component of the orthopaedic residency program that demands dexterity, precise motor skills, and anatomical awareness. Historically, bone drilling has been taught in an apprenticeship setting with an emphasis on optimizing postoperative outcomes and minimizing complications[1]. Despite the benefits of this mode of learning, traditional faculty evaluations are generally subjective and nonstandardized, making them vulnerable to experiential bias and limiting their utility as a reliable method of assessment[2,3]. Furthermore, studies have shown that faculty evaluations are effective for learning higher-order decision-making skills but lack the precision needed to evaluate basic surgical tasks[4,5].

To address the shortcomings of traditional faculty evaluations, several studies have explored quantitative assessment techniques, primarily focused on the use of hand motion

Disclosure: The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (http://links.lww.com/JBJSOA/A583).

tracking to evaluate drilling performance[6,7]. For example, Zirkle et al. used a VR simulation model to assess temporal bone drilling[8]. The study evaluated 3 parameters (the number of movements, distance traveled, and drilling time) and showed that the simulator assessments were more sensitive than evaluator assessments. Close et al. used video tapes of mastoidectomy procedures to compare performance evaluations by experienced evaluators and computer tracking software[3]. The study revealed that computer tracking software was more accurate than human evaluators in differentiating skill levels. These studies demonstrated the potential of using quantitative assessment techniques to accurately evaluate surgical drilling performance. However, there is no standardized platform or metrics to assess bone drilling. Furthermore, VR and physical bone surrogates (e.g., plastic pipes or wood plates) do not offer a realistic tactile environment, which can compromise the reliability of quantitative performance assessment[9].

To address these issues, the authors in a previous study introduced a hybrid drilling system[10]. The system used 3D-printed bone surrogates that replicated a healthy human cortical bone to provide realistic drilling haptics and motion/force sensors to provide quantitative data. The preliminary investigation was conducted to identify appropriate performance parameters by comparing motion and force data from expert surgeons and novice residents. The results showed that the experts generally exhibited a lower positional error, higher force, and a shorter drilling time with better repeatability. However, the results also showed a strong interdependence, for example, a high positional error is correlated to a shorter time. This finding suggests that assessing performance using a single parameter can compromise another aspect.

For this reason, this current article proposes a comprehensive approach that involves the visualization of all performance indicators on a global-normalized scale using radar plots. Although the radar plots have previously been used to comprehensively evaluate residents' clinical competency[11] and health care data[12,13], their application for assessing bone drilling performance is new. In the context of drilling, the previous study had identified that overshoot, drilling time, and drilling force are effective metrics to distinguish expert and novice skill levels[10]. The current study extends on this research by incorporating stability and repeatability as additional parameters, resulting in a 5-factor radar plot for assessment. The objective was to determine whether the proposed methodology using radar plots can detect noticeable differences between novice and expert groups (i.e., construct validity). If so, these plots can then be used as guidance for drilling training toward the expert level.

Methods

Physical Setup Design and Testing Procedure

As illustrated in Figure 1-A, a physical setup was developed with the capability to simultaneously record motion and force data during the drilling operation. A custom-made 3D-printed bone surrogate emulating a femur shaft was chosen as a test specimen (thickness: 5 mm and diameter: 35 mm). The 3D-printed bone surrogate was manufactured using a 2-step approach, as detailed previously[10,14,15]. The base geometry was first printed using a 3D-printable plaster and then infiltrated with a commercial epoxy to mimic the hardness of a young femur. The expert group in this study confirmed that the drilling characteristics of the 3D-printed surrogate were comparable with those of a young adult femur.

In addition, the bone surrogate was attached to a 6-degree-of-freedom force sensor (ATI gamma; ATI Industrial Automation), which recorded the drilling forces along the 3 axes. A robotic arm (Geomagic Touch; 3D Systems) was used to record the 3-dimensional motion data with an accuracy of 0.5 mm[16]. A Bosch hand drill (Bosch) was used with surgical-grade 2.5-mm drill bits (Synthes, Switzerland) at 1,200 rpm. This hand drill was chosen because of its close resemblance to a typical surgical hand drill in size, weight, and rotating speed. Also, the 2.5-mm bit was investigated throughout the study to make sure no significant tool wear.

This study was approved by the Institutional Review Board (HSC-MS-19-0361). Participants were recruited both

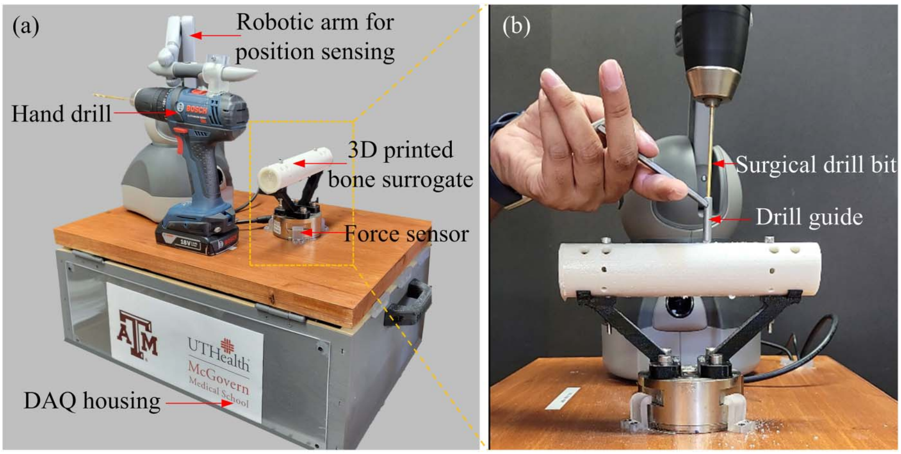


Fig. 1

(**Fig. 1-A**) Bone drilling experimental setup; (**Fig. 1-B**) vertical drilling on a 3D-printed bone surrogate.

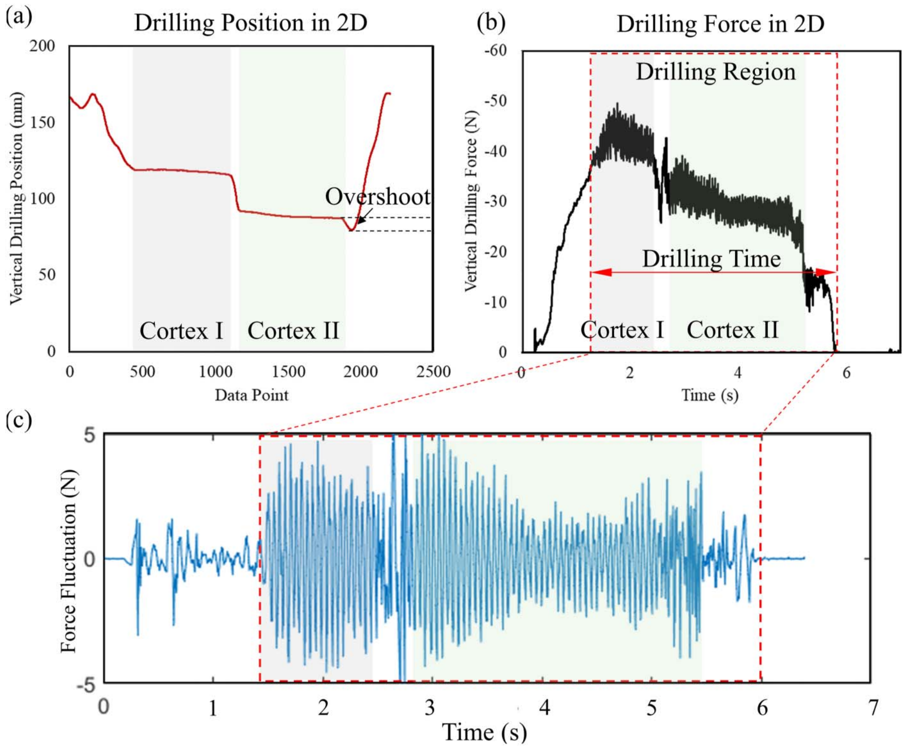


Fig. 2

Sample data plots for vertical drilling: Gray and green shaded sections indicate first and second cortical sections, respectively. (**Fig. 2-A**) Sample position plot for vertical drilling; (**Fig. 2-B**) sample force plot for vertical drilling; and (**Fig. 2-C**) extracted force fluctuation profile.

within and outside of the authors' institutions. Two groups (experts and novices) of the same size were established for the construct validity study, whereas a third group was added to test whether this method can further differentiate the intermediate skill level.

Expert group (N = 10): graduated and practicing orthopaedic surgeons;

Novice group (N = 10): residents postgraduate year-1 and medical students;

Intermediate group (N = 5): residents postgraduate year-2 to postgraduate year-5.

The testing procedure involved drilling vertically through the bone surrogate, as shown in Figure 1-B. To prevent any potential bias in participants' behavior, no data examples or metrics were provided. The participants were only instructed to drill 5 consecutive holes at the maximum rotational speed (about 1,200 rpm) through both cortical sections as if they were in a clinical setting. Participants also needed to pause for a few seconds between each drilling. Furthermore, 2 or more practice trials were given at the start of the task to familiarize participants with the ergonomics of the setup and thus make sure the system successfully characterized their drilling. No data or feedback was provided to the participants for these practice trials as that could have affected the performance of the subsequent data collection.

Data Recording and Extraction

Typical position and force data for an individual vertical drilling trial are illustrated in Figures 2-A and 2-B. The cortical drilling was identified by 2 steady slopes in the position plot and 2 steady peaks in the force plot (shaded in the figure). Three data features were extracted from a single trial, namely overshoot, drilling time, and force fluctuation. The "overshoot" represents the distance traveled by the drill tip beyond the far cortical section. It was identified by the downward drop after the second cortical section in the position plot. The "drilling time" represents the total duration for which the drill bit is in contact with the bone surrogate, excluding the setup time of the drill and guide. This was identified from the start of the drilling force to its end (as indicated by the vibrations in the force profile). Finally, force fluctuation is a measure of the subtle oscillations in the drilling region of the force profile, caused by unsteady hand movements or reluctance during the drilling trial. To evaluate this, the 1-dimensional wavelet decomposition technique was employed using MATLAB, which separates the force profile into several subcomponents that represent the overall shape of the profile and the minor fluctuations within it. By excluding the dominant profile shape and the consistent noise of the drill motor, the resulting force fluctuations attributable to unsteady hand movements were isolated, as shown in Figure 2-C. The force fluctuation was quantified by computing the root mean square of the resultant force fluctuation profile, which provides a measure of the average amplitude variation of a signal.

Data Visualization and Statistics

Three data features, namely overshoot, drilling time, and force fluctuation, were computed for each individual trial. The

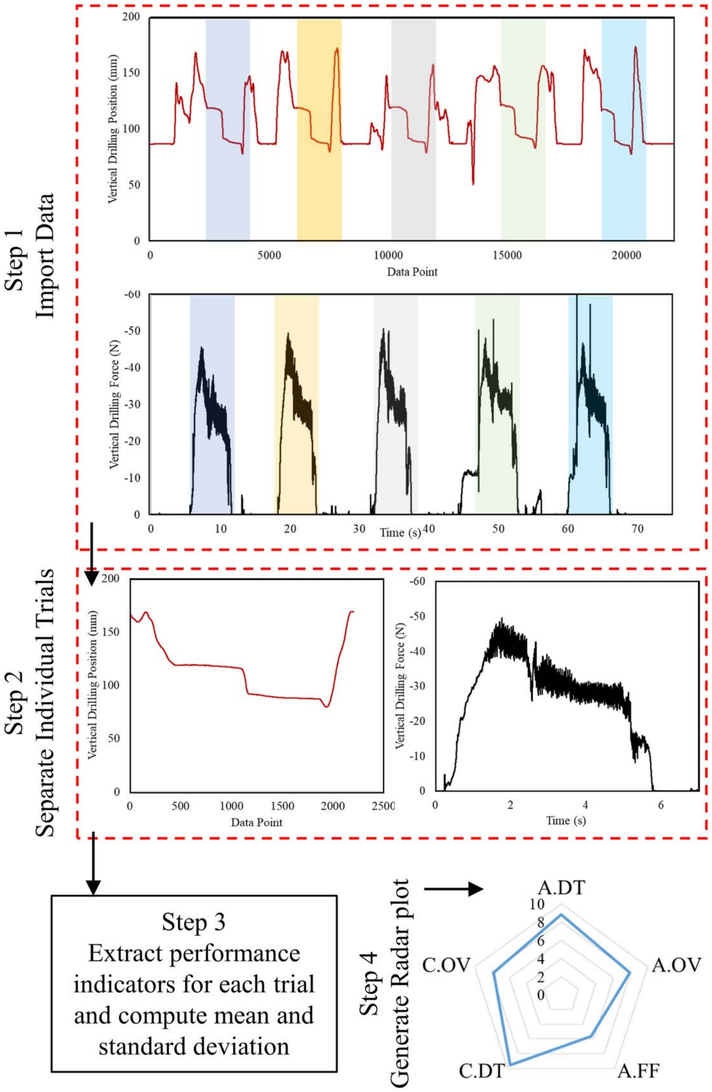


Fig. 3

Flowchart outlining the steps involved in the data analysis process to generate the radar plots (shaded region indicates the different drilling trials).

subsequent step involved computing the average and SD of overshoot and drilling time to assess consistency across the 5 trials. This approach led to the identification of 5 performance indicators, namely, average overshoot (AOV), average drilling time (ADT), overshoot consistency (COV), time consistency (CDT), and average force fluctuation (AFF).

The data analysis process comprised 4 major steps, as illustrated in Figure 3. In Step 1, the position and force data were imported into MATLAB code. These data sets represented 5 distinct drilling trials merged together for a specific participant. In Step 2, the individual drilling trials were extracted from these data sets and analyzed independently. In Step 3, overshoot, drilling time, and force fluctuation were extracted for each trial, and the average and SD for overshoot and drilling time were calculated. Finally, in Step 4, all individual performance indicators were normalized (scored from 0 to 10), and radar plots were generated.

One of the most important aspects of the methodology was the ability to visualize performance outcomes on a normalized scale. Therefore, the scoring for each performance indicator should not only recognize the skill level but also be sensitive enough to visually detect minor performance differences within a skill level. This study used a database of 25 participants, consisting of 10 experts, 5 intermediates, and 10 novices, and thus a total of 125 drilling trials to define the limits of the assessment metric. For each performance indicator (x), the lower limit (L) was the minimum value in the data set, whereas the upper limit (U) was computed using the mean plus 2 SDs (95% confidence interval) in the data set. This mathematical setting was chosen to avoid negative L values. Individual performance indicators were scored based on Equation 1:

$$N = 10 \cdot \left[\frac{U - x}{U - L}\right] \qquad (1)$$

TABLE I Reference Values Used to Obtain Normalized Scoring for Each Performance Indicator*

Score	ADT (s)	AOV (mm)	AFF (N)	CDT	COV
10	3.27	6.22	0.51	0.14	0.53
9	5.07	7.00	0.71	0.38	0.88
8	6.87	7.78	0.90	0.63	1.23
7	8.68	8.56	1.10	0.87	1.58
6	10.48	9.34	1.30	1.12	1.93
5	12.29	10.11	1.50	1.36	2.28
4	14.09	10.89	1.69	1.60	2.62
3	15.89	11.67	1.89	1.85	2.97
2	17.70	12.45	2.09	2.09	3.32
1	19.50	13.23	2.28	2.33	3.67
0	21.31	14.01	2.48	2.58	4.02

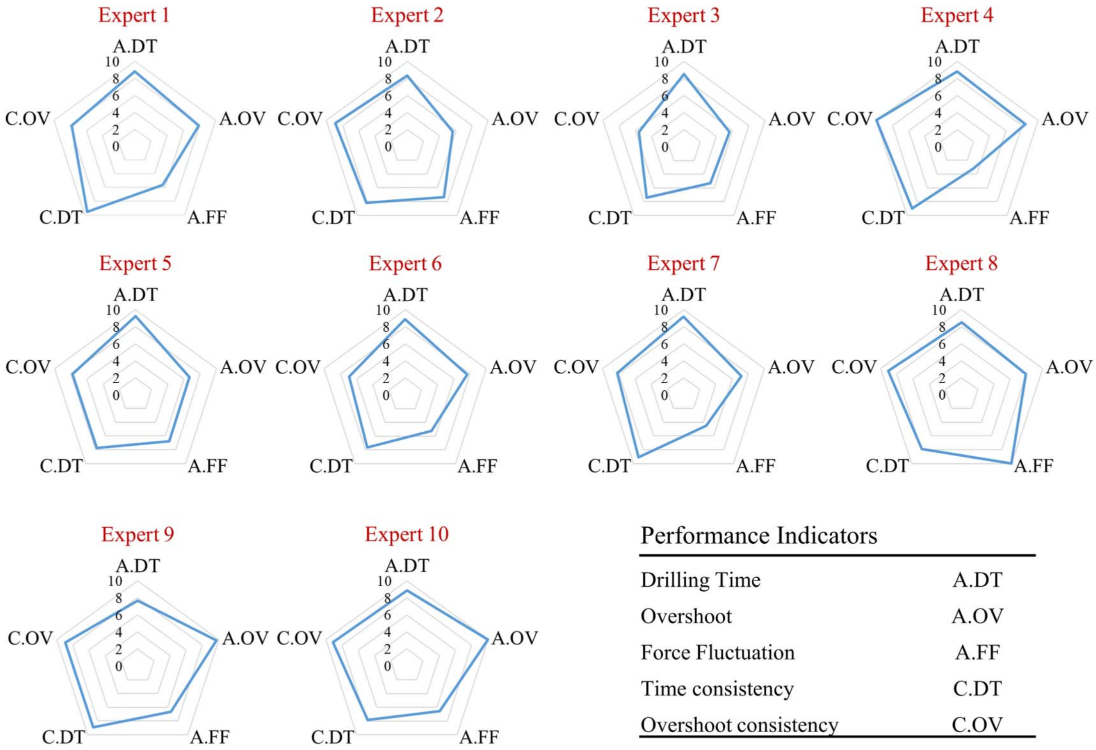
*Five performance indicators are included: average overshoot (AOV), average drilling time (ADT), overshoot consistency (COV), time consistency (CDT), and average force fluctuation (AFF).

The scoring ranged linearly from 0 to 10, with 0 being the lowest and 10 being the highest score. Note the scores are simply references and do not have physical meaning of being good or bad. The score-to-performance-indicator conversion is listed in Table I.

Results

The expert results based on 5 performance indicators: AOV, A.DT, COV, CDT, and AFF are plotted in Figure 4. They demonstrated a superior performance across all performance indicators, with the radar plots closely resembling a well-formed pentagon. To visualize the difference in performance between the expert participants and the others, an "expert band" was created using the expert scores by adding and subtracting one SD from the mean values.

As evident from the radar plots in Figure 5, the novice performances were much different than the experts. The novice participants demonstrated a selective focus on either drilling time or overshoot, leading to suboptimal outcomes. For instance, Novices 3, 5, 6, 7, 9, and 10 displayed higher scores in drilling time, yet their overshoot and force fluctuation were comparatively weaker. On the other hand, Novices 1 and 2 displayed higher scores in overshoot but were much weaker in drilling time. Novice 4 demonstrated a well-balanced performance, similar to an expert, but still remained outside the bounds of the expert band. It is also noted that the expert group does not necessarily have a higher score in force fluctuation. This is because of large variations in overall AFF data and a high correlation with drilling time in the novice group. Therefore, the balance of the pentagon shape is important. If using the pentagon area as a sole parameter to present the shape, the difference between the novices (N = 10) and experts (N = 10) is statistically significant (p value = 0.00008) at the statistical power of 88.9% for the 95% confidence.

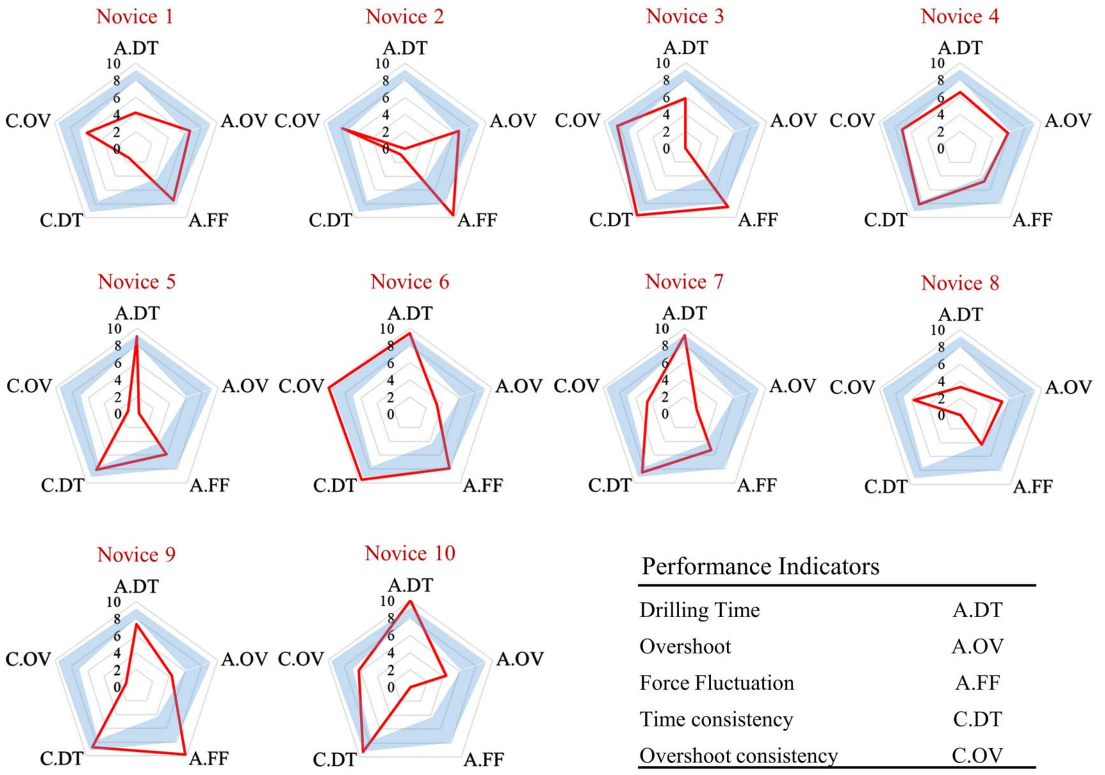


Performance Indicators

Drilling Time	A.DT
Overshoot	A.OV
Force Fluctuation	A.FF
Time consistency	C.DT
Overshoot consistency	C.OV

Fig. 4

Radar plots of expert participants.



Performance Indicators

Drilling Time	A.DT
Overshoot	A.OV
Force Fluctuation	A.FF
Time consistency	C.DT
Overshoot consistency	C.OV

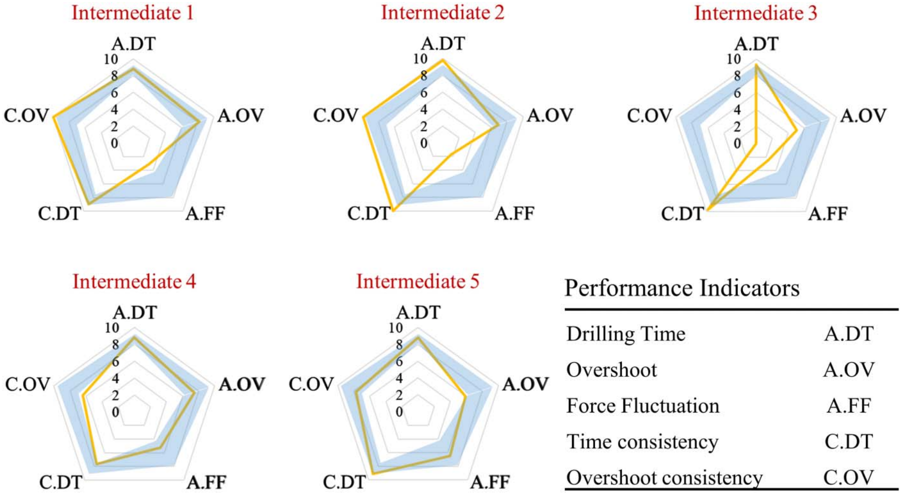
Fig. 5

Radar plots of novice participants in red; blue shaded region depicts the expert band.

Figure 6 shows the data from the intermediate group. They participants displayed improved performance, with their radar plots more closely resembling a regular pentagon (Fig. 6). Intermediate participants 1 and 2 achieved a good balance between drilling time and overshoot, but their scores in force fluctuation were comparatively weaker. Intermediates 4 and 5 demonstrated an overall balanced performance on all performance indicators, yet still fell slightly outside the expert band.

Discussion of Results

The radar plots provide an easy-to-interpret visualization to holistically evaluate the differences within and among different skill levels. The results of this study revealed that nonexpert groups exhibited more distorted radar plots, in contrast to the well-formed pentagons observed in the case of expert participants. In addition to the visual differences in Figures 4–6, individual indicators were also statistically assessed between groups using the expert group as the baseline. Two-sample *t* test confirmed the difference (p value <0.05)



Performance Indicators

Drilling Time	A.DT
Overshoot	A.OV
Force Fluctuation	A.FF
Time consistency	C.DT
Overshoot consistency	C.OV

Fig. 6

Radar plots of intermediate participants in yellow; blue shaded region depicts the expert band.

in drilling time, overshoot, and force fluctuation, justifying their use as performance indicators in the radar plot.

On the other hand, it is essential to acknowledge that the use of radar plots exhibits variability, depending on the definition of the upper and lower bounds of scoring, as well as the arrangement of each performance indicator on the plot. Various approaches can be used to define the scoring system, but given that all performance indicators are uniformly scaled, the intended function of the radar plot or statistical analysis remains unaffected. The most important factor in creating the radar plot is its visual representation. Therefore, the scoring must be sufficiently sensitive to perceive variation while also being wide enough to encompass most scenarios. In addition, it should be noted that although there is currently no rigorous mathematical framework for analyzing the radar plot, the use of the "expert band" provides a relatively effective means of quantifying the comprehensive comparison across various skill levels. Other shape parameters besides the area, such as aspect ratio or angle variation, can be used to gauge the shape distortion and run statistical comparison.

In practice, the radar plots may have psychological benefits, providing more meaningful guidance to novice trainees. First, although radar plots present the same information as their parent raw data, they are more visually appealing. In cognitive science, several studies have endorsed the use of visual aids and graphical representation to enhance the effect of quantitative data on human perception[17-19]. Second, radar plots can help to highlight patterns and relationships in multivariate data that may not be easily discernible from a table of numbers, such as the relationship between overshoot and drilling time. This can aid trainees in quickly understanding the overall picture of the data and improving their performance. In addition, the presence of an expert band on a radar plot may foster competition among junior residents, encouraging repeated practice and ultimately improving performance in the presence of quantitative data.

The current teaching scenario in orthopaedic residency programs lacks an objective methodology for providing corrective feedback to improve surgical skills[20]. The proposed methodology and use of radar plots offer 2 potential approaches to enhance drilling skills in a training setting. First, experts can use radar plots to demonstrate proper drilling techniques and factors contributing to optimal performance. Novices can observe and learn from the radar plot under expert guidance and practice consistently to improve their skills. Second, novices can use the expert radar plot as a benchmark for skill mastery. The current system costs around $10,000 USD with the force sensor, robotic arm, data acquisition, and computer, whereas this can vary depending on the precision and accuracy needed for all the analyses. The ideal implementation model is a self-directed learning scheme based on the feedback from the system (radar plots).

This study has several limitations, which could shape future research endeavors. First, the setup is relatively simple and only focuses on vertical drilling. It should be noted that the computational core of the setup could be extended to oblique drilling (or more complex surgical procedures). Additional performance parameters, such as approach angle and angular deviation, can be integrated into the performance matrix and radar plots with polygons. Second, a more uniform set of expert participants may help narrow the expert band for a better representation of a benchmark for novice participants. Finally, it is imperative to examine the long-term effects of this training and assessment platform on junior residents by conducting a multiphase feedback study that evaluates the influence of quantitative data in the form of radar plots on sustained performance improvement.

Conclusion

The research suggests that the use of radar plots provided an easy-to-interpret method to collectively visualize all performance indicators and deficiencies on 1 scale. Results showed that expert participants had a more balanced control compared with intermediate and novice participants when considering the shape of all performance indicators on the plot. The findings further endorse the construct validity of radar plots in highlighting the disparities among and within different skill levels. Future work will concentrate on refining the expert band and examining the long-term effects of this comprehensive methodology on junior residents' drilling performance over time. This work also serves as the foundation for the future clinical trial in residency training. ■

Note: This study was funded by the Orthopedic Research and Education Foundation (OREF-0014566A).

Aman Nigam, PhD[1]
James F. Kellam, MD[2]
Catherine G. Ambrose, PhD[2]
Bruce L. Tai, PhD[1]

[1]Department of Mechanical Engineering, Texas A&M University, College Station, Texas

[2]Department of Orthopaedic Surgery, University of Texas Health Science Center at Houston (UTHealth), Houston, Texas

E-mail address for B.L. Tai: btai@tamu.edu

References

1. Ding A, Marmor M. Drilling technique can minimize plunging. J Orthopaedic Trauma. 2019;33(8):e309-e312.

2. Zirkle M, Taplin MA, Anthony R, Dubrowski A. Objective assessment of temporal bone drilling skills. Ann Otol Rhinol Laryngol. 2007;116(11):793-8.

3. Close MF, Mehta CH, Liu Y, Isaac MJ, Costello MS, Kulbarsh KD, Meyer TA. Subjective vs computerized assessment of surgeon skill level during mastoidectomy. Otolaryngol Head Neck Surg. 2020;163(6):1255-7.

4. Nash R, Sykes R, Majithia A, Arora A, Singh A, Khemani S. Objective assessment of learning curves for the Voxel-Man TempoSurg temporal bone surgery computer simulator. J Laryngol Otol. 2012;126(7):663-9.

5. Feldman LS, Hagarty SE, Ghitulescu G, Stanbridge D, Fried GM. Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. J Am Coll Surg. 2004;198(1): 105-10.

6. Vankipuram M, Kahol K, McLaren A, Panchanathan S. A virtual reality simulator for orthopedic basic skills: a design and validation study. J Biomed Inform. 2010;43(5):661-8.

7. Kazum E, Dolkart O, Rosenthal Y, Sherman H, Amar E, Salai M, Maman E, Chechik O. A simple and low-cost drilling simulator for training plunging distance among orthopedic surgery residents. J Surg Educ. 2019;76(1):281-5.

8. Zirkle M, Roberson DW, Leuwer R, Dubrowski A. Using a virtual reality temporal bone simulator to assess otolaryngology trainees. Laryngoscope. 2007;117(2):258-63.

9. Lin Y, Chen H, Yu D, Zhang Y, Yuan W. A predictive bone drilling force model for haptic rendering with experimental validation using fresh cadaveric bone. Int J Comput Assist Radiol Surg. 2017;12(1):91-8.

10. Nigam A, Mohanty RR, Kellam JF, Ambrose CG, Krishnamurthy VR, Tai BL. An objective assessment for bone drilling: a pilot study on vertical drilling. J Orthop Res. 2022;41(2):378-85.

11. Harrington DT, Miner TJ, Ng T, Charpentier KP, Richardson P, Cioffi WG. What shape is your resident in? Using a radar plot to guide a milestone clinical competency discussion. J Surg Educ. 2015;72(6):e294-e298.

12. Morales-Silva DM, McPherson CS, Pineda-Villavicencio G, Atchison R. Using radar plots for performance benchmarking at patient and hospital levels using an Australian orthopaedics dataset. Health Informatics J. 2020;26(3):2119-37.

13. Stafoggia M, Lallo A, Fusco D, Barone AP, D'Ovidio M, Sorge C, Perucci CA. Spie charts, target plots, and radar plots for displaying comparative outcomes of health care. J Clin Epidemiol. 2011;64(7):770-8.

14. Tai BL, Kao YT, Payne N, Zheng Y, Chen L, Shih AJ. 3D Printed composite for simulating thermal and mechanical responses of the cortical bone in orthopaedic surgery. Med Eng Phys. 2018;61:61-8.

15. Tai BL, Wang AC, Joseph JR, Wang PI, Sullivan SE, McKean EL, Shih AJ, Rooney DM. A physical simulator for endoscopic endonasal drilling techniques: technical note. J Neurosurg. 2016;124(3):811-6.

16. Mohanty RR, Vyas S, Nigam A, Tai BL, Krishnamurthy VR. Orthopedic bone-drilling assessment through laplacian-based trajectory noise characterization. Proceedings of the ASME 2021 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 2: 41st Computers and Information in Engineering Conference (CIE). Virtual, Online. August 17–19, 2021. V002T02A060. ASME.

17. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns. 2006;61(2):173-90.

18. Borgo R, Abdul-Rahman A, Mohamed F, Grant PW, Reppa I, Floridi L, Chen M. An empirical study on using visual embellishments in visualization. IEEE Trans Vis Comput Graph. 2012;18(12):2759-68.

19. Vande Moere A, Tomitsch M, Wimmer C, Christoph B, Grechenig T. Evaluating the effect of style in information visualization. IEEE Trans Vis Comput Graph. 2012;18(12):2739-48.

20. Carpenter JE, Hurwitz SR, James MA, Jeffries JT, Marsh JL, Martin DF, Murray PF, Parsons BO, Pedowitz RA, Toolan BC, Van Heest AE, Wongworawat AD. ABOS surgical skills modules for PGY-1 residents. Available at: https://www.abos.org/residents/residency-skills-modules/abos-surgical-skills-modules-for-pgy-1-residents/. Accessed May 9, 2023.