***Title:*** **Using Eye-Tracking Technology to Quantify the Effect of Experience and Education on Forensic Anthropological Analyses**

***Authors:*** Micayla C. Spiros, M.Sc.[1], Sherry Nakhaeizadeh, Ph.D.[2,3], Tim J.U. Thompson Ph.D.[4], Ruth M. Morgan D.Phil.[2,3], Martin Arvidsson, Ph.D.[5], Viktor Olsson, M.A.[5], Alexandra Berivoe, B.A.[5], Joseph T. Hefner, Ph.D.[1]

***Affiliations:*** [1] Department of Anthropology, Michigan State University, 655 Auditorium Drive, East Lansing, MI 48824, United States
[2] Department of Security and Crime Science, University College London, 35 Tavistock Square, London, WC1H 9EZ, United Kingdom
[3] Centre for the Forensic Sciences, University College London, 35 Tavistock Square, London, WC1H 9EZ, United Kingdom
[4] School of Health and Life Sciences, Teesside University, Middlesbrough, TS1 3BX, United Kingdom
[5] Tobii Pro Insight, Karlsrovägen 2D, 82 53 Danderyd, Stockholm, Sweden

***Corresponding author***: Micayla Spiros, Department of Anthropology, Michigan State University, 655 Auditorium Drive, East Lansing, MI 48824; Email: spirosmi@msu.edu

***Acknowledgments***: This study was determined to be exempt from human studies regulations by the Michigan State University Institutional Review Board under the Human Research Protection Program Flexibility Initiative Exemption Category 98. The authors wish to extend our thanks to all of our participants, we appreciate the time you took out of your day to take part in this study.

**ABSTRACT**: The very human interpretation of analytical outputs is a significant challenge in forensic science making it vital to explore the application of protocols as we enhance our practices. This study assesses decision making in forensic anthropological analyses utilizing eye-tracking technology to quantify an observer's estimate of confidence and reliability. Ten individuals with varying levels of education and experience were asked to score cranial morphologies for two human crania. Each participants' fixation points, fixation duration, and visit count and duration were assessed using Tobii™ Pro 2 eye-tracking glasses. Mid-facial morphologies capturing relative widths were the quickest scored traits with an overall median time of 14.59 seconds; more complex morphological assessments took longer. Using time as a proxy for confidence, Kruskal-Wallis rank sum results indicate individuals with less experience differed significantly from individuals with greater experience ($p = 0.01$) although differences in level of education were not significant. Interestingly, intraclass correlation coefficients (ICC) indicate interobserver reliability is high between observers, suggesting experience only slightly improves agreement. These preliminary results suggest experience is more important than level of education. Through empirical decision making studies, forensic anthropologists can improve practices—decreasing participant differences by targeting confusing or problematic aspects of a data collection practice and improving training protocols.

**KEYWORDS:** Forensic Anthropology, Forensic Science, Subjectivity, Objectivity, Confidence, Decision making, Protocol Improvement

Forensic anthropology is a field rooted in the visual assessment of shape, collecting and recording all forms of data related to human variation with the purpose of assisting medicolegal death investigators in the identification of a decedent. These visual assessments remain a large component of current practice when estimating aspects of the biological profile, including age (1-6), sex (7,8), and population affinity (9-13). Many of these methods utilize morphological traits to estimate some part or component of the biological profile. Some of these protocols are inherently subjective, requiring assessments based on both codified and tacit knowledge—for example, the experience of the observer (14-16). In recent years, there have been a number of studies assessing the decision making processes involved in the interpretation and analysis of skeletal remains (15-18). Many of these studies focused on the effect of cognitive and contextual biases (16-18) and most highlighted the need for developing a greater understanding of the decision making strategies involved in the collection, assessment, and interpretation of human skeletal remains (18). In forensic science, there also has been an increased engagement with—and a rapidly growing evidence-base addressing—decision making and the human factors affecting interpretations in forensic reconstruction approaches (19).

The challenges arising in human decision making have been documented in the published literature (15-19, 24, 37-47). However, evaluating decision making strategies by measuring the gaze patterns of the human actors making those assessments has not yet been fully evaluated (15). Defining an expert in forensic science varies worldwide and is not universally agreed. Generally speaking, in forensic anthropology the minimum expectation for a forensic anthropology expert witness in court is a doctoral degree or equivalent forensic anthropological experience. Some regions and countries (e.g. Latin America, United Kingdom, United States) establish forensic anthropological professional bodies and a board certification process for forensic anthropologists

(20-23). Yet, empirical studies to determine the role of expertise and experience in how visually-derived information is captured for sex, age, population affinity, trauma, or taphonomic protocols are only moderately addressed or researched. As with many expert observations (24), the data captured for forensic anthropological analyses are minutiae of visually-derived information that are difficult to teach to those with limited osteology experience. Consequently, training forensic anthropologists in visual processing protocols and describing the procedures used to capture these visual cues is difficult. Eye-tracking research can provide greater insight into forensic anthropological methods and applications. For example, we can use gaze pattern and duration to identify weaknesses in the definitions of specific morphologies or in a general protocol to refine definitions or to enhance training (21,26-30).

The aim of this paper is to demonstrate how eye-tracking technology can grant further insights into the application and, in turn, education of forensic anthropological protocols. We use cranial morphology to direct the participants, to test the eye-tracking technology, and to analyze gaze pattern data. This approach has been tested for other regions of the skeleton and using other types of data (e.g., pelvic morphology). We hypothesize that using visual methods of cranial morphology would deliver similar results to the assessment on age and sex methods (15).

New insights into the challenges faced in the collection of morphological data can be identified by studying eye gaze behavior, including shortcomings in current methods and practices. Here, we use cranial morphology to identify shifts that may be necessary when teaching students to assess and collect these types of data. We also identify whether there is a difference in the time to score and the actual scores between individuals having different levels of experience and education. This study assesses whether observers with more experience/education felt more confident in their assessments than those with less experience/education. Finally, we assessed the consistency between observers using eye-tracking technology to identify why some observers more consistently agreed and whether those differences reflected a nuanced understanding of a morphological feature, the observer's experience, or resulted from a poorly defined morphology.

*Eye-tracking research in the forensic sciences*

Eye-trackers capture eye movements and enable the collection of data related to how long participants view areas of interest (15,31-33). This technology has been applied in various fields to assess practitioner performance (31,33). Eye-tracking technology has generally dominated psychology research, but has recently been utilized within a number of other scientific disciplines for educational purposes, protocol development, proficiency testing, and cognition studies (34-37). The use of eye-tracking technology as a research tool to study decision making in criminal investigations and other forensic sciences has been utilized, but for only a few published studies covering fields such as handwriting analysis, fingerprint examination, identifications in criminal line-ups, blood spatter analysis, and general crime scene investigation (34,36,38-40). The results of these studies provide insight into an experts' gaze fixations and the areas of interest when evaluating evidence (34) and in the application of search strategies to process crime scenes (40) As an example, studies looking at forensic document examiners demonstrate experienced examiners are more accurate than lay persons; simply put they are just significantly better at identifying counterfeit signatures (39,41). In one of those studies, eye-tracking technology was used to record each participants' eye movements and their response times (39). The eye movement and search pattern data for all subjects showed similarities in search strategy (how they looked and where); however, forensic document examiners took double the amount of time to reach their final decision suggesting the key to distinguishing between forgeries and disguises in signatures is in some part related to a more careful inspection of the item and longer consideration of multiple features in the item of interest (39). Another study used eye-tracking technology to quantify consistency and variability among forensic experts, showing experts were more consistent than novices when inspecting and describing the features they used for latent fingerprint analysis (38). Additionally, search duration and search sequence between expert crime scene investigators and inexperienced novices differed significantly; experts are much more consistent in the search sequence compared to a novice group (40).

Only one study to date has used eye-tracking technology to study gaze pattern strategies among forensic anthropologists analyzing skeletal remains (15). In that study, eye-tracking technology focused on nonmetric features used in sex and age-at-death estimations. That research quantified analyst gaze fixation points, fixation duration, and visit counts for the interpretation of features on the skull and os coxa. Building on that research, we use cranial morphological features to assess gaze patterns, gaze duration, and fixation among a sample of individuals with varying levels of experience and education.

## Materials and Methodology

### *Experiment Design*

Cranial morphological data were collected from two human skulls by participants wearing Tobii™ Pro 2 eye-tracking glasses. This wearable eye-tracking camera recorded the pattern of visual attention of each participant by directing near infrared light on the eyes, identifying the focus point, duration of focus, and fixation pattern for each cranial morphology on each skull for each participant. Each participant was asked a series of questions to identify their level of experience with cranial morphological trait data (less than or greater than 2 years) and level of education (undergraduate, graduate with Master's degree, professional with PhD). The mean number of years of experience with cranial morphology was two years for all participants and, as such, was used as the sectioning point for experience. Individuals with master's degrees were further divided according to their experience assessing cranial morphology. These demographics may identify factors driving any discrepancies between observers. Each observer's level of education was also collected, to evaluate the relationship between the visual collection of cranial morphological data and an observer's level of education. These data should situate the visual gaze pattern, the visual acuity, and the duration of an observer's gaze in a broader context to identify trends.

Data collected from the glasses provide information on the users' eye fixation patterns. This includes 1) time recorded for each morphology or morphological region, 2) the overall time spent on each skull, and 3) the total duration of the analysis. In addition to these data, a visual representation of gaze fixation, visualized as a heat density map, can be generated for each participant and used to further assess gaze patterns and fixation.

### *Cranial Morphological Data*

Seventeen morphological traits of the skull were used to assess the impact of education and experience on gaze patterns (Table 1). Participants were asked to score these morphologies on two different skulls. The two skulls were selected to provide variability in trait expressions (11,44,49). Each participant was provided a data collection sheet with line drawings of each trait. This allowed the eye-tracking glasses to capture exactly what participants were looking at in real time as they made their assessments. Cranial morphological data were divided into two divisions: those assessed in a single, linear direction and related to size or breadth [unidimensional] and those assessed minimally in two directions that capture shape [multidimensional].

***[Table 1 here]***

## *Participants and Procedures*

Ten participants were asked to wear the eye-tracking technology and score the individual skulls. Participants with varying levels of experience were recruited. The participants ranged in general levels of practical (<1-32 years) and analytical experience (0-18 years). Practical experience includes all forensic anthropological experience while analytical experience only considers cranial morphology. Two undergraduate students, six graduate students with Master's degrees, and two individuals with doctoral degrees participated. Six individuals reported having less than two years of experience collecting cranial morphology data; the remaining participants had more than two years of experience. This sample represents a preliminary usability study and is suitable for assessing data collection protocols and protocol efficiency (52).

Each participant conducted the analysis separately in a laboratory. All necessary equipment was provided. The two skulls were situated on a table and presented to the participants simultaneously. The mandible was present for both even though no mandibular traits were considered. To minimize any potential influence on the decision making process, participants in this study were not told to start the analysis on a specific skull. Instead, participants were free to choose. Participants were asked to use the scoring sheet to record all answers. The scoring sheet presented seventeen cranial morphological traits in alphabetical order. Each participant was asked to provide their confidence for each score to quantify self-assurance in the interpretation of these morphologies. No time limit was imposed on the participants.

*Analysis*

Metadata collected with eye-tracking technology provides novel information on the collection of cranial morphological data. However, three questions can be addressed using the demographic data. First, can we identify differences in the time to completion and the morphological scores between individuals with more experience? Second, are participants with more experience always more confident? And finally, regardless of confidence, are scoring procedures consistently reproducible across participants?

To facilitate analysis, we also generated an image in a vector graphics editor to highlight the area around each of the cranial morphological features or areas (Figure 1). This allowed the recording of metrics and count data for each region, using the images as a baseline for reference. Visualizations and metrics documented where participants were looking (gaze fixation), how long they were looking (gaze duration), and if participants were going back to certain traits more than once (visit counts). These data are used to generate a heat map to visualize gaze patterns.

***[Figure 1 here]***

Several measures of confidence were used to assess how each participant assessed cranial morphology and whether their reported confidence matched their gaze pattern. These include: 1) heat maps to visualize education/experience-level variation; 2) fixation duration as a proxy for decision making measured as the overall time to completion and the amount of time spent on each cranial morphology; and, 3) real-time decision making and confidence assessed through the eye tracking software with ad-hoc confidence scores situating the implicit and explicit assurance in the collection cranial morphological data. Finally, after data collection we calculated an intraclass correlation coefficient (ICC) to quantify the association between participant scores. ICC does not require a 'correct' score, rather ICC assesses the reliability among all individuals and within each sub-group of the data (i.e., education or experience levels).

*Statistical Analyses*

Using time-to-score as a proxy to measure the observer's level of confidence, a Kruskal-Wallis rank sum test assessed differences between the various groups. Kruskal-Wallis is a non-parametric multiple-comparison test approximating a chi-square distribution to compare two or more groups. Summary statistics were calculated by fixation duration to understand variation among confidence levels, by trait. A two-way mixed-effects model for ICC was applied to assess observer agreement between the cranial morphological traits for multiple participants, and assessed following Koo and Li (53). Interclass correlation coefficients can range from 0.0 to 1.0 (where 1 is perfect agreement between observers). To assess self-reported confidence rates, observers were also asked to provide a measure of their confidence between 1 (not confident) and 10 (very confident) for each morphology. The medians of each trait were recorded by years of experience (less than 2 years or more than 2 years).

## Results

### *Density maps*

Density maps were created by concatenating the fixation and duration times of all analysts. These gaze patterns were combined into a single density map by individual, by education level, and finally by years of experience (Figure 2) to visualize eye-tracking data. Darker areas indicate higher levels of attention.

***[Figure 2 here]***

### *Fixation Duration*

The results for the Kruskal-Wallis test indicate experience is the only variable with significantly different duration times (Table 2).

***[Table 2 here]***

Participants with a master's degree were slightly faster than those with doctorate degrees and both were faster than the undergraduate cohort (Figure 3). Separating the masters-level group into two subgroups (one with < 2 years of experience and one with > 2 years of experience), those with more experience were faster than those with less experience (Figure 4). Although these differences do not reach the level of statistical significance they may indicate participants with more experience arrive at a decision faster than others (Figure 5).

***[Figure 3 here]***

***[Figure 4 here]***

***[Figure 5 here]***

Next, each cranial morphological trait was analyzed individually, using time as a proxy to measure confidence. Figure 6 highlights summary statistic data for each trait. The slowest trait for the participants to score covers a larger area of the midfacial region and is more akin to multidimensional, morphological data (NBC) compared to the fastest which captures unidimensional, linear data (SNS). The median time to score was 27.97 seconds. Of the seven unidimensional traits, four traits (NAW, SNS, NO, ANS) fell below the median speed and three traits (PZT, MT, IOB) fell above the median. Of the ten multidimensional traits, four (NFS, OBS, ZS, NAS) fell below the median speed while six traits (PS, INA, PBD, NBS, TBS, NBC) were scored at a slower pace falling above the median. To most accurately capture confidence, all potential outliers (identified in Figure 6) were retained as they offer great insight into the variance between observers.

***[Figure 6 here]***

### *Individual Confidence Ratings*

Self-reported confidence, divided by the median score for each cranial morphology, (Figure 7), illustrates variability in various levels of experience scoring these traits. More experienced participants were most confident (6.5 to 9). Those with less experience had median self-reported confidence levels ranging from 5 to 8. Interestingly, participants with more experience were only more confident for 14 of the 17 traits. The three traits those with less experience were more or equally confident in compared to more experienced raters were multidimensional traits (OBS, INA, and PBD).

***[Figure7 here]***

*Interobserver Reliability*

Table 3 provides the ICC data. The correlation coefficients ranged from 0.72 to 0.96. (Table 3).

***[Table 3 here]***

## Discussion

This study assessed eye-tracking technology as a tool to quantify how experience and education influence participant decision making and to visualize their gaze patterns when assessing cranial morphology. Acknowledging that participant sample sizes were limited, though appropriate for protocol efficiency testing, the results of this study still demonstrate how the level of experience with scoring protocols has a direct impact on fixation and duration times. The eye-tracking data was used to visualize gaze fixation and to generate data for quantifying gaze fixation and duration for all participants and for each cranial morphology.

To identify differences in the amount of time it takes to score each skull and the scores each observer selected, fixation duration by group was analyzed. Individuals with more experience (>2 years) elicited quicker response times and it appears experience is an important contributing factor to the decision making process. Individuals with more than two years of experience were overall faster than individuals with less than 2 years of experience. Assessing the overall individual morphologies, median time to score all traits suggests unidimensional traits are more likely to be scored faster, or more confidently. All observers took considerably longer to score traits that assessed broad regions or had more complex, multidimensional morphologies, some more than 35 seconds longer than others (e.g. NAW and NBC). The results of the Kruskal-Wallis tests indicate statistically significant differences between fixation duration and experience ($p = 0.01$).

The unidimensional traits are generally quicker to score. However, individuals may be slower at scoring multidimensional traits, but those with less experience seem to have more confidence in their scores. Three of the four traits the less experienced raters reported the highest confidence are multidimensional (OBS, PBD, TPS) and yet, two of the three traits they are least confident scoring are unidimensional (NO, MT). More experienced observers were most confident scoring multidimensional (NFS, NBS, TPS, NBC) traits. Although, unlike less experienced observers, more experienced raters were less confident scoring a number of multidimensional traits (OBS, INA, PBD). These traits (OBS, INA, PBD) were the only three traits that less experienced raters were equally or more confident in scoring than the more experienced raters.

There are two possible explanations for these differences. First, multidimensional traits potentially take longer to score due to complexity (various angles or using tools). So, while they take longer to score observers feel more confident having conducted a more thorough analysis. Conversely, this observation may be an example of the Dunning-Kruger effect: participants with less experience or knowledge do not have insights into their potential shortcomings leading to more confidence than more experienced participants (54). Less experienced raters were most confident in multidimensional traits that took longer for them to score. In the current context, experienced raters potentially have seen more human variation giving them insight into the true range of variation meaning that that the drawings for these traits potentially do not encompass the true range of variation, influencing a subconscious bias that the more experienced users have when scoring the regions.

Finally, regardless of each observer's confidence, the derived scores for all of the cranial morphological traits were consistent. The ICC results indicate moderate reliability (ICC=0.72-0.74), particularly when comparing individuals with similar education (ICC=0.86) or experience levels (ICC= 0.86-0.92). When dividing participants into four groups based on both experience *and* education, the ICC results indicate good agreement (ICC = 0.82-0.88). These results all suggest experience is a key consideration for higher reliability between participants and suggests increases in experience lead to more consistent visualization of cranial morphological features.

## Conclusion

How experts make decisions, process visual cues, and interpret evidence is influenced by intrinsic and extrinsic factors (55). To understand these factors we can apply modern technologies like eye-tracking capabilities. Such efforts will allow us to quantify the degree of influence that experience and education have on practitioners and to develop more transparent approaches for forensic inference (15). New technologies are increasingly fusing the physical, digital, and biological realms. This fusion is exciting and will generate novel opportunities for research addressing human identification using, for example, the automated pattern analysis (56,57). In forensic anthropology, machine learning, including deep learning algorithms, now facilitate automated decisions on skeletal remains (58-60). However, to fully apply these technologies to the improvement of procedures in forensic anthropology, we must understand the factors that play a role in the interpretation process.

Similar to previous research using eye-tracking to assess aspects of the forensic sciences (15,34,36,38-40), this study documented the usability of eye-tracking technology as a research tool. That technology offers significant potential to understand the importance of certain factors (like experience) when observers are collecting subjective, morphological data. Undertaking further empirical research building on these data will provide insight into those factors impacting the decision making and interpretative processes involved in forensic anthropological methods.

Future research should test these factors using a larger number of participants, incorporating a broader variety of experience levels, and measuring what, if any, effect training has on confidence and consistency. Beyond cranial morphology, postcranial data and dental variation should be similarly treated, potentially even in combination with cranial morphological approaches (10,13,61). Eye-tracking data, in relation to multiple scoring modalities will be helpful to assess how scoring methods vary across different data modalities such as photos, 3D models, CT scans, and other virtual data (62-66). Finally, the pedagogical implications of these results require further exploration. The impact of modifying the teaching and training of forensic anthropologists in visual techniques needs robust assessment and remediation.

Eye tracking technology is the only way to objectively record, analyse, and interpret visual gaze behaviours. Without this technology, quantifying the time a researcher spends assessing a particular feature, trait, or region of the skull would not be possible. With eye-tracking technology we have been able to study and quantify each observer's eyes during data collection. The insight these data provide into the cognitive processes underlying any forensic anthropological analysis is exciting and has great potential to reveal patterns and analyst gaze behaviours heretofore unconsidered and most definitely unmeasured.

## References


1. Buckberry JL, Chamberlain AT. Age estimation from the auricular surface of the ilium: a revised method. *Amer J Phys Anthrop.*. 2002;119:231-239.

2. Hartnett KM. Analysis of age-at-death estimation using data from a new, modern autopsy sample-part I: pubic bone. *J Forensic Sci*. 2010;55(5):1145-1151.

3. Hartnett KM. Analysis of age-at-death estimation using data from a new, modern autopsy sample—Part II: Sternal end of the fourth rib. *J Forensic Sci*. 2010;55(5):1152-1156.

4. Lamendin H, Baccino E, Humbert JF, Tavernier JC, Nossintchouk RM, Zerilli A. A Simple Technique for Age Estimation in Adult Corpses: The Two Criteria Dental Method. *J Forensic Sci*. 1992 ;37(5):1373-9.

5. Langley-Shirley N, Jantz RL. A Bayesian approach to age estimation in modern Americans from the clavicle. *J Forensic Sci*. 2010;55(3):571-583.

6. Osborne DL, Simmons T, Nawrocki SP. Reconsidering the auricular surface as an indicator of age at death. *J Forensic Sci*. 2004;49(5):1-7.

7. Klales AR, Ousley SD, Vollner JM. A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *Amer J Phys Anthrop.* 2012;149(1):104-114.

8. Walker PL. Sexing skulls using discriminant function analysis of visually assessed traits. *Amer J Phys Anthrop.* 2008(136):39-50.

9. Berg GE, Kenyhercz MW. Introducing human mandible identification [(hu)MANid]: A free, web-based GUI to classify human mandibles. *J Forensic Sci*. 2017;62(6):1592-1598.

10. Edgar HJH. Estimation of ancestry using dental morphological characteristics. *J Forensic Sci*. 2013;58(s1):S3-S8.

11. Plemons A, Hefner JT. Ancestry estimation using macromorphoscopic traits. *J Academic Forensic Pathology*. 2016;6(3):400-412.

12. Scott GR, Pilloud MA, Navega D, d'Oliveira J, Cunha E, Irish DJ. rASUDAS: A New Web-Based Application for Estimating Ancestry from Tooth Morphology. *Forensic Anthropology*. 2018;1:18-31.

13. Spiros MC. Standardization of Postcranial Nonmetric Traits and Their Utility in Ancestry Analysis. *Forensic Anthropology*. 2019;2(1):29-44.

14. Morgan RM. Conceptualising forensic science and forensic reconstruction. Part I: a conceptual model. *Sci Justice*. 2017;57(6):455-459.

15. Nakhaeizadeh S, Morgan RM, Olsson V, Arvidsson M, Thompson T. The value of eye-tracking technology in the analysis and interpretations of skeletal remains: A pilot study. *Sci Justice*. 2020;60(1):36-42.

16. Warren MW, Friend AN, Stock MK. Navigating cognitive bias in forensic anthropology. Forensic Anthropology: Theoretical Framework and Scientific Basis. 2018:39-51.

17. Klales AR, Lesciotto KM, editors. The "science of science": examining bias in forensic anthropology. American Academy of Forensic Sciences; 2016 2016; Las Vegas, NV.

18. Nakhaeizadeh S, Morgan RM, Rando C, Dror IE. Cascading bias of initial exposure to information at the crime scene to the subsequent evaluation of skeletal remains. *J Forensic Sci*. 2018;63(2):403-411.

19. Earwaker H, Nakhaeizadeh S, Smit NM, Morgan RM. A cultural change to enable improved decision-making in forensic science: A six phased approach. *Sci Justice*. 2020;60(1):9-19.

20. Bartelink EJ, Boyd DC, France DL, Pokines JT, Prince-Zinni D. The American Board of Forensic Anthropology Turns 40: Historical Perspectives and Current Trends in Certification for Forensic Anthropology. *Forensic Anthropology*. 2020;3(2):103.

21. Boyd DC, Bartelink EJ, Passalacqua NV, Pokines JT, Tersigni-Tarrant M. The American board of forensic anthropology's certification program. *Forensic Anthropology*. 2020;3(2):112.

22. Kranioti E, Paine R. Forensic anthropology in Europe: an assessment of current status and application. *J Anthropol Sci*. 2011;89:71-92.

23. Passalacqua NV, Pilloud MA. The need to professionalize forensic anthropology. *Eur J Anat*. 2021;25:35-47.

24. Kneebone R. *Expert: Understanding the Path to Mastery*, Penguin UK, 2020.

25. Almazrouei MA, Dror IE, Morgan RM. The forensic disclosure model: what should be disclosed to, and by, forensic experts? *Int J Law*. 2019;59:100330.

26. Langley N, Tersigni-Tarrant M. Education, Training, and Competencies in Forensic Anthropology. *Forensic Anthropology*. 2020;3(2):65-65.

27. Langley NR, Tersigni-Tarrant M. Core Competencies in Forensic Anthropology: A Framework for Education, Training, and Practice. *Forensic Anthropology*. 2020;3(2):76.

28. Love JC, Langley NR. Exploring Educational Needs beyond Technical Competency: Laboratory Management, Court Testimony, and Vicarious Trauma. *Forensic Anthropology*. 2020;3(2):82.

29. Passalacqua NV, Pilloud M. Education and training in forensic anthropology. *Forensic Anthropology*. 2020;3(2):66.

30. Pinto DC, Pierce ML, Wiersema JM. A Model for Forensic Anthropology Training. *Forensic Anthropology*. 2020;3(2):91.

31. Duchowski AT. Eye movement analysis. *Eye Tracking Methodology*: Springer; 2017;141-158.

32. Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J. *Eye tracking: A comprehensive guide to methods and measures*: OUP Oxford, 2011.

33. Horsley, Mike, Matt Eliot, Bruce Allen Knight, and Ronan Reilly, eds. *Current trends in eye tracking research*. Springer Science & Business Media, 2014.

34. Arthur RM, Hoogenboom J, Green RD, Taylor MC, de Bruin KG. An eye tracking study of bloodstain pattern analysts during pattern classification. *Int J Legal Med*. 2018;132(3):875-885.

35. Ashraf H, Sodergren MH, Merali N, Mylonas G, Singh H, Darzi A. Eye-tracking technology in medical education: A systematic review. *Med Teach*. 2018;40(1):62-69.

36. Hicklin RA, Ulery BT, Busey TA, Roberts MA, Buscaglia J. Gaze behavior and cognitive states during fingerprint target group localization. *Cognitive research: principles and implications*. 2019;4(1):12.

37. Paul IJL, Sasirekha S, Maheswari SU, Ajith KAM, Arjun SM, Kumar SA. Eye gaze tracking-based adaptive e-learning for enhancing teaching and learning in virtual classrooms. *Information and Communication Technology for Competitive Strategies*: Springer; 2019;165-176.

38. Busey T, Yu C, Wyatte D, Vanderkolk J, Parada F, Akavipat R. Consistency and variability among latent print examiners as revealed by eye tracking methodologies. *J Forensic Ident*. 2011;61(1):60-91.

39. Dyer AG, Found B, Rogers D. An insight into forensic document examiner expertise for discriminating between forged and disguised signatures. *J Forensic Sci*. 2008;53(5):1154-1159.

40. Watalingam RD, Richetelli N, Pelz JB, Speir JA. Eye tracking to evaluate evidence recognition in crime scene investigations. *Forensic Sci Int*. 2017;280:64-80.

41. Dyer AG, Found B, Rogers D. Visual attention and expertise for forensic signature analysis. *J Forensic Sci*. 2006;51(6):1397-1404.

42. Cole SA. Acculturating Forensic Science: What is Scientific Culture, and How Can Forensic Science Adopt It. *Fordham Urb LJ*. 2010;38:435.

43. Morgan RM. Forensic science needs both the 'hedgehog'and the 'fox'. *Forensic Sci Int*. 2018;292:e10-e12.

44. Hefner JT. Cranial nonmetric variation and estimating ancestry. *J Forensic Sci*. 2009;54(5):985-995.

45. Hefner JT. *Assessing Nonmetric Cranial Traits Currently Used in the Forensic Determination of Ancestry* [MA]. Gainesville, FL: University of Florida, 2003.

46. Hefner JT. *The Statistical Determination of Ancestry Using Cranial Nonmetric Traits* [PhD]. Gainesville, FL: University of Florida, 2007.

47. Kamnikar KR, Plemons AM, Hefner JT. Intraobserver error in macromorphoscopic trait data. *J Forensic Sci*. 2018;63(2):361-370.

48. Klales AR, Kenyhercz MW. Morphological Assessment of Ancestry using Cranial Macromorphoscopics. *J Forensic Sci*. 2015;60(1):13-20.

49. Dunn RR, Spiros MC, Kamnikar KR, Plemons AM, Hefner JT. Ancestry estimation in forensic anthropology: A review. *Wiley Interdisciplinary Reviews: Forensic Science*. 2020:e1369.

50. Atkinson ML, Tallman SD. Nonmetric Cranial Trait Variation and Ancestry Estimation in Asian and Asian-Derived Groups. *J Forensic Sci*. 2020;65(3):692-706.

51. Kenyhercz MW, Klales AR, Rainwater CW, Fredette SM. The optimized summed scored attributes method for the classification of US blacks and whites: A validation study. *J Forensic Sci*. 2017;62(1):174-180.

52. Faulkner L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput. 2003;35(3):379-383.

53. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155-163.

54. Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the "signal extraction" counterexplanation for the Dunning Kruger effect in self-evaluation of performance. *J Econ Psych*. 2013; 39,85-100.

55. Georgiou N, Morgan RM, French JC. Conceptualising, evaluating and communicating uncertainty in forensic science: Identifying commonly used tools through an interdisciplinary configurative review. *Sci Justice*. 2020.

56. Rajpurkar P, Irvin J, Bagul A, Ding D, Duan T, Mehta H, et al. Mura: Large dataset for abnormality detection in musculoskeletal radiographs. *Conference on Medical Imaging with Deep Learning*, 2018; Amsterdam, Netherlands. 2018.

57. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*, 2015; Sand Diego, CA. 2015.

58. Bewes J, Low A, Morphett A, Pate FD, Henneberg M. Artificial intelligence for sex determination of skeletal remains: application of a deep learning artificial neural network to human skulls. *J Forensic Leg Med*. 2019;62:40-43.

59. Peña-Solórzano CA, Albrecht DW, Harris PC, Bassed RB, Gillam J, Dimmock MR. Semi-supervised labelling of the femur in a whole-body post-mortem CT database using deep learning. *Comput Biol Med*. 2020:103797.

60. Pozzi A, Raffone C, Belcastro MG, Camilleri-Carter T-L. Sex estimation in cranial remains: A comparison of machine learning and discriminant analysis in Italian populations. *bioRxiv*; 2020.

61. Pilloud MA, Adams DM, Hefner JT. Observer error and its impact on ancestry estimation using dental morphology. *Int J Legal Med*. 2019;133(3):949-962.

62. Carew RM, Errickson D. An Overview of 3D Printing in Forensic Science: The Tangible Third-Dimension. *J Forensic Sci*. 2020;65(5):1752-1760.

63. Carew RM, Morgan RM, Rando C. A preliminary investigation into the accuracy of 3D modeling and 3D printing in forensic anthropology evidence reconstruction. *J Forensic Sci*. 2019;64(2):342-352.

64. Carew RM, Morgan RM, Rando C. Experimental assessment of the surface quality of 3D printed bones. *Australian J Forens Sci*. 2020:1-18.

65. Carew RM, Viner MD, Conlogue G, Márquez-Grant N, Beckett S. Accuracy of computed radiography in osteometry: a comparison of digital imaging techniques and the effect of magnification. *J. Forensic Radiol Imaging.* 2019;19:100348.

66. Robles M, Carew RM, Morgan RM, Rando C. A step-by-step method for producing 3D crania models from CT data. *Forensic Imaging*. 2020;23:200404.