ORIGINAL PAPER

Anthropology

JOURNAL OF
FORENSIC SCIENCES · AAFS

Evaluating expertise in forensic anthropology

Nicholas V. Passalacqua PhD[1] | **Natalie R. Langley PhD**[2] | **Marin A. Pilloud PhD**[3] |
MariaTeresa A. Tersigni-Tarrant PhD[4,5]

[1]Anthropology and Sociology Department, Western Carolina University, Cullowhee, North Carolina, USA

[2]Mayo Clinic, Scottsdale, Arizona, USA

[3]Anthropology Department, University of Nevada, Reno, Reno, Nevada, USA

[4]Saint Louis University, Saint Louis, Missouri, USA

[5]SNA International in Support of the Defense POW/MIA Accounting Agency, Hickam Airforce Base, Hawaii, USA

Correspondence
Nicholas V. Passalacqua, Anthropology and Sociology Department, Western Carolina University, Cullowhe, NC, USA.
Email: passala5@gmail.com

Abstract

The medicolegal system relies on the ability of experts and non-experts alike to make judgments about expertise and use those judgments to reach consequential decisions. Given the lack of standard criteria, mandatory certification, or licensure for establishing expertise required to practice forensic anthropology and testify as an expert witness, we sought to understand how individuals assess and identify expertise in forensic anthropology by using a social science tool called the Imitation Game. This tool assesses immersion in a specific area of study via discourse, with the premise that some individuals lacking expertise themselves imitate or attempt to pass as experts. For this project we recruited volunteers with varying expertise in forensic anthropology to participate in interviews which asked questions about the practice and structure of the discipline. Those interviews were transcribed, anonymized, and evaluated by other recruited individuals with varying expertise in forensic anthropology. Results found that judges who were experts in forensic anthropology performed better than non-expert judges in determining who was *not* an expert in forensic anthropology based on their anonymized responses; however, nearly half of the non-experts were still able to pass as experts in forensic anthropology. The difficulties in assessing expertise based on discourse interactions demonstrates the value and need for well-defined credentials and mandatory certification to practice forensic anthropology. This study demonstrates that accurately identifying expertise in forensic anthropology may be challenging for both experts and non-experts, especially when relying solely on interactional expertise rather than formal assessments of competency which directly elucidate contributory expertise.

KEYWORDS
competency, education, expertise, forensic anthropology, knowledge, professional identity, professionalism, training

Highlights

- The Imitation Game was used to evaluate how individuals assess forensic anthropology expertise.
- Experts and non-experts alike struggled to assess forensic anthropology expertise accurately.
- The use of jargon and length of responses were considered informative indicators of expertise.

- Expert judges put greater weight on references to theory in their assessments than non-experts.
- This study highlights the importance of certification and credentials to demonstrate expertise.

1 | INTRODUCTION

Assessing or measuring expertise is not always a straightforward task. This is especially true in professions where experts make multiple complex decisions in response to various stimuli, the outcomes of which may not have obviously correct or incorrect solutions [1, 2]. However, the medicolegal system relies on the ability of experts and non-experts alike to make judgments about expertise and use those judgments to reach decisions of great consequence. Hence the term "expert witness." So, what is expertise, what qualifies someone as an expert, and how do practitioners, medicolegal experts, and the public assess expertise? This study investigates the relevance and impact of these questions to forensic anthropology, specifically, and the medicolegal system more broadly.

1.1 | What is expertise?

Expertise is knowledge about a particular topic or area (i.e., domain) and the relevant behaviors for its practice, including conduct concerning application, as well as behavioral norms within a domain. Ericsson et al. [3] propose three tests to assess domain expertise: "First, it must lead to performance that is consistently superior to that of the expert's peers [meaning non- or less competent experts]. Second, real expertise produces concrete results.... Finally, true expertise can be replicated and measured in the lab." Importantly, expertise is not simply an accumulation of knowledge, facts, or skills within a domain, but also the ability to organize and apply that knowledge, often without conscious effort [4]. Collins [5: pg. 68] argues that: "To become an expert in some domain is a matter of becoming embedded in the social life of the domain, acquiring what is to a large extent, tacit knowledge, so as to internalize the associated concepts and skillful actions to the point of fluency." Domain-specific expertise does not generalize to other domains, activities altered from a given domain, or to complex problems broadly [e.g., 6–8]. This is true even for domains which appear similar, such as different games which use the same game board [9]. For example, within the broad domain of anthropology, expertise in a specialty such as evolutionary anthropology, does not make someone an expert in a related domain like bioarcheology or forensic anthropology.

Philosophers, psychologists, and sociologists approach expertise from different vantage points. Our study is primarily informed by the sociological literature, but we will attempt to summarize each discipline's approach to understanding expertise. Philosophers focus on the nature of expertise (i.e., the phenomenological analysis of expertise) and how expertise functions within various social conditions [e.g., 10–13]. The philosophical literature discusses the development of expertise, embodiment of expertise by experts, and the varying roles experts and expertise play in societies. For example, Selinger and Crease [12] discuss the differences between an "expert in x" versus an "expert x," with the example of a Secretary of Agriculture who might be an *expert in farming* who also lacks the expertise to actually operate a farm, while an individual competently operating a farm is an *expert farmer*.

Psychologists are concerned primarily with how exceptional performers develop expertise. "Superior expert performance," as opposed to minimum expertise to participate in a domain, reflects national or international recognition and takes years to develop [14]. Ericsson and Towne [2: pg. 408] note: "No experts began their training as superior performers, nor did they spontaneously obtain high levels of performance. Only after years of gradual improvement do they obtain elite status." In fact, Malcom Gladwell's pseudoscientific "10,000-h rule" [15] is a misinterpretation of Ericsson et al. [16] because Gladwell emphasized *time spent* developing expertise, rather than *deliberate practice* as a method of developing expertise [14, 17]. Deliberate practice is "engagement in highly structured activities that are created specifically to improve performance in a domain through immediate feedback, that require a high level of concentration, and that are not inherently enjoyable" [17: pg. 35].

Much of the sociological literature focuses on modeling expertise broadly and understanding how expertise is used for decision making. This body of research falls under Studies of Expertise and Experience (SEE), which differs from other approaches in that it does not consider expertise as mono-focal or scalar. In other words, SEE acknowledges an individual may have various *expertises* (e.g., language, music, math, driving, and athletics) which may be unrelated, and expertise that is ubiquitous in one context may be valuable in another (e.g., Vietnamese language fluency in Vietnam vs. Vietnamese language fluency in the United States) [5]. SEE also recognize that the public has increasing access to all forms of disciplinary knowledge; however, this access to knowledge does not mean the public should be considered experts [5, 18]. The philosophical literature refers to these concepts as "knowing that" versus "knowing how" [see 19].

The sociological literature on expertise largely stems from the work of Collins and Evans [20], whose "periodic table of expertise" categorizes various forms of expertise used to make judgments in a variety of contexts. This periodic table of expertise includes "specialist expertise" which represents the competent practice of professional domains (e.g., academic disciplines, medical professions; Table 1). Collins and Evans' [20] model for specialist expertise begins with ubiquitous tacit knowledge about a domain (i.e., omnipresent or highly accessible knowledge), and progresses towards competence as an active contributing member of a domain. Collins [5: pg. 72] notes simply being familiar with primary domain sources is not adequate: "there are published papers

TABLE 1 Specialist expertises based on Collins and Evans [20].

Specialist expertises	Ubiquitous tacit knowledge			Specialist tacit knowledge	
	Knowledge that is easily accessible (i.e., ubiquitous)			Exclusive knowledge that must be acquired via interactions and enculturation with practicing professionals	
	Beer mat knowledge	Popular understanding	Primary source knowledge	Interactional expertise	Contributory expertise
	Knowledge of very superficial facts about a topic such that you might find on a beer mat/coaster	Knowledge based on popular non-fiction books and the general media	Knowledge based on engaging with the primary literature.	Having enough expertise about a discipline to interact with contributory experts performing their work but lacking the technical knowledge to perform it.	Having enough expertise to contribute to a discipline through technical and scholarly practice

TABLE 2 Acquisition of expertise (based on Collins and Evans [20]: 24–25).

Stage of expertise	Description
Stage 1—Novice	Skills are performed "mechanically" with targeted or specific decisions or reactions. The individual follows rules strictly and does not consider the nuances of variable contexts or conditions, or the whole of the situation.
Stage 2—Advanced beginner	Skills are performed more fluidly, often in response to unexplained but specific features or conditions of the environment. However, there is still a lack of appreciation for the full context of the environment and all decisions (both conscious and unconscious) are still in response to specific stimuli.
Stage 3—Competence	Skills are performed more intuitively than intentionally. The individual is beginning to appreciate the whole context of the environment with more unconscious decision making.
Stage 4—Proficiency	Skills are performed with a holistic recognition of the environment; however, elements of conscious choice and intentional analysis still guide decisions.
Stage 5—Expert	Skills are performed fluidly in an unselfconscious manner based on holistic environmental cues which may be impossible to articulate by the practitioner.

that are indistinguishable from all the others in a journal yet which the professionals in the domain simply ignore. The public have no chance of understanding this, and that is one reason why obtaining knowledge from the internet without further back-up is unreliable." Additionally, Collins and Evans [20] point out that familiarity with the primary literature may provide a false sense of technical mastery which cannot be achieved from familiarity with literature alone.

In contrast to ubiquitous tacit knowledge, specialist tacit knowledge is exclusive domain knowledge acquired only through acculturation with practicing experts [20]. Specialist tacit knowledge is subdivided into interactional expertise and contributory expertise (see Table 1). Interactional expertise is the ability to interact meaningfully with contributory experts performing their work, but lacking the technical knowledge to perform it oneself (i.e., "expert in x") [21]. Contributory expertise represents sufficient expertise to contribute to a discipline through technical and scholarly practice (i.e., "expert x") [22]. Contributory expertise is "close to the common sense meaning of 'expert'" [23: pg. 520], and Collins and Evans [18: pg. 255] note that interactional expertise may not be required to develop contributory expertise. The progression of the acquisition of expertise is further outlined in Table 2 as described by Collins and Evans [20].

1.2 | What qualifies someone as an expert?

Forensic experts are practitioners who may be called upon to testify as expert witnesses and provide an opinion about evidence or events based on their specialized knowledge or expertise in a domain. Criteria used to establish expertise include education, training, experience, and credentials (e.g., certification and licensure). In a court of law, the judge may reference Daubert guidelines [24] to determine if testimony given by an expert witness is based on scientifically valid reasoning, theories, or techniques. Usually, qualifying as a forensic anthropologist expert witness has more to do with education, training, credentials, and casework experience rather than a focus on the theories and techniques used. Certification in forensic anthropology (e.g., American Board of Forensic Anthropology, Forensic Anthropology Society of Europe, and Asociación Latinoamericano de Antropología Forense) may also be used as a credential to demonstrate expertise, but no standard set of criteria or licensure in forensic anthropology currently allows the medicolegal community to determine who is and who is not qualified to undertake forensic anthropology casework and testify as an expert witness.

1.3 | How is expertise assessed?

Contributory expertise may best be assessed via competency or proficiency testing with questions designed to assess competence and differentiate interactional and contributory expertise. Accredited forensic anthropology laboratories typically administer proficiency tests to demonstrate their staff can perform their job duties competently, however these are also not standardized. Additionally, certification demonstrates expertise at the level of competency to practice.

A useful method for assessing interactional expertise is *The Imitation Game* [25]. This tool assesses immersion in a domain via discourse, with the premise that some individuals lacking expertise themselves imitate or attempt to pass as experts. Collins et al. [25: pg. 660] explain that the game "compare[s] the domain specific linguistic abilities of interactional experts with those of contributory experts and those of non-experts." Different forms of the game exist and vary in structure, approach, and participant roles [23]. However, typical roles are:

> Interrogator – individuals who generate/ask questions.
> Pretender – individuals lacking interactional expertise in a particular domain (also responder).
> Non-Pretender – individuals possessing interactional expertise in a particular domain (also responder).
> Judge – individuals who assess responses.

Imitation Game studies [e.g., 25] have found that judges typically base their assessments of expertise on technical content and style (e.g., length, tone, and use of jargon) of the answers.

We emulated an Imitation Game to determine how individuals assess and identify expertise in forensic anthropology. Given the lack of standard criteria, mandatory certification, or licensure for establishing the expertise required to practice forensic anthropology and testify as an expert witness, we wondered what factors experts, other medicolegal professionals, and laypersons alike use to identify an expert. We (as the interrogators) developed a set of questions designed to distinguish interactional expertise and recruited volunteer interviewees (pretenders and non-pretenders) and judges with varying degrees of experience, expertise, and exposure to forensic anthropology. The authors conducted the interviews, created anonymous transcripts for the judges to score, and assigned a level of expertise (i.e., expert or non-expert) to interviewees and judges based on pre-defined criteria. These criteria included terminal degree, experience writing case reports, and certification in forensic anthropology. We hypothesized that individuals identified as experts could assess interactional expertise more accurately than non-experts and detect true experts (non-pretenders) from non-experts (pretenders). We anticipated the results would highlight the need for objective criteria to qualify practitioners as experts, as many individuals who must assess expertise in forensic anthropology are not domain experts.

2 | Materials

The authors initially developed a set of 13 questions about the discipline and practice of forensic anthropology. Questions were structured such that an interviewee may construct answers in various ways depending on their knowledge or familiarity with forensic anthropology; in other words, questions did not have "yes" or "no," or obviously correct or incorrect answers. The questions were beta-tested on several volunteers with varying degrees of forensic anthropology knowledge, training, and experience. These individuals also provided feedback about the questions, noting if questions were confusing, poorly worded, or otherwise problematic. The authors integrated this feedback into the final set of 11 questions (Table 3).

The authors also developed a set of "background information" questions to gather data about education, training, certification, experience, and continuing education in forensic anthropology. This information was used by the authors to assign an expertise level for each study participant (respondents and judges). These questions also asked participants to self-assess their level of expertise using a five-point scale. These expertise self-assessments were compared to the author-assigned expertise levels of all study participants (based on training, experience in the field, and certification), as well as the expertise assigned to the respondents by the judges. The five-point scale was also modified into a dichotomous variable where scores of 1 or 2 were coded as "expert" and scores of 3–5 were coded as "non-expert" (Table 4).

All participants (i.e., judges and responders) were recruited by distributing a request for participation to biological and forensic anthropology listservs and members of various professional organizations. Participation was open to any individual aged 18 years or older, and participants were notified that their participation would remain confidential and anonymous regardless of their participant role. Interested individuals directly emailed the project Principal Investigators (PIs) and notified them of their preferred role (respondent, judge, or no preference). Prior to recruitment, this project was reviewed and approved by the Internal Review Boards of all authors' institutions (Mayo Clinic; Saint Louis University; University of Nevada, Reno; and Western Carolina University).

2.1 | Respondents

A total of 27 participants with varying degrees of expertise in forensic anthropology volunteered to be respondents. Prior to being asked any questions, respondents were read an informed consent and reassured their identities would be kept confidential. Interviewers explained that the goal of the project was to examine how individuals with varying backgrounds in forensic anthropology would answer these questions. All questions were asked in the same order to all respondents, and respondents were allowed to ask for clarification about questions as needed. All interviews took place over the communication and collaboration platform Zoom, which was also used

TABLE 3 Questions used to assess interactional expertise in forensic anthropology.

Question order	Question
1.	Please define forensic anthropology.
2.	What is a taphonomic agent? Please provide an example.
3.	What does it mean to determine forensic significance?
4.	If you were to examine remains known to be a soldier who died during the Korean War what method(s) would you use to estimate age? Why?
5.	When assessing the skeleton of a sub-adult, which part(s) of the biological profile cannot be ascertained with reasonable certainty? Why?
6.	You are presented with a modern forensic case, consisting of complete skeletal remains from an adult, how would you estimate sex?
7.	What is FORDISC and how does it work?
8.	What organization(s) are currently working to develop standards within forensic anthropology?
9.	While taking the standard cranial measurements, you notice that the adult cranium you are measuring is completely edentulous with significant maxillary alveolar resorption. Which measurements would be affected?
10.	You observe a defect in the right parietal bone and ascertain it is likely a gunshot wound. What feature(s) do you use to determine if the defect is an entrance or exit wound?
11.	What is the difference between analytical/bench notes and reports?

TABLE 4 Ordinal and dichotomous expertise scales.

Score	Description given to raters in survey	Rating used by authors	Dichotomous scale
1	Extensive knowledge about forensic anthropology	Individual who is board certified and has a PhD	Expert
2	Considerable knowledge about forensic anthropology	Individual is a PhD student with significant experience or has a PhD with focus on forensic anthropologist and is not board certified	Expert
3	Moderate knowledge of forensic anthropology	Master's student, bioarcheologist with little forensic anthropology experience, forensic pathologist	Non-expert
4	Little knowledge about forensic anthropology	Undergraduate level	Non-expert
5	Minimal knowledge about forensic anthropology	No knowledge of forensic anthropology	Non-expert

to record the interviews. Zoom's automated transcription feature was used to create a draft transcript, which was then edited by the PI who conducted the interview, using the recording to ensure the transcript was correct. All potentially identifying information about respondents was removed, and recordings were destroyed after redacted transcripts were prepared.

2.2 | Judges

A total of 11 judges with varying degrees of expertise in forensic anthropology volunteered to review the written anonymized responses (judges did not have access to any video or audio recordings of the interviews). Because of the time required to perform the evaluations, each judge was assigned ~15 respondent transcripts to evaluate. Judges were assigned transcripts randomly; however, the same five respondent transcripts were assigned as part of the total transcripts to all evaluators to examine evaluator agreement. These five universally assigned transcripts included expert and non-expert respondents.

Judges used the five-point scale ranging from extensive to minimal knowledge to assess each of the 11 answers individually from each respondent (see Table 4). Judges then provided an overall assessment of perceived expertise in forensic anthropology for the respondent. Judges were also asked to identify cues or information they used to reach their assessments for each question, as well as for their overall assessment of respondent expertise using the following prompt:

What information/variables did you use to arrive at this assessment? Check all that apply

1. *Discipline-specific terminology*
2. *Length of answers*
3. *Span of knowledge across topics*

TABLE 5 Evaluation performance—ordinal.

	n	Frequency correct	Percent correct	Frequency incorrect	Percent incorrect	Pass rate
Total	155	59	38%	96	62%	120%
Expert judges	72	34	47%	38	53%	110%
Non-expert judges	83	25	30%	58	70%	140%

TABLE 6 Evaluation performance—dichotomous (expert vs. non-expert).

	n	Frequency correct	Percent correct	Frequency incorrect	Percent incorrect	Pass rate
Total	155	115	74%	40	26%	52%
Expert judges	72	57	79%	15	21%	42%
Non-expert judges	83	58	70%	25	30%	60%
					Sensitivity	91%
					Specificity	54%
Positive predictive value						81%
Negative predictive value						73%

4. *Knowledge of the current state of practice – familiarity with current methods and advanced research within the discipline*

5. *Historical perspective of the discipline*

6. *Knowledge of theory*

7. *Knowledge of professional practice of forensic anthropology and relevant laws/guidance governing practice*

8. *Use of discipline-specific example(s)*

9. *Other: _____*

After all evaluations were completed, judges completed a survey about their experience performing the respondent transcript evaluations.

3 | METHODS

Judge performance was examined using percent correct, percent incorrect, and pass rate. Percent correct corresponds to a judge's ability to correctly determine a respondent's author-assigned expertise in forensic anthropology; percent incorrect corresponds to an incorrect assessment. Pass rate reflects the percent of non-expert respondents (i.e., pretenders) passing as experts based on judge assessments [e.g., 26]. The pass rate is calculated as: 1- (right guesses—wrong guesses/total guesses). A 0% pass rate means a judge was correct in all their assessments (no pretenders passed as experts); a 100% pass rate means a judge was only correct in half of their assessments, and a 200% pass rate means a judge was incorrect in all their assessments. All statistical analyses were done in the statistical computing environment R [27]. Consistency between judge ratings was calculated using a weighted Cohen's kappa. Polychoric and tetrachoric correlations, and Cohen's kappa were used to examine the relationship between judge assessments of expertise and author-assigned expertise. The irr package was used to calculate Cohen's

kappa [28]. The polychoric correlation was calculated in the polycor package [29]. The tetrachoric correlation was calculated in the psych package [30].

4 | RESULTS

Results showed little to no agreement between respondents' actual expertise (as assigned by the authors) and judges' assessment of expertise when expertise was coded as a five-point ordinal score (polychoric correlation=−0.027; Cohen's kappa=0.32). With these ordinal data, neither group of judges (experts or non-experts) performed above 50% correct in assessing expertise (overall 38% correct); however, judges who were experts in forensic anthropology did outperform non-experts in their evaluations (Table 5); thus, our hypothesis was supported.

Correlations were moderate when expertise was simplified into a dichotomous variable of expert versus non-expert (tetrachoric correlation=0.45; Cohen's kappa=0.49). Additionally, a moderate correlation was found between scores for each answer and overall expertise rating. All judges performed better when expertise was coded dichotomously, with an overall percent correct of 74%; judges who were experts in forensic anthropology again performed better than non-experts in their evaluations (Table 6). The pass rate for non-experts passing as experts was 52% overall. Judges who were experts in forensic anthropology performed better (42% pass rate) than non-expert judges (60% pass rate) in determining who was *not* an expert in forensic anthropology; however, data were variable, and nearly half of the non-experts were able to pass as experts. Table 7 provides individual judge data for dichotomous expertise evaluations. These findings suggest interactional expertise in forensic anthropology is difficult to assess consistently and may not be the best indicator of true domain expertise.

TABLE 7 Individual judge performance.

ID	n	Frequency correct	Percent correct	Frequency incorrect	Percent incorrect	Pass rate
1 (expert)	14	10	71%	4	29%	57%
2 (expert)	14	11	79%	3	21%	43%
3 (expert)	14	10	71%	4	29%	57%
4 (expert)	14	9	64%	5	36%	71%
5 (expert)	14	10	71%	4	29%	57%
6 (expert)	13	8	62%	5	38%	77%
7 (non-expert)	15	12	80%	3	20%	40%
8 (non-expert)	15	13	87%	2	13%	27%
9 (non-expert)	15	13	87%	2	13%	27%
10 (non-expert)	12	6	50%	6	50%	100%
11 (non-expert)	15	13	87%	2	13%	27%

TABLE 8 Variables the judges used to arrive at their assessment of the respondents answer to question #7. See text for description of variable numbers.

	1	2	3	4	5	6	7	8	9
Expert %	20.74	22.58	3.69	23.96	3.69	11.52	4.15	2.76	6.91
Non-expert %	25.95	23.78	5.95	19.46	3.24	2.70	8.11	3.78	7.03
All judges	23.13	23.13	4.73	21.89	3.48	7.46	5.97	3.23	6.97

The questions most highly correlated with author-rated expertise focused on methods (e.g., FORDISC—question #7 and subadult age estimation—question #5) and standards development within forensic anthropology (question #8). Inter-rater agreement was variable, ranging from concordance smaller than chance to a weighted kappa of 0.62, and was unrelated to judge expertise. Factors most often reported by judges used to assess respondent expertise included the use of discipline-specific jargon and the overall length of the answers. In an analysis of just question #7 which was highly correlated with overall expertise rating by the judges, the factors being weighted by the judges were separated by expert and non-expert judges. Each judge could select multiple answers; therefore, percentages were calculated as a percentage of the total variables indicated by each judge. Both the expert and the non-expert judges weighted terminology use (#1), length of answers (#2), and knowledge of the current state of discipline (#4) as the most important factors in assessing the answers of the respondents. The experts weighted knowledge of theory (#6) as much higher than the non-expert judges (Table 8).

5 | DISCUSSION

The most important conclusion of the current study is that accurately identifying expertise in forensic anthropology may be challenging for both experts and non-experts, especially when relying solely on interactional expertise rather than formal assessments of competency which directly elucidate contributory expertise and are typically associated with credentials (e.g.,

certification by the ABFA). Overall, judges who were experts in forensic anthropology outperformed non-expert judges in detecting non-expert respondents. However, a considerable number of non-experts passed as experts, regardless of the expertise level of the judge. Collectively, the evidence suggests that criminal justice and medicolegal death investigation professionals need reliable, non-subjective criteria to assess forensic anthropology expertise accurately and consistently.

Gatekeeping of knowledge and access to practice a domain provides a means of differentiating experts. Within professional domains, gatekeeping by educators and professional bodies ensures graduates and practitioners are fit to practice and screens out unqualified individuals who may cause harm to clients and stakeholders [31, 32]. Sowbel [32] argues that gatekeeping a professional domain "is a fundamental ethical obligation," but various factors affect successful gatekeeping, including fear of litigation from individuals considered unfit or lacking expertise; vague or ill-defined suitability criteria for characteristics of individuals to perform within a domain; conflicting educator/mentor roles for screening out unqualified individuals; training with unqualified individuals to attain competencies necessary to contribute to the domain; and lack of measures or protocols for evaluating performance/expertise within a domain. Sowbel's [32] definition of professional gatekeeping is different from gatekeeping in which individuals are prevented from participating due to discrimination based on perceived sex, gender, race, religion, socioeconomic status, ability status, among many other self-recognized identities, as well as theoretical perspectives. In contrast, professional gatekeeping is meant to prevent unqualified individuals from practicing and posing as experts.

Many professions use licensure as a means of gatekeeping and ensuring discipline-related work is performed competently and ethically by qualified, vetted experts (e.g., physicians, nurses, dentists, plumbers, electricians, aestheticians, etc.). Licensed professionals who perform work outside the scope of their license may lose their license to practice and/or face legal repercussions. An added benefit of requiring certification or licensure to practice is that re-certifying or renewing a license usually requires continuing education, ensuring experts keep current on the latest developments and best practices within their domain. Research has shown that time spent practicing a domain often is unrelated to improvements in performance, and time since graduation, which is highly correlated with experience, is associated with *decreasing* performance [2: pg. 405 and references therein]. Additionally, no association exists between amount of experience and performance, and perceived expertise and performance [33], possibly reflecting the Dunning–Kruger Effect, or cognitive bias where people incorrectly overestimate their knowledge, abilities, and/or expertise in a domain [34]. Importantly, commitment to maintaining and/or increasing one's expertise through deliberate practice appears to be associated with maintaining or increasing one's expertise, or rebounding from periods of inactivity [14]. The difficulties and inconsistencies in assessing expertise revealed by our study underscore the importance of professional gatekeeping mechanisms (e.g., licensure and certification) to demonstrate expertise for the purposes of forensic anthropological casework and expert witness testimony. This is especially important as forensic anthropology has diverged from biological anthropology broadly and bioarcheology specifically, now functioning as a unique discipline with specialized methods relevant to medicolegal applications [35, e.g., 36–39].

Forensic anthropology has made considerable progress in professionalization and validation over the last 15 years, largely in response to the 2009 National Research Council report which underscored issues with a lack of defined qualifications for many forensic disciplines and practitioners' inability to understand or explain methods used routinely in their work [40]. This progress includes the ongoing development of consensus-based standards for forensic anthropology practice, including forthcoming standards for education, training, and qualifications [41]. Consensus exists for implementing accreditation for graduate education and training programs [42], but standardization in these areas is lacking. Defining, implementing, and assessing core competencies would provide a framework for standardizing education, training, and certification and lay the foundation for establishing forensic anthropology expertise [43]. The American Board of Forensic Anthropology is undertaking this effort as part of implementing a multilevel certification system for forensic anthropologists. Multiple levels of certification will ensure credentials, training, experience, and expertise align with certification. However, though it is best practice to demonstrate disciplinary expertise via certification by an accredited certifying body [35, 44, 45], certification is not required to practice forensic anthropology. This leaves stakeholders with the burden of identifying experts from a pool of practitioners who may or may not have the requisite expertise.

This study has a couple of limitations. First, the sample size is relatively small; however, this is common in studies requiring considerable time from volunteer research subjects. Second, the authors recruited participants with varying degrees of forensic anthropology knowledge and experience, but we did not include subjects with no knowledge of forensic anthropology (due to limitations in recruitment strategies). Presumably these individuals would have been less adept at assessing expertise than participants with domain knowledge, but such data would be of value, as members of a jury may fall into this category. Though not a study limitation, one unexpected outcome was that research participants' self-reported expertise, frequently did not correspond to their knowledge or experience. For example, two ABFA-certified individuals working as full-time forensic anthropologists rated themselves as "2—Has considerable knowledge about forensic anthropology" rather than "1—Has extensive knowledge about forensic anthropology." The potential reasons for biases in self-reporting of expertise are many and may include imposter syndrome for the under-reporting of expertise, and overconfidence (e.g., The Dunning–Kruger effect) for over-reporting of expertise. While this discrepancy may not have affected the dichotomous scale analyses, it would have impacted five-point scale analyses, so we chose to assign expertise to participants based on data in the "background information" questionnaire about education, training, certification, credentials, and experience (see Table 4).

6 | CONCLUSIONS

This study of assessing interactional expertise demonstrated that individuals with expertise in forensic anthropology can distinguish experts from non-experts more reliably than individuals without expertise in forensic anthropology. However, in all cases all judges were incorrect in at least some of their assessments, showing that individuals with some knowledge of forensic anthropology can emulate expertise conversationally. The difficulties in assessing expertise based on discourse interactions demonstrates the value and need for well-defined credentials and mandatory certification to practice forensic anthropology. Forensic anthropology certification is a voluntary activity; therefore, many practitioners remain uncertified. Licensure would ensure medical examiner's offices, law enforcement agencies, legal professionals, and the public that forensic anthropology casework is conducted by qualified experts.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT
No authors do not have any conflicts of interest to report.

REFERENCES

1. Mockus A, Herbsleb JD, editors. Expertise browser: a quantitative approach to identifying expertise. Proceedings of the 24th International Conference on Software Engineering; 2022 May 25; Orlando, FL. Piscataway, NJ: IEEE; 2022.
2. Ericsson KA, Towne TJ. Expertise. WIREs Cognit Sci. 2010;1(3):404–16. https://doi.org/10.1002/wcs.47
3. Ericsson KA, Prietula MJ, Cokely ET. The making of an expert. Harv Bus Rev. 2007;85(7–8):114–21.
4. Bedard J, Chi MT. Expertise. Curr Dir Psychol Sci. 1992;1(4):135–9.
5. Collins H. Studies of expertise and experience. Topoi. 2018;37(1):67–77. https://doi.org/10.1007/s11245-016-9412-1
6. Voss JF, Greene TR, Post TA, Penner BC. Problem-solving skill in the social sciences. In: Bower GH, editor. Psychology of Learning and Motivation. Cambridge, MA: Academic Press; 1983. p. 165–213.
7. Hagemann N. The advantage of being left-handed in interactive sports. Atten Percept Psychophys. 2009;71(7):1641–8. https://doi.org/10.3758/APP.71.7.1641
8. Kirkman MA. Deliberate practice, domain-specific expertise, and implications for surgical education in current climes. J Surg Educ. 2013;70(3):309–17. https://doi.org/10.1016/j.jsurg.2012.11.011
9. Eisenstaedt M, Kareev Y. Aspects of human problem solving: the use of internal representation. In: Noman D, Rumelhart D, editors. Exploration in Cognition. San Francisco, CA: W.H. Freeman; 1975. p. 308–46.
10. Dreyfus H, Dreyfus S. Mind over Machine. New York, NY: Free Press; 1986.
11. Mialet H. Do angels have bodies? Two stories about subjectivity in science: the cases of William X and mister H. Soc Stud Sci. 1999;29(4):551–81.
12. Selinger EM, Crease RP. Dreyfus on expertise: the limits of phenomenological analysis. Cont Philos Rev. 2002;35(3):245–79. https://doi.org/10.1023/A:1022643708111
13. Luntley M. Understanding expertise. J Appl Philos. 2009;26(4):356–70. https://doi.org/10.1111/j.1468-5930.2009.00468.x
14. Ericsson KA, Harwell KW. Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: why the original definition matters and recommendations for future research. Front Psychol. 2019;10:2396. https://doi.org/10.3389/fpsyg.2019.02396
15. Gladwell M. Outliers: The Story Of Success. New York, NY: Little, Brown, and Co; 2008.
16. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev. 1993;100(3):363–406. https://doi.org/10.1037/0033-295X.100.3.363
17. Hambrick DZ, Oswald FL, Altmann EM, Meinz EJ, Gobet F, Campitelli G. Deliberate practice: is that all it takes to become an expert? Dermatol Int. 2014;45:34–45. https://doi.org/10.1016/j.intell.2013.04.001
18. Collins HM, Evans R. The third wave of science studies: studies of expertise and experience. Soc Stud Sci. 2002;32(2):235–96. https://doi.org/10.1177/0306312702032002003
19. Dreyfus HL, Dreyfus SE. What is moral maturity? A phenomenological account of the development of ethical expertise. In: Rasmussen D, editor. Universalism vs. Communitarianism: Contemporary Debates in Ethics. Cambridge, MA: MIT Press; 1990. p. 237–64.
20. Collins H, Evans R. Rethinking Expertise. Chicago, IL: University of Chicago Press; 2007.
21. Collins H, Evans R. Expertise revisited, part I—interactional expertise. Stud Hist Philos Sci Part A. 2015;54:113–23. https://doi.org/10.1016/j.shpsa.2015.07.004
22. Collins H, Evans R, Weinel M. Expertise revisited, part II: contributory expertise. Stud Hist Philos Sci Part A. 2016;56:103–10. https://doi.org/10.1016/j.shpsa.2015.07.003
23. Collins H, Evans R, Weinel M, Lyttleton-Smith J, Bartlett A, Hall M. The imitation game and the nature of mixed methods. J Mix Methods Res. 2017;11(4):510–27. https://doi.org/10.1016/j.shpsa.2015.07.003
24. Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579; 1993.
25. Collins H, Evans R, Ribeiro R, Hall M. Experiments with interactional expertise. Stud Hist Philos Sci Part A. 2006;37(4):656–74. https://doi.org/10.1016/j.shpsa.2006.09.005
26. Collins H, Evans R, Hall M, O'Mahoney H, Weinel M. Bonfire night and burns night: using the imitation game to research English and Scottish identities. In: Caudill D, Conley S, Me G, Weinel M, editors. The Third Wave in Science and Technology Studies. Cham: Palgrave Macmillan; 2019. p. 109–31.
27. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021 http://www.R-project.org. Accessed 11 Apr 2023
28. Gamer M, Lemon J, Fellows I, Singh P. irr: various coefficents of interrater reliability and agreement. R package version 0.84; 2012. https://CRAN.R-project.org/package=irr. Accessed 11 Apr 2023
29. Fox J, Dusa A. Polychoric and polyserial correlations. R package version 0.8–1; 2022. https://cran.r-project.org/web/packages/polycor/polycor.pdf. Accessed 11 Apr 2023
30. Revelle W. psych: procedures for personality and psychological research. Version 1.8.4; 2018. https://CRAN.R-project.org/package=psych. Accessed 11 Apr 2023
31. Moore LS, Urwin CA. Gatekeeping: a model for screening baccalaureate students for field education. J Soc Work Educ. 1991;27(1):8–17. https://doi.org/10.1080/10437797.1991.10672164
32. Sowbel LR. Gatekeeping: why shouldn't we be ambivalent? J Soc Work Educ. 2012;48(1):27–44. https://doi.org/10.5175/JSWE.2012.201000027
33. Ericsson KA, Whyte J IV, Ward P. Expert performance in nursing: reviewing research on expertise in nursing within the framework of the expert-performance approach. Adv Nurs Sci. 2007;30(1):E58–71. https://doi.org/10.1097/00012272-200701000-00014
34. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol. 1999;77(6):1121–34. https://doi.org/10.1037/0022-3514.77.6.1121
35. Passalacqua NV, Pilloud MA, Congram D. Forensic anthropology as a discipline. Biology. 2021;10(8):691–708. https://doi.org/10.3390/biology10080691
36. İşcan MY. Rise of forensic anthropology. Yearb Phys Anthropol. 1988;31:203–30. https://doi.org/10.1002/ajpa.1330310510
37. Krogman WM. A guide to the identification of human skeletal material. Federal Bureau Invest Law Enforce Bull. 1939;8(8):3–31.
38. Dirkmaat DC, Cabo LL, Ousley SD, Symes SA. New perspectives in forensic anthropology. Yearb Phys Anthropol. 2008;51:33–52. https://doi.org/10.1002/ajpa.20948
39. Ubelaker D. The historical development of forensic anthropology. Forensic Anthropol. 2020;3(3):164–70. https://doi.org/10.5744/fa.2020.1022
40. National Research Council. Strengthening Forensic Science in the United States: A Path Forward. Washington, DC: National Academy Press; 2009.
41. Passalacqua NV, Pilloud MA. Ethics and Professionalism in Forensic Anthropology. San Diego, CA: Academic Press; 2018.
42. Passalacqua NV, Pilloud MA. Education and training in forensic anthropology. Forensic Anthropology. 2020;3(3):65–74. https://doi.org/10.5744/fa.2020.1013
43. Langley N, Tersigni-Tarrant M. Core competencies in forensic anthropology: a framework for education, training, and practice.

Forensic Anthropology. 2020;3(2):75–80. https://doi.org/10.5744/fa.2020.1014

44. Pilloud MA, Passalacqua NV, Philbin CS. Caseloads in forensic anthropology. Am J Biol Anthropol. 2022;177(3):556–65. https://doi.org/10.1002/ajpa.24471

45. Passalacqua NV, Pilloud MA. The need to professionalize forensic anthropology. Eur J Anat. 2021;25(S2):35–47.

How to cite this article: Passalacqua NV, Langley NR, Pilloud MA, Tersigni-Tarrant MA. Evaluating expertise in forensic anthropology. J Forensic Sci. 2023;00:1–10. https://doi.org/10.1111/1556-4029.15262