Improving the Relationship Between Confidence and Competence: Implications for Diagnostic Radiology Training From the Psychology and Medical Literature

Oren Jaspan, MD, Anna Wysocka, BS, Carmen Sanchez, Ph.D, Andrew D. Schweitzer, MD

The focus of diagnostic radiology training is on creating competent professionals, whereas confidence and its calibration receive less attention. Appropriate confidence is critical for patient care both during and after training. Overconfidence can adversely affect patient care and underconfidence can create excessive costs. We reviewed the psychology and medical literature pertaining to confidence and competence to collect insights and best practices from the psychology and medical literature on confidence and apply them to radiology training.

People are rarely accurate in assessments of their own competence. Among physicians, the correlation between perceived abilities and external assessments of those abilities is weak. Overconfidence is more prevalent than underconfidence, particularly at lower levels of competence. On the individual level, confidence can be calibrated to a more appropriate level through efforts to increase competence, including sub-specialization, and by gaining a better understanding of metacognitive processes. With feedback, high-fidelity simulation has the potential to improve both competence and metacognition. On the system level, systems that facilitate access to follow-up imaging, pathology, and clinical outcomes can help close the gap between perceived and actual performance.

Appropriate matching of trainee confidence and competence should be a goal of radiology residency and fellowship training to help mitigate the adverse effects of both overconfidence and underconfidence during training and independent practice.

KEY WORDS: Calibration; Confidence; Competence; Education.

INTRODUCTION

The primary goal of post-graduate training in diagnostic radiology is to produce physicians who are competent to perform unsupervised interpretation of diagnostic imaging studies. In order to achieve this, residents must receive gradual incremental increases in responsibility and autonomy over the course of their training. In order for these increases in autonomy to be safe, in addition to appropriate supervision, resident confidence and competence must be appropriately calibrated. Calibration is defined as the appropriate matching of competence and confidence (1,2). A resident whose confidence matches their competence is well-calibrated. Under- and overconfidence in one's own competence can have serious negative consequences both during and after the completion of training.

Both of these states of miscalibration expose individuals to various decision-making biases and behaviors which may foster faulty medical decision-making, diagnostic errors, inappropriate communication of diagnostic uncertainty, and the incurrence of excess healthcare costs. In radiology trainees specifically, miscalibration may contribute to substandard patient care, as well as undermine their development and clinical autonomy. With a pledge to *do no harm* to our patients, and a premium on providing *value-based* healthcare in a responsible and judicious manner, appropriate calibration of individual residents should be an explicit goal of residency training. In order to achieve this, we must first identify individuals who are most susceptible to miscalibration, explore the detrimental impacts of miscalibration on physician performance and project how these effects may manifest in radiology specifically, and apply best practices from the relevant psychology and medical literature to tailor radiology training programs to optimize trainee calibration.

Acad Radiol 2022; 29:428–438

From the Department of Radiology, Weill Cornell Medicine / NewYork-Presbyterian Hospital, New York, New York (O.J., A.W.t.e.S.b.s.W., A.D.S.); University of Illinois at Urbana Champaign, Champaign, Illinois (C.S.). Received October 20, 2020; revised November 23, 2020; accepted December 11, 2020.
Address correspondence to: O.J. e-mail: oren.jaspan@gmail.com

https://doi.org/10.1016/j.acra.2020.12.006

WHAT IS *CALIBRATION* AND WHY SHOULD IT BE PURSUED?

In general, people are not accurate judges of their own competence. Specifically, there is a tendency to estimate one's knowledge, abilities, and positive attributes relative to others as slightly above average. This occurs across all levels of competence and is referred to as the "better-than-average" effect (3).

The largest mismatch between confidence and competence occurs in those with the lowest levels of performance. That is, overconfidence is the most prominent among those with limited competence and/or experience in a particular domain (4,5) and is exacerbated as tasks become more challenging (6).

In this context, *calibration* is defined as the appropriate matching of competence and confidence (7−15). A mismatch in this respect may then be referred to as *miscalibration*. Miscalibration manifests as either over- or under-confidence relative to an individual's true competence. Such miscalibration of confidence and competence is pervasive across many domains including in laboratory experiments assessing humor, written grammar, logical reasoning, and probabilistic learning (4,5) as well as countless real-world applications such as personal financial literacy, stock market trading, negotiation, higher education, aviation, and driving (16−20). Unsurprisingly, such miscalibration of confidence and competence also pervades the medical field, often affecting trainees and expert physicians to different extents (6,7,12−15,21,22). In medicine, as within the other domains examined, miscalibration is associated with certain maladaptive behaviors, discussed subsequently, which hinder physicians' ability to perform autonomously as trainees, and to provide safe, efficacious, and efficient care as their careers progress.

WHO IS MOST SUSCEPTIBLE TO MISCALIBRATION?

In order to improve radiology training programs, it is useful to gain a better understanding of who is most susceptible to miscalibration, and which factors and contexts moderate their degree of miscalibration. There has been substantial research which elucidates the variables which have an association with miscalibration. These may be divided broadly into 2 overarching categories (*Table 1*): *individual* factors such as competence (4), experience (5,7,13,20−22), personality (23), reliance upon various cognitive biases (10,24,25) sex/gender (25−28), culture (27−29), and *context* or *task-specific* factors, such as task difficulty (6,30) outcome predictability (30,31) and feedback (8,9,32,33).

TABLE 1. Factors contributing to miscalibration

Individual factors
Competence
Experience
Metacognition
Cognitive biases and heuristics
Personality
Sex/Gender
Nationality
Context-dependent factors
Feedback (availability, quality, and timing)
Task difficulty
Outcome predictability
Population epidemiology
Educational methods

In terms of performance, on average most people rate their performance as better than average. However, the greatest mismatch between confidence and performance occurs in those in the bottom quartile on knowledge, a phenomenon that has been coined the *Dunning-Kruger* effect (4): people with low competence tend to be the most overconfident. They hypothesize that this mismatch is largely due to a deficit of *metacognition* - an awareness and understanding of one's thought processes which allow one to know how well they are performing, to be able to discern how likely they are to be accurate or inaccurate, and the ability to recognize competence in others to improve their own performance. On the other end of the spectrum, experts tend to be underconfident, and paradoxically become deficient in their ability to recognize the relative incompetence of others, taking their competence for granted (34). These dynamics may account for the findings of a review assessing the accuracy of physician self-assessment compared to observed measures of their competence, where the least-expert physicians were found to be the most overconfident in their self-assessments (7). Analogous research relating radiologist performance and confidence is sparse. However, the Dunning-Kruger effect was replicated in an experiment in the field of radiology which analyzed the performance of radiologists and non-radiologist physicians with regards to their interpretation of a set of chest radiographs, and also measured their confidence levels (14). While there was a wide range in actual performance, there was a more narrow range in confidence levels, with worse performers reporting higher levels of confidence than their more accurate counterparts.

There is vast research demonstrating the robustness of the Dunning-Kruger effect. However, it primarily focuses on the differences in overconfidence between people performing tasks with which they already have varying degrees of familiarity. Sanchez and Dunning (5) expand upon this work by exploring changes in confidence as people actively learn a novel skill that is probabilistic in nature, while receiving feedback about their learning. They demonstrate that rank beginners are not initially overconfident, but after just a few learning experiences surge to a *beginner's bubble* of overconfidence. An experiment exploring the calibration between attending physician, senior resident, and medical student diagnostic accuracy on a set of challenging medical cases and their confidence in their ability to make the correct diagnosis, found that while most of the subjects were overconfident, trainees were the *most* overconfident group, followed by attending physicians (13). Perhaps analogous to Sanchez and Dunning's *rank beginners*, not only were medical students the least confident group, but their confidence and competence levels were the most optimally-calibrated. A pattern reflective of a beginner's bubble has also manifested in other disciplines, medical and otherwise, sometimes with morbid and even fatal consequences. Upon receiving minimal thresholds of experience, new pilots (20), "medium trained" endoscopists (22), and spine surgeons learning a new technique (21) are responsible for a disproportionate number of fatal accidents and procedural errors, respectively. Radiology residents may be particularly susceptible to the beginner's bubble phenomenon, as they likely begin their training with limited interpretive competence and appropriately-calibrated low confidence, but may rapidly develop confidence which outpaces their true skills early in training.

WHY ARE PEOPLE MISCALIBRATED?

Human decision-making and the mechanisms that underlie faulty reasoning and bias have been studied formally for decades. Many of these mechanisms have been proposed to contribute to overconfidence in decision-makers.

Heuristics, Dual-processing, and Bias

One of the foundational studies in this domain was the landmark paper by Tversky and Kahneman in 1974 (35). They describe a set of mental shortcuts, or *heuristics*, that individuals use to facilitate decision-making in probabilistic tasks in the setting of uncertainty. Although many such heuristics allow for rapid and accurate decisions to be made in most circumstances, they, along with various cognitive biases, have been implicated in the formation of overconfidence in general, and in medicine specifically (10,12,36,37). Such biases include motivational biases (3,38,39), self-serving trait definitions (40,41) selective recall of past behavior (42) a tendency to ignore proficiency of peers (43,44) recency bias (11) and confirmation bias (9,12) among others. Many such biases are known to potentiate interpretative error in radiology specifically, the most prevalent and perhaps well-known of which being the satisfaction of search bias (45−47).

Subsequently, various *dual process* models of cognition were developed, which hypothesize that people think and make decisions using two separate thought processes, one automatic and quick, which occurs subconsciously, and one more deliberate and contemplative, which occurs consciously. These processes eventually became known as *System 1* and *System 2*, respectively (48) the nomenclature which reached mainstream notoriety via Kahneman in 2011 (49). It stands to reason that the automatic processing which characterizes *System 1* thinking relies upon heuristics to allow for efficient problem solving. The ways in which such automatic (i.e., *System 1*) processing has been implicated in clinician overconfidence have been explored in the literature. Reliance upon a prevalent set of heuristics (Appendix A) is demonstrated to contribute to overconfidence and diagnostic error, likely in part because of premature closure of the differential diagnosis and under-consideration of additional diagnostic possibilities (10,12,37). There is additional thought that use of *System 1* thinking, in and of itself, may contribute to overconfidence in one's judgment because of its association with positive emotions like excitement and enthusiasm (10,50).

The reliance upon a set of heuristics, or *personal decision rules* (36), as they have been called in the clinical context, is exacerbated in settings of increased uncertainty, even when it may contribute to miscalibration of confidence and competence and occasionally result in suboptimal outcomes (51). This is particularly relevant when considering the implications for miscalibration of clinicians, because uncertainty underpins most medical decision making scenarios to some degree (51−54), marking medical professionals as a particularly susceptible group to miscalibration. Ironically, the medical field has long considered the acknowledgment of this uncertainty as a sign of vulnerability, both implicitly and explicitly, and has fostered a culture in which confidence is valued over the recognition of uncertainty (10,55).

Metacognition and Exuberant Theorizing

The term *metacognition* refers to a conscious understanding of one's thought process. A deficit in metacognitive abilities is proposed as a driving mechanism of miscalibration, particularly in less-competent individuals. This is demonstrated in experiments assessing individuals' ability to recognize competence in themselves and others, as well as their ability to gain insight regarding their own level of competence through social comparison information (4). In each case, less competent individuals are less likely to recognize true competence, and to use social comparison information to improve their own competence. Essentially, if one does not know what they do not know, they lack the ability to recognize that they do not know, to recognize those who know more, and to learn from the better performance of others. As previously discussed, this dynamic has been replicated in a chest radiograph interpretation task (14) as well as within physician self-assessment surveys (7) both of which demonstrate the highest confidence levels among the least competent.

Similarly, in early learners, faulty logic and theory-formation are implicated in the development of miscalibration, such as through the *beginner's bubble* phenomenon (5). This is attributed to the "exuberant and error-filled theorizing" novices perform as they gain early experience with a task, the levels of which correlate with an individual's level of overconfidence. As discussed previously, there are many examples both within and outside of medicine in which individuals with an early-to-intermediate degree of experience with a task are least calibrated (13) and disproportionately responsible for errors, relative to those with either more or less experience with the same task (20−22).

Feedback

Feedback refers to the communication of decision outcomes, and it is an essential component of learning. In the clinical setting, examples of feedback include diagnostic test results, pathology findings, and patient outcomes. The emergency medicine literature illuminates the critical role that effective feedback plays in the calibration of physicians, and many of the infrastructural and logistical obstacles which impede effective feedback in the clinical setting. Such obstacles include time pressure, infrequency of various pathologies, and deficient systems for the reporting of near-misses, errors, and adverse events (8). The relative deficit of effective and timely feedback resulting from these limitations of the clinical environment is implicated in physician miscalibration.

The proposed link between feedback and calibration is through the development of various feedback loops. When a decision results in a positive outcome, a positive feedback loop is engaged which reinforces the antecedent decision-making process and behavior. Conversely, when a negative outcome is realized, a negative feedback loop is initiated which promotes recalibration and behavioral change. Given the barriers to consistent feedback in the clinical setting, however, what happens in the absence of feedback is of particular relevance to clinician calibration. Croskerry (8) proposes that when feedback is missing and outcomes are unknown, the preceding cognitive processes and behaviors are perpetuated - an unknown outcome is processed identically to a positive outcome. Therefore, lack of feedback, and any environmental or cultural obstacles preventing sufficient feedback, contribute to physician miscalibration.

Context

In addition to the wide array of individual factors which contribute to miscalibration, there are many contextual, environmental, or task-related factors which are also implicated in fostering miscalibration in decision-makers. Berner and Graber (12) describe *context errors*, characterized by the early categorization of a clinical problem based on the available clinical history, a patient's previously established diagnoses, or other contextual information. While this approach is sensible, if unchecked it may promote dependence on the *representativeness* heuristic (Appendix A), and lead the physician to seek confirming evidence (confirmation bias) for a potentially incorrect diagnosis, and to prematurely close the differential diagnosis instead of considering other possible diagnoses.

Similarly, Cifu (11) elaborates on numerous *static* and *dynamic* influences on a physician's calibration (Appendix A). Both categories are dominated by effects of the environment and context in which a physician practices. These factors may promote reliance upon various heuristics, which may in turn result in the overvaluing of general epidemiological trends or overestimation of the pretest probability of diagnoses which have recently been encountered.

Another dynamic influence on calibration is the mental state of the physician at the time of decision-making. An experiment in primary care medicine demonstrates that inappropriate prescription of antibiotics increases later in the day, possibly implicating decision-fatigue in miscalibration (56), a phenomenon that is also known to influence decision-making of judges in the court of law (57). It should be noted that in contrast to all of the previously described influences on miscalibration which promote overconfidence, decision-fatigue appears to promote decision-making behavior consistent with underconfidence - it is in keeping with underconfidence to over-test or over-evaluate patient concerns (11).

Context-related influences on calibration also include case difficulty - as cases become more challenging, calibration is worse (6). Additionally, certain teaching tools frequently used in medical education, such as retrospective case analyses, may promote overconfidence by facilitating hindsight bias, wherein individuals' decision-making and thought processes are influenced by knowing the final outcome before working through a complex problem, which is not how problem solving realistically occurs in the clinical setting (36,51).

THE CONSEQUENCES OF MISCALIBRATION

The tendency of humans to be miscalibrated is not an isolated psychological phenomenon, but one which has implications on their decision-making and behavioral patterns, often to great detriment. The deleterious effects of overconfidence have been substantially examined in both the medical and non-medical domains. Outside of medicine, miscalibration is associated with excessive financial risk-taking and unsuccessful stock trading (16,58). In an extreme example, miscalibration is implicated in politicians "smooth[ing] the tragic road to war" (5,59).

Diagnostic Error

Within the realm of medicine, via a variety of mechanisms, physician miscalibration contributes to diagnostic error (12). One way in which this occurs is through the neglect of tools designed to improve decision-making accuracy and standardization. In a laboratory experiment examining individuals performing a non-medical actuarial judgment task, overconfidence contributed to subjects' reluctance to use an actuarial decision-making aid (60). In analogous research within medicine, miscalibration is implicated in physicians' unreliable judgment of when they do or do not need to consult external clinical decision support or other sources of information (13), likely contributing to their tendency to eschew easily-accessible clinical decision support tools and to buck national clinical guidelines, resulting in suboptimal patient outcomes (12,61,62).

Another mechanism by which miscalibration contributes to diagnostic error is through insufficient updating of *diagnostic schema* in overconfident individuals (63). As in the model of feedback proposed by Croskerry (8), feedback is necessary for physicians to reinforce and/or update their diagnostic schemas as necessary. If feedback data is unavailable or ineffectively communicated, the physician proceeds as if they have received positive feedback, updating of diagnostic schema does not occur, and confidence increases. This *self-confirming bias* acts to further suppress the drive to update diagnostic schema, thus promoting a "vicious reinforcing cycle that erroneously amplifies confidence" (63).

Additionally, all of the individual and contextual factors which contribute to miscalibration, as described in the previous section, conspire to facilitate error-prone clinical decisions. For example, the automatic processing, reliance on heuristics, and cognitive biases all promote physicians to either misemphasize, miscategorize, or completely disregard certain clinical information, leading to incorrect diagnoses (12). A synergistic deleterious effect on diagnostic accuracy is achieved when these faulty cognitive processes are used to solve difficult and uncertain diagnostic problems, all in an environment which does not promote communication of that uncertainty or of effective feedback. The faulty reasoning and the resulting diagnostic errors are then implicitly reinforced, and diagnostic errors are perpetuated. These effects are likely even further magnified in postgraduate trainees, who lack the competence, experience, and metacognition to effectively calibrate.

Excess Healthcare Costs

While no research studies directly assessing the link between miscalibration and healthcare costs were identified, indirect evidence does exist. Overconfident physicians tend to underevaluate patient concerns, while underconfident physicians err towards overevaluation (11). While both of these tendencies may contribute to diagnostic error, overtesting and overtreatment are largely responsible for excessive healthcare costs (64,65). Moreover, in the setting of the overconfident physician, patient concerns which may initially be under-evaluated occasionally manifest at a more advanced stage, contributing to increased healthcare costs, with greater morbidity and mortality conferred to the patient - the rationale for instituting many preventive care policies (66).

APPROACHES TO IMPROVING CALIBRATION

Miscalibration has a profound impact on human behavior, contributing to irrational and inefficacious decision-making and sometimes resulting in suboptimal outcomes, and its correction has long been pursued. Many unsuccessful attempts in the laboratory setting have established that improving calibration is challenging. Moreover, even in instances where recalibration was successfully improved for one task, the improvement would not reliably generalize to different tasks (33). As has been reviewed, miscalibration is a multifaceted and deeply-ingrained component of the human condition, with numerous intrinsic and extrinsic factors which foster its prominence. As such, there are many potential targets at which to aim interventions.

Approaches to improving calibration may be broadly divided into strategies that focus on the individual, and those that address systems (12) (*Table 2*). This framework is particularly useful because it explicitly accounts for both the individual and context-dependent factors which contribute to miscalibration.

Individual-based Approaches

Strategies which focus on the individual postulate that the physician's cognition and/or metacognition must be improved in order to make them more competent, less subject to biases, and more cognizant of their limitations, thereby optimizing their calibration. One of the potential means to this end is through increasing physicians' level of expertise, as those who are more competent are better calibrated (4,13). The most common avenue by which physicians increase their level of expertise is through subspecialization, which has been demonstrated to improve patient care in an array of clinical contexts, but may also facilitate the potential pitfalls of overspecialization (67). A medical education technique with a lot of promise for improving physician competence and expertise is simulation, which is a burgeoning area of interest within radiology training (68−77). Simulation allows for increased exposure to high-fidelity clinical scenarios in a low-stakes setting where instant feedback is possible, thus addressing two key components of calibration - experience and rapid feedback, which evidence supports is more effective than delayed feedback (32).

A deficit or failure of metacognition is often cited as a major culprit behind miscalibration, robbing physicians of the ability to sense when their intuition may be misguided or that they must seek additional sources of information. Several strategies have been proposed to augment individuals' metacognitive aptitude. These are based on the hypothesis that gaining an understanding of one's tendencies to be miscalibrated, particularly when these are based on automatic subconscious processing, can result in improved calibration. This is the same principle which underlies the finding that education on biases in radiology makes individuals less likely to make mistakes based on those biases (45−47), an approach promoted by some as an integral component of medical education at large (36). Examples of strategies for improving metacognition include *enhanced calibration feedback* (60), *cognitive forcing strategies* (9), and *prospective hindsight* (78) (Table 2; Appendix A).

Another technique closely related to these metacognitive training strategies is *reflective practice*, which has been demonstrated as a contributing mechanism in expertise acquisition (79), and may improve diagnostic accuracy in complex scenarios (12,80) (Table 2; Appendix A).

System-based Approaches

System-based approaches to improving calibration are directed at the environment in which the diagnosis is made. Through the optimization of this environment to maximize quantity, quality, and appropriateness of the data provided to the physician, calibration is improved.

Feedback provides information which helps highlight and close the gap between perceived and actual performance. Therefore, systems built to optimize prompt and accurate feedback should aid in the reduction of miscalibration. An

TABLE 2. Approaches to improving calibration

Approach	Summary	Potential Application in Radiology
Individual-based approaches		
Increasing expertise	Individuals who are more competent are better-calibrated	Encourage mini-fellowships and fellowship training to deepen knowledge base in specific domains
Simulation	Educational technique which provides high-fidelity clinical experience in a low-stakes environment	Residents review a curated set of anonymized cases at a PACS station and record their findings and impressions, followed by a review session
Improving metacognition		
Enhanced Calibration Feedback	Improve awareness of miscalibration by providing data regarding one's calibration, and then having them answer a series of questions regarding their degree of miscalibration	During case readouts, ask trainees to indicate their level of confidence. Further explore instances of trainee miscalibration. Incorporate confidence measurements during simulation or other assessment/educational exercises
Cognitive forcing strategies	Combat specific biases and avoid pitfalls by forcing oneself to perform a "metacognitive step" upon an otherwise subconscious decision-making process	Combat the satisfaction of search bias by forcing oneself to complete a thorough search pattern on every case Ask "what diagnoses can I not afford to miss?"
Prospective hindsight	Consider a potential future event as a certainty, like you are looking back on it from a future time point as if it has already happened	Imagine that one's diagnosis is inaccurate, explain why this may be, and broaden the differential diagnosis
Reflective practice	When faced with a complex clinical scenario, search for alternative hypotheses or solutions, consider their possible consequences, test predictions, and remain open to reflection	When reviewing a challenging case with uncertain findings, develop a differential diagnosis, consider the possible consequences of the diagnoses, and reflect on the decision-making process
System-based approaches		
Optimizing feedback	Timely, accurate feedback is necessary for optimizing calibration and diagnostic schema accuracy	May be delivered on a daily basis from attendings to trainees, as part of simulation exercises, and more globally as part of required milestone meetings
Facilitating clinical follow-up	As a type of feedback, knowledge of subsequent clinical data such as pathology or operative reports is essential for optimizing diagnostic calibration	Develop and implement software which allows radiologists to easily access subsequent clinical data for cases they read as it becomes available

example of feedback infrastructure in the clinical radiology setting is the American College of Radiology (ACR) RADPEER program, a system which aims to promote nonpunitive peer learning through peer review of prior reports (81). Alternative peer-review feedback systems with increased customization have been created, allowing for the delivery of more detailed and instantaneous feedback (82). Similar feedback systems have also been implemented in breast imaging, as well as in the field of pathology, with varying degrees of participation and success (12).

Follow-up of clinical outcomes is a specific form of feedback which helps close the gap between diagnostic accuracy and confidence in that accuracy, and is therefore a critical component to improving physician calibration (11). As has been reviewed, the absence of timely clinical feedback reinforces diagnostic schema as if positive outcomes were achieved, serving to falsely increase miscalibration and diminish the drive to adapt the diagnostic schema. Therefore, systems that facilitate and streamline access to pertinent clinical follow-up data as it becomes available in real time are paramount to improving radiologist calibration. One institution with such software studied the utility of automated radiology-pathology feedback by having attending radiologists mark pathology results as concordant or discordant with their imaging interpretation (83). Of the hundreds of discordant cases, the radiologists indicated that they would not have otherwise followed up on the pathology in approximately 70% of cases, and further analysis demonstrated that approximately 40% of discordances were due to missed or misinterpreted findings. Perhaps just as important to the potential improvement of calibration, the radiologists reported learning from ~90% of the discordant cases, with ~40% of the cases influencing future interpretations.

DISCUSSION

Miscalibration is ubiquitous. Its prevalence and magnitude are influenced by numerous intrinsic and extrinsic factors, many of which are fundamental to the clinical setting and the physicians who practice therein. Due to their relative lack of experience, domain-specific competence, metacognitive skills, and the potential for disproportionate gains in confidence which rapidly outpace gains in competence early in the learning process, we hypothesize that radiology residents may be particularly susceptible to miscalibration and its effects.

While the relatively limited body of research on the effects of miscalibration in medicine focuses on its direct and indirect impacts on diagnostic accuracy, we believe that there are additional potential consequences of resident miscalibration. For example, an overconfident resident may not appropriately escalate a patient case or question to an attending, even in situations where patient care would be improved by such an escalation. This is analogous to the finding that overconfident individuals are less likely to consult decision or diagnostic aids (13,60). Overconfident residents may offer overconfident misinterpretations of studies, resulting in suboptimal patient care, either through providing too narrow of a differential diagnosis (or no differential diagnosis), or by failing to properly acknowledge and/or communicate the uncertainty present in the case. Furthermore, radiologists who are overconfident may agree to interpret modalities or studies for which they are inadequately skilled, and may not effectively communicate the degree of uncertainty in their findings.

The potential direct consequences of underconfidence among physicians are essentially unexplored in the literature, perhaps because underconfidence seems to be much less prevalent than overconfidence, and because the consequences may not lead to as many adverse patient outcomes. The indirect link between physician underconfidence and overtesting contributing to increased healthcare costs has been discussed. However, we speculate that there are additional negative consequences to radiologist underconfidence. For example, excessive escalation of cases and questions to attendings by underconfident residents may perpetuate their underconfidence and stymie their developing independence. After residency, the same individuals may tend to consult colleagues for second opinions too frequently, placing excess time demands on others and furthering their own underconfidence. More germane to patient outcomes, underconfident radiologists may overestimate the degree of uncertainty of their findings, and produce reports which are not sufficiently assertive. This can impede or prevent appropriate management (84).

For these reasons, resident calibration should be an explicit goal of radiology training programs. We believe that this can be accomplished through the application of insights and best practices from the vast psychology literature on calibration, and from the medical literature where available. A necessary first step is to develop mechanisms by which to measure resident miscalibration and to communicate the results. A recent publication reviews, numerous statistical and graphical methods to analyze radiologist calibration with regards to their probabilistic diagnostic judgments, and presents a framework for their implementation in practice (85).

Feedback for radiology trainees comes in many forms, and has received substantial attention in the radiology literature in recent years (86−88). Feedback regarding calibration may be implemented on both micro- and macroscopic levels. For example, in day-to-day practice, attendings may regularly challenge trainees with whom they are working to express their level of confidence in their diagnosis, and then discuss in more depth instances in which the trainee is miscalibrated, analogous to *enhanced calibration feedback* (60). Promoting such a routine dialogue is in keeping with the idea that immediate feedback is more effective than delayed feedback (32). On a more global level, end-of-rotation evaluations and the ACGME Milestone meeting provide settings in which feedback may be delivered to trainees (84). Trainees may be overconfident in some competencies and underconfident in others, and providing them feedback on their performance relative to their perceived abilities offers an implicit suggestion that they are miscalibrated. Miscalibration should then become an explicit point of emphasis, as supported by the data demonstrating the efficacy of *enhanced feedback* over basic outcome feedback in the improvement of calibration.

Numerous additional educational techniques with potential to improve calibration should be implemented, and their efficacy investigated. For example, metacognitive debiasing techniques such as cognitive forcing strategies (ie, "what diagnosis can I not afford to miss?"), prospective hindsight, and reflective practice can be introduced early on in the residency curriculum (Table 2, Appendix A). These techniques should at the very least increase awareness of the concept of miscalibration, which may in and of itself improve calibration, similar to the debiasing effect of education on cognitive biases.

Simulation is another educational technique which may aid in efforts to improve calibration. We find that residents enjoy the opportunity to learn from a curated set of real cases at a PACS workstation in an efficient and low-stakes environment, followed by small-group review sessions. Perhaps diagnostic confidence can be recorded within simulations so that calibration may be studied. A metacognitive technique like enhanced calibration feedback could be incorporated into the experience by providing data about an individual's confidence relative to their performance, and subsequently prompting them to answer questions about the nature of their calibration.

Many educational methods already commonly employed by radiology training programs may also improve calibration as an unintended benefit. Examples include "hot-seat" case conferences (89,90) electronic audience response systems such as *RSNA Diagnosis Live* (91) and *Poll Everywhere* (92), dashboards to track and classify trainee case volume (93) and resident assessments which offer varying degrees of feedback such as *ACR RadExam* (94), the *ACR Diagnostic Radiology In-Training Exam* (95), and the *ABR Core Exam* (96). The impact of these tools on calibration should be examined to further optimize their value to resident education.

System-level interventions should also be employed within radiology departments where feasible, and their impact on calibration investigated. Outside of interpersonal feedback, it is important for radiologists to receive clinical and pathology follow-up for studies they interpret. A potentially elegant solution to provide this information to radiologists is through software integrated into PACS which alerts the radiologist and links them to pertinent clinical follow-up as it becomes available (83,97−101). As the emerging data demonstrates, without software facilitating this type of follow-up, radiologists are likely missing out on innumerable learning opportunities and inappropriately reinforcing faulty diagnostic schema, contributing to overconfidence and miscalibration. As we have at our institution, further implementation and investigation of the impact of such software should be sought.

There are additional topics which warrant further investigation beyond the efficacy of the various individual- and system-level interventions discussed previously (84). These include the exploration of factors that are associated with overconfidence and underconfidence in radiology trainees. Is it possible to identify the individuals at greatest risk for miscalibration earlier in their training so that the effects of various recalibration interventions can be maximized? Can specific causative factors of miscalibration be identified, so that they can be mitigated to the best of our abilities? Furthermore, as confidence may be closely linked with “positive emotions” (50), the impact of confidence, and by extension, confidence-altering interventions, on resident well-being should be explored.

CONCLUSION

People are rarely accurate in assessments of their own competence, and have a tendency to be overconfident relative to their actual abilities. This tendency is exacerbated by many internal and external factors which physicians, particularly trainees, contend with during their practice. Physician miscalibration can have dramatic consequences, including suboptimal clinical outcomes and excess healthcare costs. Therefore, appropriate matching of trainee confidence and competence should be an explicit goal of radiology residency and fellowship programs. Through various individual- and system-level interventions focused on providing effective feedback, improving competence and metacognition, and providing more reliable access to clinical follow-up, physician calibration may be improved, thereby mitigating the adverse effects of both overconfidence and underconfidence during training and independent practice.

REFERENCES

1. Lichtenstein S, Fischhoff B, Phillips LD. Calibration of Probabilities: The State of the Art. *Decis Mak Change Hum Aff*. 275–324. doi:10.1007/978-94-010-1276-8_19.
2. Keren G. Calibration and probability judgements: conceptual and methodological issues. Acta Psychol (Amst) 1991; 77(3):217–273. doi:10.1016/0001-6918(91)90036-Y.
3. Alicke MD, Govorun O. The better-than-average effect. The Self in Social Judgment. Psychology Press, 2005:85–106.
4. 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; 77:1121–1134. doi:10.1037/0022-3514.77.6.1121.
5. Sanchez C, Dunning D. Overconfidence among beginners: Is a little learning a dangerous thing? J Pers Soc Psychol 2018; 114(1):10–28. doi:10.1037/pspa0000102.
6. Meyer AND, Payne VL, Meeks DW, et al. Physicians’ diagnostic accuracy, confidence, and resource requests: a vignette study. JAMA Intern Med 2013; 173(21):1952. doi:10.1001/jamainternmed.2013.10081.
7. Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA 2006; 296:1094–1102. doi:10.1001/jama.296.9.1094.
8. Croskerry P. The feedback sanction. Acad Emerg Med 2000; 7 (11):1232–1238. doi:10.1111/j.1553-2712.2000.tb00468.x.
9. Croskerry P. Cognitive forcing strategies in clinical decision making. Ann Emerg Med 2003; 41(1):110–120. doi:10.1067/mem.2003.22.
10. Croskerry P, Norman G. Overconfidence in clinical decision making. Am J Med 2008; 121(5):S24–S29. doi:10.1016/j.amjmed.2008.02.001.
11. Cifu AS. Diagnostic errors and diagnostic calibration. JAMA 2017; 318 (10):905. doi:10.1001/jama.2017.11030.
12. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med 2008; 121(5):S2–S23. doi:10.1016/j.amjmed.2008.01.001.
13. Friedman CP, Gatti GG, Franz TM, et al. Do physicians know when their diagnoses are correct?: implications for decision support and error reduction. J Gen Intern Med 2005; 20(4):334–339. doi:10.1111/j.1525-1497.2005.30145.x.
14. Potchen EJ. Measuring observer performance in chest radiology: some experiences. J Am Coll Radiol 2006; 3(6):423–432. doi:10.1016/j.jacr.2006.02.020.
15. Hodges B, Regehr G, Martin D. Difficulties in recognizing one’s own incompetence: novice physicians who are unskilled and unaware of it. Acad Med 2001; 76(10):3.
16. Barber BM, Odean T. Trading is hazardous to your wealth: the common stock investment performance of individual investors. J Finance 2000; 55(2):773–806.
17. Thompson L, Loewenstein G. Egocentric interpretations of fairness and interpersonal conflict. Organ Behav Hum Decis Process 1992; 51 (2):176–197. doi:10.1016/0749-5978(92)90010-5.
18. Falchikov N, Boud D. Student Self-Assessment in Higher Education: A Meta-Analysis: *Rev Educ Res*. doi:10.3102/00346543059004395 Published online.
19. Svenson O. Are we all less risky and more skillful than our fellow drivers? Acta Psychol (Amst) 1981; 47(2):143–148. doi:10.1016/0001-6918(81)90005-6.
20. Craig P.*The killing zone: how & why pilots die*. 2nd ed. McGrawHill Education; 2013.
21. Schatlo B, Martinez R, Alaid A, et al. Unskilled unawareness and the learning curve in robotic spine surgery. Acta Neurochir (Wien) 2015; 157(10):1819–1823. doi:10.1007/s00701-015-2535-0.
22. O’Callaghan E, Miyamoto A, Takahashi H, et al. Do beginner endoscopists make erroneous diagnosis? Showa Univ J Med Sci 1990; 2(2):121–125. doi:10.15369/sujms1989.2.121.
23. Landazabal MG. Psychopathological symptoms, social skills, and personality traits: a study with adolescents. Span J Psychol 2006; 9 (2):182–193.
24. Detmer DE, Fryback DG, Gassner K. Heuristics and biases in medical decision-making. - PubMed - NCBI. J Med Educ 1978; 58:682–683. doi:10.1097/00001888-197808000-00012.
25. Deaux K, Farris E. Attributing causes for one’s own performance: The effects of sex, norms, and outcome. J Res Personal 1977; 11(1):59–72. doi:10.1016/0092-6566(77)90029-0.
26. Lundeberg MA, Fox PW, Puncochaf J. Highly confident but wrong: gender differences and similarities in confidence judgments. J Educ Psychol 1994; 86:114–121.
27. Lundeberg MA, Fox PW, Brown AC, et al. Cultural influences on confidence: country and gender. J Educ Psychol 2000; 92(1):152–159. doi:10.1037/0022-0663.92.1.152.
28. Muthukrishna M, Henrich J, Toyokawa W, et al. Overconfidence is universal? Elicitation of Genuine Overconfidence (EGO) procedure reveals systematic differences across domain, task knowledge, and incentives in four populations. PLOS ONE 2018; 13(8):e0202288. doi:10.1371/journal.pone.0202288.

29. Moore DA, Dev AS, Goncharova EY. Overconfidence across cultures. Collabra Psychol 2018; 4(1):36. doi:10.1525/collabra.153.
30. Fischhoff B, Slovic P. A little learning...: confidence in multicue judgment tasks. Defense Adv Res Projects Agency 1978: 1–59.
31. Griffin D, Tversky A. The weighing of evidence and the determinants of confidence. Cognit Psychol 1992; 24(3):411–435. doi:10.1016/0010-0285(92)90013-R.
32. Duffy FD, Holmboe ES. Self-assessment in lifelong learning and improving performance in practice: physician know thyself. JAMA 2006; 296(9):1137. doi:10.1001/jama.296.9.1137.
33. Lichtenstein S, Fischhoff B. Training for calibration. Organ Behav Hum Perform 1980; 26(2):149–171. doi:10.1016/0030-5073(80)90052-5.
34. Camerer C, Loewenstein G, Weber M. The curse of knowledge in economic settings: an experimental analysis. J Polit Econ 1989; 97 (5):1232–1254. doi:10.1086/261651.
35. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science 1974; 185(4157):1124–1131. doi:10.1126/science.185.4157.1124.
36. Hall KH. Reviewing intuitive decision-making and uncertainty: the implications for medical education. Med Educ 2002; 36(3):216–224. doi:10.1046/j.1365-2923.2002.01140.x.
37. Bornstein BH, Emler AC. Rationality in medical decision making: a review of the literature on doctors' decision-making biases. J Eval Clin Pract 2001; 7(2):97–107. doi:10.1046/j.1365-2753.2001.00284.x.
38. Brown JD. Evaluations of self and others: self-enhancement biases in social judgments. Soc Cogn 1986; 4(4):353–376. doi:10.1521/soco.1986.4.4.353.
39. Taylor SE, Brown JD. Illusion and well-being:a social psychological perspective on mental health. Psychol Bull 1988; 103:193–210. doi:10.1037/0033-2909.103.2.193.
40. Dunning D, Cohen GL. Egocentric definitions of traits and abilities in social judgment. J Pers Soc Psychol 1992; 63(3):341–355. doi:10.1037/0022-3514.63.3.341.
41. Dunning D, Meyerowitz JA, Holzberg AD. Ambiguity and self-evaluation: the role of idiosyncratic trait definitions in self-serving assessments of ability. J Pers Soc Psychol 1989; 57(6):1082–1090. doi:10.1037/0022-3514.57.6.1082.
42. Sanitioso R, Kunda Z, Fong GT. Motivated recruitment of autobiographical memories. J Pers Soc Psychol 1990; 59(2):229–241. doi:10.1037/0022-3514.59.2.229.
43. Klar Y, Medding A, Sarel D. Nonunique invulnerability: singular versus distributional probabilities and unrealistic optimism in comparative risk judgments. Organ Behav Hum Decis Process 1996; 67(2):229–245. doi:10.1006/obhd.1996.0076.
44. Kruger J. Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. J Pers Soc Psychol 1999; 77(2):221–232. doi:10.1037/0022-3514.77.2.221.
45. Gunderman RB. Biases in radiologic reasoning. Am J Roentgenol 2009; 192(3):561–564. doi:10.2214/AJR.08.1220.
46. Waite S, Scott J, Gale B. Interpretive error in radiology. Am J Roentgenol 2017; 208(4):739–749. doi:10.2214/AJR.16.16963.
47. Busby LP, Courtier JL, Glastonbury CM. Bias in radiology: the how and why of misses and misinterpretations. RadioGraphics 2018; 38(1):236–247. doi:10.1148/rg.2018170107.
48. Stanovich KE, West R. Individual differences in reasoning: Implications for the rationality debate? Behav Brain Sci 2000; 23(5):645–665.
49. Kahneman D. Thinking, Fast and Slow. Farrar, Straus and Giroux, 2011:499.
50. Tiedens LZ, Linton S. Judgment under emotional certainty and uncertainty: the effects of specific emotions on information processing. J Pers Soc Psychol 2001; 81(6):973–988. doi:10.1037/0022-3514.81.6.973.
51. Arkes HR, Dawes RM, Christensen C. Factors influencing the use of a decision rule in a probabilistic task. Organ Behav Hum Decis Process 1986; 37(1):93–110. doi:10.1016/0749-5978(86)90046-4.
52. Fox RC. The evolution of medical uncertainty. Milbank Mem Fund Q Health Soc 1980; 58(1):1–49.
53. Fox RC. Medical uncertainty revisited. Handbook of social studies in health and medicine. SAGE Publications Ltd, 2000:409–425. doi:10.4135/9781848608412.
54. Beresford Eb. Uncertainty and the shaping of medical decisions. Hastings Cent Rep 1991; 21(4):6. doi:10.2307/3562993.
55. Katz J. Why doctors don't disclose uncertainty. Hastings Cent Rep 1984; 14(1):35–44. doi:10.2307/3560848.
56. Linder JA, Doctor JN, Friedberg MW, et al. Time of Day and the Decision to Prescribe Antibiotics. JAMA Intern Med 2014; 174(12):2029. doi:10.1001/jamainternmed.2014.5225.
57. Danziger S, Levav J, Avnaim-Pesso L. Extraneous factors in judicial decisions. Proc Natl Acad Sci 2011; 108(17):6889–6892. doi:10.1073/pnas.1018033108.
58. Broihanne MH, Merli M, Roger P. Overconfidence, risk perception and the risk-taking behavior of finance professionals. Finance Res Lett 2014; 11(2):64–73. doi:10.1016/j.frl.2013.11.002.
59. Johnson D. Overconfidence and war: the havoc and glory of positive illusions. Harvard University Press, 2004.
60. Sieck WR, Arkes HR. The recalcitrance of overconfidence and its contribution to decision aid neglect. J Behav Decis Mak 2005; 18(1):29–53. doi:10.1002/bdm.486.
61. Berner ES, Maisiak RS, Heudebert GR, et al. Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system. AMIA Annu Symp Proc 2003; 2003:76–80.
62. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. N Engl J Med 2003; 348(26):2635–2645. doi:10.1056/NEJMsa022615.
63. Rudolph JW, Morrison JB. Sidestepping superstitious learning, ambiguity, and other roadblocks: a feedback model of diagnostic problem solving. Am J Med 2008; 121(5):S34–S37. doi:10.1016/j.amjmed.2008.02.003.
64. Carpenter CR, Raja AS, Brown MD. Overtesting and the downstream consequences of overtreatment: implications of "preventing overdiagnosis" for emergency medicine. Acad Emerg Med 2015; 22(12):1484–1492. doi:10.1111/acem.12820.
65. Berwick DM, Hackbarth AD. Eliminating waste in US health care. JAMA 2012; 307(14):1513–1516. doi:10.1001/jama.2012.362.
66. Cohen JT, Neumann PJ, Weinstein MC. Does preventive care save money? health economics and the presidential candidates. N Engl J Med 2008; 358(7):661–663. doi:10.1056/NEJMp0708558.
67. Oren O, Gersh BJ, Bhatt DL. On the pearls and perils of sub-subspecialization. Am J Med 2020; 133(2):158–159. doi:10.1016/j.amjmed.2019.06.052.
68. Desser TS. Simulation-based training: the next revolution in radiology education? J Am Coll Radiol 2007; 4(11):816–824. doi:10.1016/j.jacr.2007.07.013.
69. Towbin AJ, Paterson B, Chang PJ. A computer-based radiology simulator as a learning tool to help prepare first-year residents for being on call. Acad Radiol 2007; 14(10):1271–1283. doi:10.1016/j.acra.2007.06.011.
70. Wang CL, Schopp JG, Petscavage JM, Paladin AM, Richardson ML, Bush WH. Prospective randomized comparison of standard didactic lecture versus high-fidelity simulation for radiology resident contrast reaction management training. Am J Roentgenol 2011; 196(6):1288–1295. doi:10.2214/AJR.10.5618.
71. Wang CL, Chinnugounder S, Hippe DS, et al. Comparative effectiveness of hands-on versus computer simulation−based training for contrast media reactions and teamwork skills. J Am Coll Radiol 2017; 14 (1):103–110. doi:10.1016/j.jacr.2016.07.013.
72. Chetlen AL, Mendiratta-Lala M, Probyn L, et al. Conventional medical education and the history of simulation in radiology. Acad Radiol 2015; 22(10):1252–1267. doi:10.1016/j.acra.2015.07.003.
73. Picard M, Curry N, Collins H, et al. Comparison of high-fidelity simulation versus didactic instruction as a reinforcement intervention in a comprehensive curriculum for radiology trainees in learning contrast reaction management: does it matter how we refresh? Acad Radiol 2015; 22(10):1268–1276. doi:10.1016/j.acra.2015.01.021.
74. Cook TS, Hernandez J, Scanlon M, et al. Why isn't there more high-fidelity simulation training in diagnostic radiology? results of a survey of academic radiologists. Acad Radiol 2016; 23(7):870–876. doi:10.1016/j.acra.2016.03.008.
75. Klein KA, Neal CH. Simulation in radiology education: thinking outside the phantom. Acad Radiol 2016; 23(7):908–910. doi:10.1016/j.acra.2016.02.013.
76. Roark AA, Ebuoma LO, Ortiz-Perez T, et al. Impact of simulation-based training on radiology trainee education in ultrasound-guided breast biopsies. J Am Coll Radiol 2018; 15(10):1458–1463. doi:10.1016/j.jacr.2017.09.016.
77. Mirza S, Athreya S. Review of simulation training in interventional radiology. Acad Radiol 2018; 25(4):529–539. doi:10.1016/j.acra.2017.10.009.

78. Mitchell DJ, Edward Russo J, Pennington N. Back to the future: temporal perspective in the explanation of events. J Behav Decis Mak 1989; 2 (1):25–38. doi:10.1002/bdm.3960020103.
79. Mamede S, Schmidt HG. The structure of reflective practice in medicine. Med Educ 2004; 38(12):1302–1308. doi:10.1111/j.1365-2929.2004.01917.x.
80. Soares S. Reflective practice in medicine (PhD Thesis). 2006.
81. Goldberg-Stein S, Frigini LA, Long S, et al. ACR RADPEER committee white paper with 2016 updates: revised scoring system, new classifications, self-review, and subspecialized reports. J Am Coll Radiol 2017; 14(8):1080–1086. doi:10.1016/j.jacr.2017.03.023.
82. Swanson JO, Thapa MM, Iyer RS, et al. Optimizing peer review: a year of experience after instituting a real-time comment-enhanced program at a children's hospital. Am J Roentgenol 2012; 198(5):1121–1125. doi:10.2214/AJR.11.6724.
83. Doshi AM, Huang C, Melamud K, et al. Utility of an automated radiology-pathology feedback tool. J Am Coll Radiol 2019; 16(9):1211–1217. doi:10.1016/j.jacr.2019.03.001.
84. Schweitzer AD. The importance of appropriate matching of confidence and competence in radiology training and beyond. Clin Imaging 2020; 66:64–66. doi:10.1016/j.clinimag.2020.04.031.
85. Amini B, Bassett RL, Haygood TM, et al. Confidence calibration: an introduction with application to quality improvement. J Am Coll Radiol 2020; 17(5):620–628. doi:10.1016/j.jacr.2019.12.009.
86. Gill AE, Wong PK, Mullins ME, et al. Missed case feedback and quality assurance conferences in radiology resident education: a survey of United States radiology program directors. Curr Probl Diagn Radiol 2018; 47(4):209–214. doi:10.1067/j.cpradiol.2017.06.008.
87. Choi HH, Clark J, Jay AK, et al. Minimizing barriers in learning for on-call radiology residents—end-to-end web-based resident feedback system. J Digit Imaging 2018; 31(1):117. doi:10.1007/s10278-017-0015-1.
88. Chen P-H, Scanlon MH. Teaching radiology trainees from the perspective of a millennial. Acad Radiol 2018; 25(6):794–800. doi:10.1016/j.acra.2018.02.008.
89. Redmond CE, Healy GM, Clifford S, et al. Radiology resident education. Acad Radiol 2017; 24(10):1327. doi:10.1016/j.acra.2017.04.015.
90. Taylor-Cho MW, Ronald J, Grimm LJ, et al. Knowledge retention in radiology residents for audience response system versus traditional hot-seat conference. J Am Coll Radiol 2020:S1546144020307535. doi:10.1016/j.jacr.2020.07.008. Published online.
91. Awan OA, Shaikh F, Kalbfleisch B, et al. RSNA diagnosis live: a novel web-based audience response tool to promote evidence-based learning. RadioGraphics 2017; 37(4):1111–1118. doi:10.1148/rg.2017160107.
92. Engaging 100 radiology students with innovative polling. Poll everywhere. Accessed November 22, 2020. Available at: https://www.polleverywhere.com/case-studies/interactive-lecture
93. Chen H, Gangaram V, Shih G. Developing a more responsive radiology resident dashboard. J Digit Imaging 2019; 32(1):81–90. doi:10.1007/s10278-018-0123-6.
94. Lewis PJ, Nyberg E, Cayere J, et al. Educational crowdsourcing: developing radexam. J Am Coll Radiol 2017; 14(6):800–803. doi:10.1016/j.jacr.2017.01.033.
95. Monticciolo DL. The ACR diagnostic radiology in-training examination: evolution, current status, and future directions. J Am Coll Radiol 2010; 7(2):132–137. doi:10.1016/j.jacr.2009.10.010.
96. Guiberteau MJ, Balfe DM, Jackson VP, et al. The ABR core exam. Acad Radiol 2016; 23(9):1180–1182. doi:10.1016/j.acra.2016.06.003.
97. Kovacs MD, Mesterhazy J, Avrin D, et al. Correlate: a PACS- and EHR-integrated tool leveraging natural language processing to provide automated clinical follow-up. RadioGraphics 2017; 37(5):1451–1460. doi:10.1148/rg.2017160195.
98. Dibble EH, Binns E, Ellermeier A, et al. Automated delivery of clinical follow-up to the radiologist via e-mail: feasibility study of a new information technology algorithm. Am J Roentgenol 2018; 211(2):400–404. doi:10.2214/AJR.17.19375.
99. Moore W, Doshi A, Bhattacharji P, et al. Automated radiology-operative note communication tool; closing the loop in musculoskeletal imaging. Acad Radiol 2018; 25(2):244–249. doi:10.1016/j.acra.2017.08.016.
100. Dane B, Doshi A, Gyftopoulos S, et al. Automated radiology-pathology module correlation using a novel report matching algorithm by organ system. Acad Radiol 2018; 25(5):673–680. doi:10.1016/j.acra.2017.11.009.
101. Filice RW. Radiology-pathology correlation to facilitate peer learning: an overview including recent artificial intelligence methods. J Am Coll Radiol 2019; 16(9):1279–1285. doi:10.1016/j.jacr.2019.05.010.

APPENDIX A - GLOSSARY

Heuristics

Anchoring heuristic – Acceptance and reliance upon the first piece of information received before making a decision.

Availability heuristic - The reliance on recent examples that come to mind when confronted with a situation perceived as similar; closely related to recency bias.

Representativeness heuristic - The assessment of similarities of items based on their characteristics and classifying them based on ideal category prototypes

Dynamic influence – A situation-specific influence on physician calibration (i.e., a recent diagnostic experience, which may lead to the overestimation of the pretest probability of a diagnosis which was recently encountered in a previous patient)

Static influence – An influence on physician calibration which is constant from patient to patient and from day to day (i.e., the epidemiology of a physician's patient population, which may lead to over- or under-consideration of certain diagnoses, based on conditions which have been diagnosed in the past within that population)

Metacognitive & Related Techniques:

Enhanced calibration feedback – Improve awareness of miscalibration by providing data regarding one's calibration, followed by a series of questions intended to reinforce their level of miscalibration, intended to reduce overconfidence and improve calibration. Builds upon basic outcome feedback and basic calibration feedback.

Outcome feedback – Providing data regarding only task performance and accuracy

Calibration feedback – Providing data regarding an individual's performance relative to their confidence, without any additional reinforcement exercise

Cognitive forcing strategies – The forced performance of a conscious metacognitive step upon the subconscious decision-making process, such as asking the questions "what diagnoses can I not afford to miss?" or "why might a different diagnosis be correct?" By forcing the consideration of possible alternative solutions or findings, individuals are debiased and made better-calibrated

Prospective hindsight – Debiasing technique which entails developing an explanation for a potential future event as if it has already happened.

Reflective practice – The conscious reflection and critical thinking about one's thought processes and professional activities, intended to help navigate complex and/or uncertain scenarios through the exploration of alternative solutions, considering their possible consequences, testing predictions, and reflecting upon the problem-solving process