american physiological society

ADVANCES IN PHYSIOLOGY EDUCATION

Adv Physiol Educ 47: 71–81, 2023.
First published August 19, 2022; doi:10.1152/advan.00156.2021

EDUCATION RESEARCH

A metacognitive confidence calibration (MCC) tool to help medical students scaffold diagnostic reasoning in decision-making during high-fidelity patient simulations

Luciana S. Garbayo,[1,2,4] David M. Harris,[1,4] Stephen M. Fiore,[1,3] Matthew Robinson,[1,4,5] and Jonathan D. Kibble[1,4]

[1]*Department of Medical Education, University of Central Florida, Orlando, Florida;* [2]*Department of Philosophy, University of Central Florida, Orlando, Florida;* [3]*Institute for Simulation and Training, University of Central Florida, Orlando, Florida;* [4]*College of Medicine, University of Central Florida, Orlando, Florida; and* [5]*Burnett School of Biomedical Sciences, University of Central Florida, Orlando, Florida*

Abstract

The purpose of this study was to *1*) help novice students scaffold problem-solving and engage safely in the deliberate practice of diagnostic reasoning and medical decision-making in real time; *2*) assess how accurately students gather and apply data in medical reasoning and treatment during high-fidelity patient simulations (HFPSs); *3*) identify students' scientific misconceptions related to the case; *4*) promote student metacognitive processing, self-assessment, and self-efficacy; and *5*) facilitate the explicit calibration of student confidence in deliberate reasoning with patient outcomes. In a mixed-method design, a metacognitive calibration self-assessing (MCC) survey tool was applied to HFPS (n = 80, 20 teams of 6 medical students) and semistructured interviews were conducted with faculty (n = 5). When scored by faculty with a rubric, the mean student accuracy ranged from 23% to 74%, whereas their self-assessment of confidence ranged from 71% to 86%. This result revealed overconfidence bias in novice students regarding the correctness of their wrong responses. The most common misconception identified was inverting cause and effect: metabolic acidosis was pointed to as the cause of the patient's problems rather than a consequence of untreated diabetes mellitus. The most common treatment error was overtreatment, with unnecessary added medication. Interviews with faculty suggested that the MCC tool improved the team process by slowing students down, requiring them to think through their answers, and that overall the tool improved their critical thinking. This study demonstrated the feasibility of using a metacognitive confidence calibration tool to assist novice students in learning safely to make deliberate diagnostic reasoning and decisions on patient care in real time during complex simulations while observing objectively their levels of psychological confidence against patient outcomes.

NEW & NOTEWORTHY This study demonstrates the feasibility of a metacognitive confidence calibration tool (MCC) to assess and promote novices in the learning of diagnostic reasoning and treatment decisions on patient care in real time during high-fidelity patient simulations while comparing confidence and accuracy data and identifying students' scientific misconceptions. Results revealed the presence of overconfidence bias, overtreatment, and the misconception of metabolic acidosis as the cause of the patient's problems rather than a consequence of untreated diabetes mellitus.

diagnostic reasoning; ketoacidosis overtreatment; metacognition; overconfidence bias; scientific misconceptions

INTRODUCTION

One of the main goals of medical educators is to train future physicians to successfully adapt and respond to challenging clinical situations under constraints in practice. An important barrier is the recognition that medicine, based on "flesh and blood" decision-making (1), includes both cognitive and affective biases as additional challenges to the development of effective medical education strategies under constraints (2).

One factor contributing to the quality of a practitioner's medical decision-making is the constant ability to self-assess against uncertainty and ignorance on a case-by-case basis. Metacognition is a necessary skill for successful self-monitoring and refining action with the avoidance of unnecessary and/or harmful medical interventions (3, 4). Within the breadth of metacognition, critical reasoning, as the ability to reflect and apply normative standards of reasoning to problem-solving, is a fundamental skill for avoiding many biases. In particular, the Dunning–Kruger effect (5), the bias of overconfidence, is of interest in the context of a practitioner's self-assessment. This bias prevails when the subject disregards the gaps in the knowledge base and acts confidently nonetheless, displaying a

Correspondence: L. S. Garbayo (Luciana.Garbayo@ucf.edu).
Submitted 20 September 2021 / Revised 3 August 2022 / Accepted 11 August 2022

higher level of psychological confidence—against prudence and safety.

To develop a strong metacognitive, critical reasoning capacity, medical students need to both *1*) revise their confidence level in every decision, or learning response, in the context of the relevant piece of training in a progression of expertise development and *2*) be able to learn productively from mistakes in different case scenarios, in practice, while not harming patients. Yet, given the uncertain nature of the clinical contexts as well as the presence of personal biases (6), it is often very difficult for medical students to correctly calibrate their knowledge under constraints while also becoming systematically aware of metacognitive challenges to their medical decision-making performance.

To help students safely and ethically self-assess and learn from their mistakes during the deliberate practice of diagnostic reasoning, we have developed an innovative student self-assessment metacognitive calibration instrument (the MCC tool for short). This tool was designed to help students learn to sustain deliberate awareness of their reasoning process in simulated clinical environments. The MCC design integrates elements of the concept inventory approach for the identification of scientific misconceptions (7, 8) with explicit metacognitive self-monitoring prompted by taking a survey during scenario-based simulations (9). Students identify how confident they really are about their ideas and decisions with a Likert scale and better qualify and distinguish their guesses from solid understanding in practice. The MCC is an effort to produce this assessment integration, aiming at developing metacognitive learning excellence and overcoming gaps in the literature regarding medical students' self-assessment of their diagnostic skills.

Background: Metacognition

Metacognition was originally defined as the knowledge a learner has about their cognitive activities during the learning process as well as what level of skill they have in regulating cognitive activity (10). The ideal self-regulated learner is able to use metacognitive knowledge to effectively self-assess and has the metacognitive skill to regulate effort and learning strategy (11). When this is combined with high intrinsic motivation and low anxiety, we would have the description of an ideal learner. But for many learners, this self-assessment process does not come naturally and it takes deliberate practice to improve it (12, 13). Our tool was designed to guide students through a deliberate sequence of problem-solving steps and, by gauging their confidence level at each step, leading to a deliberate and piecemeal progress by self-monitoring and self-evaluating processes (9).

Significance: Medical Error and Patient Safety in the Medical Curriculum

At the beginning of the twenty-first century, the Institute of Medicine (IOM) published a landmark report about the critical state of patient safety: *To Err is Human: Building a Safer Health System* (14). Data at that time indicated that between 44,000 and 98,000 deaths per year were caused by medical errors in United States hospitals, placing medical error in the top 10 causes of death. A decade later, the World Health Organization (WHO) summarized the evidence on patient safety (15) demonstrating the global nature of the problem, with an estimated 1 in 10 patients worldwide being harmed to some degree while receiving health care. A global burden of disease study produced world estimates that further integrated adverse data on morbidity and mortality to medical errors (16). Other estimates were proposed (17), ranking medical error as the third leading cause of death in the United States. Against this backdrop, medical schools are now striving to improve education and training in patient safety reasoning and overall best practices. Accordingly, the curriculum inventory submitted for medical school accreditation by the Liaison Committee on Medical Education (LCME) now includes patient safety instruction as a defined category (18).

The aforementioned WHO and World Alliance for Patient Safety Report (2008) pointed, in particular, to the lack of appropriate knowledge and the difficulty some practitioners have in knowledge transfer to diverse clinical contexts as major structural factors leading to unsafe care. It also identified misdiagnosis as one of the five major processes that lead to unsafe care. To avoid misdiagnosis, it is also well known in medical education research that the acquisition of expertise in diagnosing requires deliberate reflective practice, with strong metacognitive regulatory control. If this is not achieved, an unreflected practice instead leads to the formation of the physician as an "experienced nonexpert," a problematic training outcome in diagnostic reasoning and medical decision-making (19, 20). An appropriate level of medical decision-making reasoning and practice is associated with the reflective, deliberate development of a highly complex cluster of skills that takes years to master and is also ethically required for a high standard of medical professionalism (21). Based on scientific methodologies to reduce uncertainty (and to improve patient engagement), students are taught in diagnostic reasoning initially the hypothetico-deductive model, in which hypotheses are generated from the medical history and physical examination. Drawing one's hypotheses further from knowledge of disease prevalence and physio-pathological reasoning, additional rational choices of laboratory tests and imaging may be used to refine and reevaluate hypotheses before a final working diagnosis is made (22). Experienced physicians develop heuristics and pattern recognition from experience that allow rapid processing of large amounts of information to expedite diagnosis (23). Decision-making errors are known to be caused by a variety of unreflected cognitive and affective biases, which required deliberate practice to endure (24, 25). For example, anchoring on an initial diagnosis leading to premature closure is a common cognitive mistake that requires deliberate self-assessment (2). Overconfidence bias, a sense of "illusory superiority" in one's ability to perform, known as the Dunning–Kruger effect (5), is another common problem. The miscalibration of "the relationship between diagnostic accuracy and confidence in that accuracy" (26) is a clear threat to patient safety and represents an opportunity for learning intervention.

High-fidelity patient simulation (HFPS) is a real life-like educational clinical situation centered around lifelike mannequins as patients, with simulated physiological reactions. Such types of simulations provide students with a situated learning opportunity to hone their clinical skills, with no risk

of harming real patients. Exposing students early in the curriculum to HFPS is a way to support deliberate metacognitive learning in compressed time and clinical context, toward improving their medical decision-making with the integration of pathophysiology with basic concepts in medical diagnosing. The overall goal of the present study was to create an educational tool for use during HFPS to help students scaffold the diagnostic process in a deliberate, self-reflective manner and to make medical decisions based on sound physiological reasoning.

MATERIALS AND METHODS

Rationale for Development of the MCC Tool

This study is the result of interdisciplinary team science medical education research efforts including contributions in the domains of medical decision-making, ethics, and epistemology (L. S. Garbayo), physiology and HFPS (J. D. Kibble and D. M. Harris), human factors/cognitive psychology/team learning (S. M. Fiore), and methods and statistical analysis (L. S. Garbayo, J. D. Kibble, D. M. Harris, S. M. Fiore, and M. Robinson).

Since medical errors are mostly derived from systemic causes rather than personal causes, medical errors are seldom individualized. There is in fact a special convergence between systemic and personal causes of medical error in physician overconfidence bias (27), which makes it a key topic for both human factors psychology research and epistemology and (medical) decision science. Although overconfidence bias may not be easily eliminated given its unconscious entrenchment, support for the deliberate practice of diagnostic reasoning and for the deliberate calibration of one's overconfidence in diagnostic reasoning against patient outcomes may provide learners with better means and enough feedback for the development of appropriate self-efficacy and reflective expertise. A metacognitive overconfidence calibration (MCC) tool may help in guiding novices to achieve over time more clarity between diagnostic reasoning process and patient outcomes, with increased self-awareness of overconfidence bias in a scientific frame for the revision of beliefs (28). The MCC tool complements other efforts such as Richie and Josephson's heuristic tool (29), which focuses on quantifying anchoring, availability, and representativeness biases to improve the calibration of medical decisions.

The MCC tool was developed within the medical curriculum of our medical school for early utilization when students begin constructing their mental models with medical physiology, which will underpin their diagnostic reasoning. The appropriate mastery of physiology knowledge and its transfer is a structural factor for the success of medical decision-making. This study also provides an opportunity to integrate patient safety curriculum in areas such as appropriate knowledge transfer, diagnostic reasoning, and critical reasoning skills whenever physiology instruction is delivered through pedagogies that require students to transfer physiological knowledge to a medical problem and whenever they are asked to attempt safely to devise a medical diagnosis and recommend a treatment.

One example of such pedagogies for erring safely with medical diagnosis is the aforementioned use of high-fidelity patient simulations (HFPSs), which are being increasingly used to teach applications of medical physiology (30–33). Such simulations provide the necessary patient outcomes in real time to help novices to improve timely medical decision-making and to illuminate their reasoning in context. We have previously shown the utility of such simulations to enhance learning in physiology (30) and have recently piloted an instrument to assess critical thinking in this setting (34).

Study Design and Institutional Review Board Approval

This mixed study encompasses both an observational analytical (quantitative) component, with a single student cohort, and a descriptive (qualitative) component, using semistructured faculty interviews. This study was exempted by the University of Central Florida (UCF) Institutional Review Board (IRB). All students were provided with approved IRB explanation of research for research participation prior to the simulation. A subset of the students agreed confidentially to become research participants and provided their informed consent (online opt in/out). The Physiology faculty had access only to the anonymized list of research participants.

Population of Interest

The population of interest was medical students (applicable to learners in other health professions as well).

Study Objectives

1) Build a survey metacognitive confidence calibration tool for use during an HFPS event to help novice students engage safely in the deliberate practice of diagnostic reasoning, problem-solving, and medical decision-making with a simulated patient in real time.
2) Assess how accurately students gather information and apply data during HFPS.
3) Make explicit students' scientific misconceptions about physiology related to the case.
4) Provide students with a sustained stimulus for metacognitive processing and increased self-assessment and self-efficacy.
5) Identify the correlation between student responses and accuracy to calibrate their level of confidence with patient outcomes and team medical decision-making.

Main Hypothesis

Medical students as novice learners shall display overconfidence bias in the process of clinical reasoning, whereas such bias shall recede with the support of a deliberate reasoning process and sustained self-assessment.

Additional Hypothesis

Systematic patterns of errors in physiology reasoning shall emerge with systematic misconceptions in the learning to critically apply physiology concepts in a clinical decision-making situation elicited in a teaching simulation condition.

Intervention: Educational Context

The University of Central Florida (Orlando) College of Medicine offers an integrated 4-yr MD program. Foundational basic medical sciences are taught in the first year in Human Body (HB) modules that leverage traditional synergies between disciplines (e.g., physiology is taught together with anatomy, microbiology is taught with immunology, etc.). A systems-based curriculum is used in the second year with a focus on the study of disease processes; students take the USMLE Step 1 at the end of the second year. The last phase of the curriculum is translation of knowledge and skills into practice and is represented by clerkship and elective rotations in the third and fourth academic years. This study was based in the Structure and Function module in the first year, which is a 16-week course covering anatomy and physiology. Pedagogy includes a wide variety of session types such as lectures, web-based self-learning modules, anatomy dissection laboratory, case-based learning, team-based learning (TBL), high-fidelity patient simulations, and ultrasound sessions. Assessment includes midterm and final multiple-choice and anatomy practical examinations.

High-Fidelity Description: Diabetic Ketoacidosis Case

The course included four high-fidelity simulations, which were originally developed as a means for students to apply their knowledge of physiology. The topics, in chronological order, were *1*) heat exhaustion, *2*) heart failure, *3*) respiratory failure, and *4*) diabetic ketoacidosis (DKA). Each simulation was designed to run in five rotations, and each rotation had 4 groups (teams) of around six students, for a total of 20 groups. The simulation activity started with the four teams in the same rotation being given a 5-min prebrief describing the context of patient admission to the emergency room. Students in each of the four teams then spent 15 min in the immersive simulation attempting to diagnose the problem and to develop a treatment plan. Faculty facilitators in each room acted as nurses who could give treatments and directed students to follow the study protocol but otherwise did not intervene in the group process. The four teams were then immediately brought back together for a 30-min after-action debrief with two lead faculty members, the purpose of which was to review the key findings, elucidate physiological mechanisms, and reason appropriate treatment. The session concluded with each team having a breakout session to do an informal self-assessment of their team process and performance.

The DKA simulation was used to pilot test the use of the MCC tool. By this stage in the course, students were familiar with the general setting of simulations and the capabilities of the mannequin and were also familiar with assigning group members to team roles. The simulated DKA patient was programmed to present with hypotension, tachycardia, and tachypnea but had normal arterial oxygen saturation and body temperature. If no treatments were selected, the patient slowly deteriorated. The patient was unconscious on presentation, but the simulation included a roommate who could provide the patient's history of recent polyuria and polydipsia if asked. A patient chart was provided that included a metabolic panel showing severe hyperglycemia as well as hyperkalemia and ketonuria. Arterial blood gas data were provided for students to interpret and revealed a severe acute metabolic acidosis. The expected initial treatments should have been intravenous isotonic saline and insulin to address the cardiovascular instability and hyperglycemia, respectively. The faculty facilitator ended the simulation if groups reached this end point. The debriefing session included further consideration of what adjustments to treatment would be needed over the next 24–48 h (i.e., potassium and glucose addition to the intravenous infusion).

Tool Development and Content Validation

The MCC tool was developed and validated for form and content by three investigators (L. S. Garbayo, D. M. Harris, and J. D. Kibble). The MCC survey tool and faculty rubric are shown in Table 1.

Data Collection

The survey was sent to students electronically via individual iPads when they entered the simulation room. The

Table 1. *Student survey and faculty scoring rubric*

Item	Question	Faculty Rubric
1	What sources of information are available in this scenario?	Mannequin, roommate, paper chart, ICU monitor, medication list, faculty member acting as a nurse. [6 points]
2	What problems require your attention?	Unconscious patient, abnormal CV vital signs, abnormal respiratory vital signs, glucose/ketones abnormal, metabolic acidosis, electrolyte disorders. [6 points]
3	Provide a priority order of the problems.	Correct priority order: *1*) low mean arterial blood pressure, *2*) hyperglycemia/ketosis, *3*) electrolyte disorder. [6 points; 2 points each for right order, 1 point each if present but not in right order]
4	Explain your ordering using physiology.	Low cardiac output is due to volume depletion from polyuria. Cardiovascular instability takes first priority. Severe hyperglycemia due to lack of insulin produces polyuria. Lack of insulin opposition to glucagon causes ketone production and acidosis.
5	Which intravenous fluids (if any) are appropriate? [Students could select from isotonic saline, half-isotonic saline, 5% dextrose, and half-isotonic + 5% dextrose]	Isotonic saline is the best choice because it is retained in the ECF and best addresses the volume depletion from renal water and sodium losses. [1 point yes, 0 points no]
6	Which medications (if any) are appropriate? [Students could select from insulin, glucagon, epinephrine, albuterol, atenolol, dobutamine, bicarbonate, and potassium]	Correct treatment: insulin only [1 point yes, 0 points no]

CV, cardiovascular; ECF, extracellular fluid; ICU, intensive care unit.

order in which survey items were presented was intended to guide students through a diagnosis reasoning process starting with information gathering and physiological reasoning before the consideration of a treatment plan was developed.

The protocol for the MCC tool deployment is shown in Fig. 1. Faculty facilitators ensured that each individual student completed the first four survey items before any group discussion occurred. This typically took around 5 min. Each group was then instructed to discuss and attempt to reach a consensus about the problem list but were told not to announce diagnosis or treatment suggestions yet. Once a group's problem list was agreed upon, students were instructed to individually answer the last two survey questions about treatments they would individually indicate. The group was then allowed to discuss for the remaining 5–7 min what they thought the correct diagnosis was and to apply any treatments they had agreed upon.

An important aspect of the survey deployment was a real-time calibration of confidence level for every question; students rated their confidence level on a 7-point Likert scale from Entirely Confident to Entirely Not Confident.

The faculty facilitators were individually interviewed through open-ended, semistructured interviews conducted and recorded by L. S. Garbayo to gain insight into the impact of introducing the MCC tool on the dynamics of a simulation session and to provide insight about whether the tool might improve reasoning, according to them. Interviews were transcribed verbatim, and a thematic analysis was performed to capture the faculty's viewpoint.

Each session was video recorded with two in situ cameras. Videos were reviewed to determine which consensus treatments were given after team discussion. L. S. Garbayo observed samples of the sessions incognito in real time through the cameras. Two expert physiology faculty members (D. M. Harris and J. D. Kibble) graded the student responses with the rubric and met to agree on a consensus score for each student.

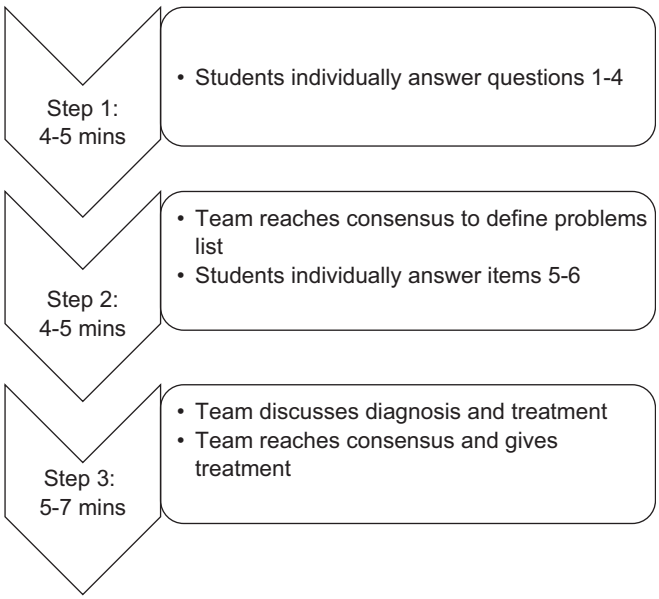

Figure 1. Schematic showing the protocol for survey tool during the high-fidelity patient simulation (HFPS).

Student Participants

This pilot study was undertaken with a single cohort of 120 first-year medical students, made up of 59 females and 61 males, with an average age of 24.7 yr and an age range of between 22 and 45 yr. Eighty students gave consent for survey data to be anonymously included in the study (37 female; 43 male).

Faculty Participants

A total of eight faculty members facilitated the HFPS lesson, of which five gave consent for interview data to be used and were interviewed; two faculty instructors were impeded as investigators and expert scorers; one instructor/teaching assistant consented but was not available to be interviewed.

Data Presentation and Statistical Analysis

This study included the following types of data: student accuracy expressed as percent task correction as scored by faculty rubric, student self-reported confidence level gathered on a 7-point Likert scale and converted to percentage of confidence for comparison to accuracy, and qualitative data from faculty interviews. Quantitative data are expressed as percentages and are reported as medians and ranges, as well as means and SDs. Descriptive statistics were also tabulated for the demographic data on student sex. When the data were not normal and heterogeneous, nonparametric tests were used for statistical analyses. To assess differences between the absolute level of percent student accuracy (i.e., rubric score) and the corresponding level of confidence on each item, the Wilcoxon signed ranks test was used separately for all item score-confidence pairs. For each survey item, correlations were also performed between accuracy and confidence, by first standardizing both variables with z scores. Finally, differences in the level of accuracy and of confidence across different questions were assessed with Kruskal–Wallis tests. Follow-up Mann–Whitney U tests were applied to locate differences between survey items; a Bonferroni correction was used to address the multiplicity of analyses. All statistical tests were two tailed; $P < 0.05$ was used for statistical significance. Statistical analyses were conducted with the Statistical Package for the Social Sciences (SPSS v. 24.0; IBM, Chicago, IL).

RESULTS

Demographic Characteristics of Study Participants and Graded Performance

The sample had 37 females and 43 males. Figure 2 compares the accuracy scores between males and females for each quantitative survey item and shows that there was no meaningful overall difference between male and female graded performance.

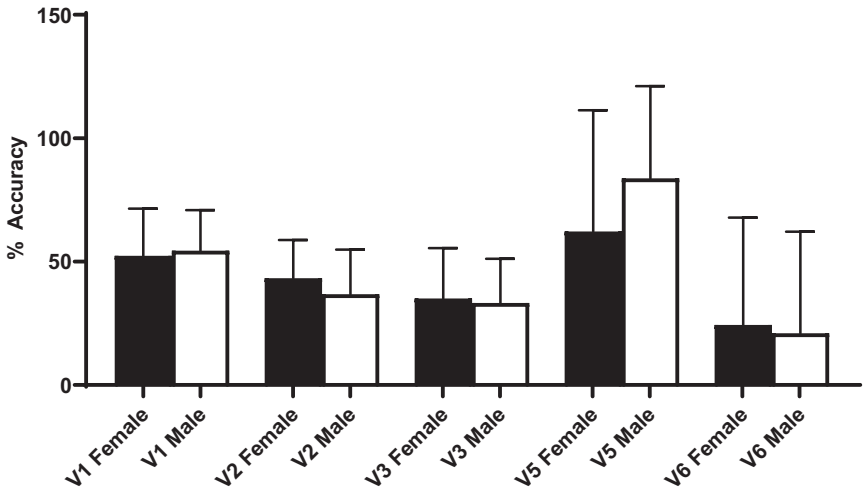

Figure 2. Comparison of graded performance outcomes between males and females. Outcomes are expressed as % accuracy with standard deviations shown for each survey question (denoted as V for "variable"). No significant differences were noted for any variable (Mann–Whitney test, n = 37 female, 43 male).

Student Accuracy Findings

When judged against the rubric, the general level of student accuracy in item responses in the survey was low, ranging from 23% to 74% (Table 2). In detail:

- The first survey item was intended to be the most straightforward, asking students to document sources of information they perceived in the room, but around only half of the major sources of information in the scenario were documented on average. The most-identified sources of information were the monitor, patient chart, simulated patient, and roommate; the medication list and the faculty member acting as a nurse were rarely listed as potential sources of information.
- Individual students found even more difficulty sorting out the abnormal data and prioritizing a problem list, with only about one-third of students answering these tasks correctly.
- After their team consultation to consolidate the problem list, individual students were more successful in identifying the correct intravenous fluid to use from a choice of four commonly used options. The most difficult item proved to be selecting the correct initial medication and avoiding the temptation of adding other unnecessary treatments or overtreatment. Less than one-quarter of students selected insulin alone. Figure 3 summarizes the frequency distribution of medication choices, where >75% of students correctly identified insulin as a needed treatment, yet most individuals then wanted to include additional unnecessary treatments. Most often, students wanted to include a bicarbonate infusion to address the metabolic acidosis, despite classroom instruction that acid-base disturbances are generally addressed by identifying and treating the underlying cause.
- Video review showed that team discussion improved the actual treatments given to the patient: 20/20 groups gave at least insulin, 17/20 groups correctly gave isotonic saline, and 10/20 groups gave the correct combination of normal saline and insulin without any additional unnecessary medications. Of the 10 groups that gave an additional treatment, 8 gave bicarbonate, 1 dobutamine (a positive inotrope), and 1 potassium chloride; no group gave more than two additional medications.

From the point of view of learning taxonomy and cognitive demand, the items presented to students in the survey prompts became progressively more challenging. The first two items required recognition and description, and the third item required ordering of priorities, leading to a requirement for explanations and decision-making in *items 4–6*. The median confidence level found in this study for the ability to describe information sources and patient problems (*items 1* and *2*) was "Mostly Confident" on the rating scale. As the items on the survey became intentionally more challenging with time, transitioning from description of sources and problems to ordering, explaining, and making decisions, there was a general decline in confidence. This trend can be appreciated from the data in Table 3, which displays the comparisons between items, with, for example, significantly different (lower) confidence scores for *items 1* and *2* compared with *items 5* and *6*, although the median confidence

Table 2. *Student accuracy and confidence levels*

	Graded Accuracy Performance			Self-Reported Student Confidence		
Item	Median accuracy, %	Range of accuracy, %	Mean ± SD accuracy, %	Median confidence, % (*P* Value)	Range of confidence, %	Mean ± SD confidence, %
1	50	0–67	53 ± 18	86 (0.000*)	57–100	86 ± 10
2	33	0–67	40 ± 17	86 (0.000*)	43–100	81 ± 11
3	33	0–67	34 ± 19	71 (0.000*)	43–100	75 ± 13
4				N/A (narrative answers)		
5	100	0–100	74 ± 44	71 (0.530)	29–100	72 ± 16
6	0	0–100	23 ± 42	71 (0.000*)	29–100	71 ± 17

N/A, not applicable. *Median confidence level significantly different from accuracy level (Wilcoxon's signed ranked test).

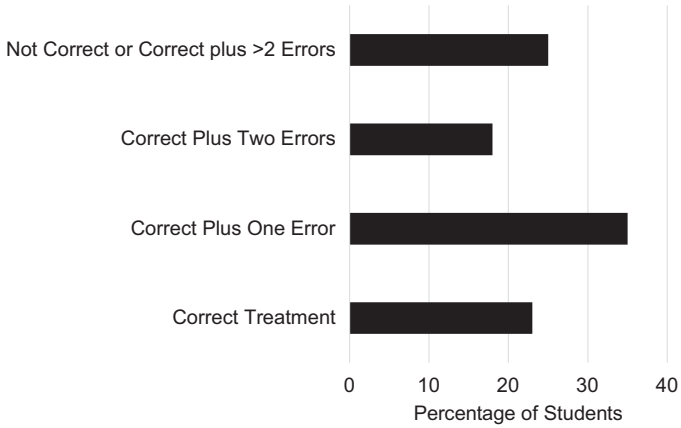

Figure 3. Frequency distribution of medication choices.

level never fell below "Somewhat Confident." Table 2 shows that in every case except for *item 5* (intravenous fluid selection) the level of confidence expressed as a percentage was significantly higher than the actual performance level, suggesting that overconfidence was prevalent.

To further investigate the relationship between student confidence and accuracy, the raw values were standardized by expressing them as *z* scores. These were plotted in a four-zone graph describing high and low levels of student confidence combined with high and low levels of accuracy of student performance (Fig. 4). A first zone situated up and right in the graph identifies students with high accuracy and high confidence, who display appropriate metacognitive tracking of their reasoning and level of performance, A second zone down and left tracks students with low accuracy and low confidence, who also display appropriate metacognitive tracking of their reasoning and level of performance. In contrast, the zone up and left is suggestive of the Dunning–Kruger effect, with students presenting high confidence and low accuracy, with a display of inappropriate metacognitive tracking of their reasoning and level of performance. A final zone situated down and right identifies those students who displayed low confidence and yet had performed with high accuracy. Those students also displayed inappropriate metacognitive tracking of their reasoning and level of performance, Such *z* scores provide a novel way to illuminate student performance

Table 3. *Adjusted P values for pairwise comparisons of confidence levels between items*

	Item No. (confidence)					
Item No. (confidence)	*1*	*2*	*3*	*4**	*5*	*6*
1						
2	0.176					
3	0.000†	0.072				
*4**	N/A	N/A	N/A			
5	0.000†	0.001†	1	N/A		
6	0.000†	0.002†	1	N/A	1	

A Bonferroni adjustment was applied. Values < 0.05 indicate statistically significantly different confidence levels between items. N/A, not applicable. *Correlation with *item 4* cannot be performed because it was narrative answer and not numeric. †Median confidence significantly differs between item pair (Kruskal–Wallis test plus post hoc Mann–Whitney tests with Bonferroni correction).

and support student learning with personalized feedback on zones of confidence-accuracy for improving metacognitive understanding of one's hot cognition and its performative outcomes in the diagnosis learning context of HFPS.

Scientific Misconception Findings

Item 4 was the space for inquiry on students' explanations and possible scientific misconceptions regarding DKA. Most students correctly identified that the patient was in shock, that low mean arterial pressure was reflecting low cardiac output, and that tachycardia was a compensatory mechanism via the baroreceptor reflex. Students demonstrated a similarly high level of understanding about the use of isotonic saline to restore extracellular fluid volume.

The most prevalent misconception found, albeit in only 6 of 80 written responses to survey *item 4*, was the notion that metabolic acidosis was the root cause of all other problems and that correcting the acidosis would be curative. Although it was not strongly represented in *item 4*, it provided an important clue to interpreting the students' responses on *items 5* and *6*. In fact, the analysis of treatment errors in the subsequent items showed that 35 individual students would incorrectly include bicarbonate in the treatment plan, a decision that persisted even after group discussion, with 8 of 20 groups giving a bicarbonate infusion. This strong finding of persistence suggests that the presence of an acid-base disturbance was strongly distracting to overall clinical reasoning and that physiological causal reasoning is not consolidated, giving rise to obstructive misconceptions that persist toward decisions of treatment. In addition, given the clear disparity between fewer students reporting the misconception explicitly and the relevant presence of said scientific misconception in the decision-making during the simulation, it suggests that such a metabolic acidosis misconception may be more pervasive than reported, and that is overall a very significant result of the study.

Faculty Interview Results

Five faculty facilitators were available and eligible for interview out of six; D. M. Harris and J. D. Kibble were not interviewed because they were investigators; one instructor was not available. Considering qualitative research, five interviews out of six subjects identified repeated themes among them. All interviewed faculty agreed that using the MCC tool had improved the overall group process, mainly by slowing student teams down and forcing students to follow through with a methodic, piecemeal reasoning process and confidence calibration. Faculty were mostly neutral when asked if they felt that using the survey had assisted group process by inhibiting dominant students and encouraging more reticent students. One faculty member thought the survey was possibly too intrusive and had inhibited spontaneous group process, whereas the other four faculty acknowledged that some intrusion had occurred.

DISCUSSION

In the past, the authors have noticed that many students are immediately drawn to an abnormal vital sign and may prematurely initiate treatment before considering underlying causes. By asking students to first develop a problem list

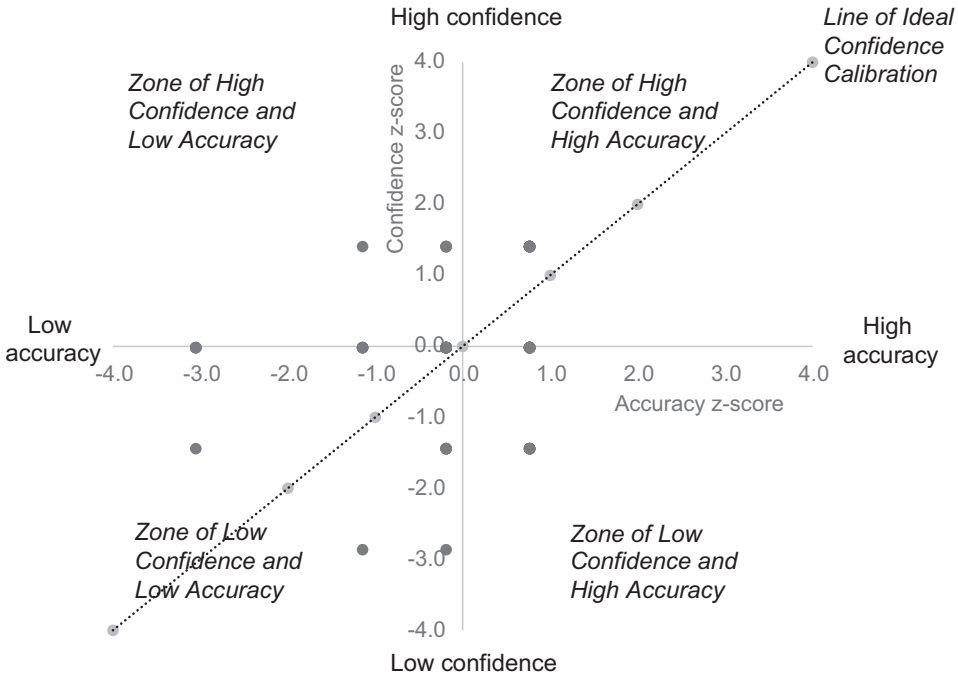

Figure 4. *z* Scores student confidence-accuracy zones. Spearman correlations were as follows: *item 1*, $r = 0.188$ ($P = 0.048$); *item 2*, $r = -0.016$ ($P = 0.445$); *item 3*, $r = 0.081$ ($P = 0.239$); *item 5*, $r = 0.34$ ($P = 0.001$); *item 6*, $r = 0.198$ ($P = 0.039$).

and then to rank problems and make visible their scientific explanations, the goal of the authors was to model a critical thinking process for deliberate practice of diagnostic reasoning and to reveal misconceptions about the underlying physiology.

In relation to the specific objectives of this study, the survey tool developed to scaffold student reasoning toward decision-making during HFPS was deemed successful (*objective 1*). The feasibility of using such a tool in real time was established since all groups completed the required tasks within the 15 min allowed. Faculty facilitators unanimously felt that students' thought processes were improved by using the tool. The study succeeded in assessing how accurately students gathered and applied data by comparing individual student responses against a faculty rubric and generally found a low level of accuracy (*objective 2*). However, by including opportunities for teams to reach negotiated understanding before decisions were made, the final team decisions were better than the average of individual decisions. The use of the protocol with both individual and team tasks was key to obtaining both personal and team-guided cooperative reasoning processes. The MCC tool provided valuable insight into what concepts students understood and what misconceptions or gaps were present (*objective 3*), contributing to a future inventory. Cardiovascular physiology concepts were well understood, but the presence of an acid-base problem was highly distracting. The study successfully gathered confidence calibration data for each item that revealed an overconfidence bias in our novice learners. We discuss further below the importance of metacognition in learning, how the instrument provided an opportunity for gaining metacognitive knowledge, and suggestions for how this could be leveraged in the future (*objective 4*). Furthermore, we discuss below the potential of this tool to provide the basis for reflective practices aimed at repairing any faulty mental physiology models, promoting metacognitive skills, and introducing novice students to the facets of safer medical decision-making (*objective 5*).

It is well known that scientific misconceptions, first emerging from physics research, are common fare in students learning physiology (35, 36). One advantage of using this new tool was to provide faculty with insight into the prevalence of student misconceptions and improve faculty opportunity to develop class plans and curricula, as well as faculty development, as is the case in other disciplines (37). It was gratifying for authors as instructors to see that all groups were correctly able to identify that the patient was in shock but did not have heart failure. Most students correctly connected tachycardia with low arterial pressure through the baroreceptor reflex. Similarly, the concepts of body fluid volumes and NaCl and water distribution were mastered well, which led to the rational use of isotonic saline infusion to increase preload and thereby restore cardiac output. Students had the most trouble with acid-base physiology. The class had recently completed a TBL activity introducing acid-base physiology, and, whereas most students were able to identify the presence of acute metabolic acidosis, putting this knowledge in context proved more difficult. A minority of students thought the acidosis was the root cause of all other problems. Forty-three percent of respondents wanted to add bicarbonate to the intravenous infusion, showing that a classroom discussion about the general principle of treating acid-base disorders by addressing the cause had not transferred into action when students were faced with making a decision. Lack of knowledge and misconceptions are a basic source of cognitive error in medical decision-making (2). The MCC tool provided ample data to objectively describe misconceptions occurring in the HFPS. There are

several potential future uses of the data to help students fix misconceptions that are recorded. For example, the data could be reviewed in class the next day, individual students could receive feedback, or it may be possible for a faculty team to review the data and discuss misconceptions in the debrief session.

The data indicated the presence of overconfidence bias in diagnostic reasoning and medical decision-making, based on the observation that confidence scores were almost uniformly significantly greater than accuracy scores. This miscalibration of beliefs and actual performance is a very common observation in all walks of life, known as the Dunning–Kruger effect (5). The theory postulates that people who are not yet competent in a particular domain have the dual problem that they make the wrong choices and at the same time their lack of competence deprives them of the ability to realize a decision is wrong. This overconfidence problem is another representation of a lack of metacognitive skill in novices and is worse in people who rely on intuition to solve problems than in those who use a slower deliberative analytical approach (38). Faculty interviews suggested that forcing students to slow down (albeit over 15 min of focused inquiry) was a key advantage of using the survey tool, which may promote such deliberative thinking. Indeed, Trowbridge (39) proposed that the simple strategy of finding ways to slow students down is a priority when teaching about avoidance of diagnostic errors. In healthcare, it is an obvious concern that overconfidence bias be rectified before learners are certified for practice without supervision. At present, there is strong momentum behind the adoption of "entrustment" frameworks for the assessment of medical trainees, in which learners are observed over extended periods in different settings until there is sufficient evidence to suggest competence (40). These data offer a powerful opportunity to open a conversation with students about the phenomenon of overconfidence bias in decision-making. The authors' use of *z* scores to standardize absolute values of accuracy and confidence yielded a side benefit of identifying individuals who have a clear divergence in their relative confidence and accuracy. In future studies the authors intend to further develop a metric considering the graph developed with the four zones of performance, to inform tailored feedback for individuals to support their learning with hot cognition, within a growth mindset.

A final insight from the data was that teams outperformed individuals. For example, survey data showed that <25% of individuals selected the correct treatment alone but 50% of teams were correct after a short consensus-building discussion. In addition, groups did better at selecting the correct intravenous fluid infusion and had a lower rate of overtreatment. However, the urge to give bicarbonate to treat the acidosis persisted at a similar level in groups and individuals (40% vs. 43%). Team-based learning literature consistently shows that teams outperform individuals (41), which the authors saw with just a minute or two of discussion. Before our first simulation, the students have a 2-h introduction to Team STEPPS, which is a training program created by the Agency for Healthcare Research and Quality and the Department of Defense that aims to improve collaboration and communication within teams and has been extensively used in healthcare to enhance patient safety (42). By the fourth simulation in the physiology series, when our study was performed, the students were proficient at assigning roles and using closed-loop communication. It was gratifying to see objective evidence that brief team interaction improved performance, and this is another potential area where data obtained from our tool could enrich debriefing about the importance of teamwork.

Implication for Stakeholders

Both medical students and patients are stakeholders in this research program: medical students may sharpen their self-awareness, calibrate their confidence aligned with patient outcomes, and improve learning outcomes with simulations as diagnosticians. Systematically improving the reflexive dimension of diagnosis with simulations has the potential to improve patient safety and to reduce the effects of misdiagnosis in the future distributed in health care, thereby ultimately benefiting patients. The methodology developed can support better tracking of learning for better performance of diagnosis with hot cognition.

Concluding Remarks and Future Work

In this study, the authors were able to demonstrate the feasibility of using a self-assessment tool during HFPS sessions. The tool provided a scaffolding opportunity for individuals and teams to follow a deliberate practice of diagnostic process in medical decision-making. The instrument makes student misconceptions salient and allows students to calibrate confidence as a means to practice the metacognitive skills needed to develop expertise. The tool can be applied to other high-fidelity patient simulations. The data identified specific issues for debriefing, such as low accuracy levels and overconfidence. The project was also an example of improved integration between physiology, critical reasoning (philosophy of medicine/medical diagnosis), and patient safety curricula.

Future studies will attempt to evaluate self-regulated learning attributes and relate them to student attitudes toward the use of the MCC tool. A follow-up study in the works, developed with the feedback of the reviewers to this first study, will further provide individualized information about the discrepancy between accuracy and confidence for students as a unified score, which is the next study goal, after achieving the validation of the MCC tool in the present study. The future debriefing session plan may also include a brief section to emphasize the importance of metacognition and the process of participating in disagreement in team collaborative reasoning, which may reveal other aspects and criteria used and its path for analytic reduction (43).

ACKNOWLEDGMENTS

We acknowledge the invaluable contributions of Bee Ben Khallouq, for early statistical support and performing SPSS statistical analysis; Bill Barker, for assistance in deploying the online survey tool at the University of Central Florida College of Medicine; and the moderators and audience comments to a conference presentation at the North American Conference at the Society of Medical Decision-Making (SMDM), Portland, OR (2019), and at the Annual Conference of the Society to Improve Diagnosis in Medicine (SIDM, 2021).

GRANTS

We acknowledge the internal competitive research award from the Medical Education Department at the University of Central Florida.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

L.S.G., D.M.H., S.M.F., and J.D.K. conceived and designed research; L.S.G., D.M.H., and J.D.K. performed experiments; L.S.G., D.M.H., S.M.F., M.R., and J.D.K. analyzed data; L.S.G., D.M.H., S.M.F., M.R., and J.D.K. interpreted results of experiments; L.S.G., M.R., and J.D.K. prepared figures; L.S.G. and J.D.K. drafted manuscript; L.S.G., D.M.H., S.M.F., M.R., and J.D.K. edited and revised manuscript; L.S.G., D.M.H., S.M.F., M.R., and J.D.K. approved final version of manuscript.

REFERENCES

1. **Reason J.** *Human Error*. Cambridge: Cambridge University Press, 1990. doi:10.1017/CBO9781139062367.
2. **Croskerry P.** Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Acad Emerg Med* 9: 1184–1204, 2002. doi:10.1111/j.1553-2712.2002.tb01574.x.
3. **Papa F**, **Harasym PH.** Medical curriculum reform in North America, 1765 to the present: a cognitive science perspective. *Acad Med* 74: 154–164, 1999. doi:10.1097/00001888-199902000-00015.
4. **Wachter RM**, **Shojania KG**, **Saint S**, **Markowitz AJ**, **Smith M.** Learning from our mistakes: quality grand rounds, a new case-based series on medical errors and patient safety. *Ann Intern Med* 136: 850–852, 2002. doi:10.7326/0003-4819-136-11-200206040-00015.
5. **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* 77: 1121–1134, 1999. doi:10.1037/0022-3514.77.6.1121.
6. **Altabbaa G**, **Raven AD**, **Laberge J.** A simulation-based approach to training in heuristic clinical decision-making. *Diagnosis (Berl)* 6: 91–99, 2019. doi:10.1515/dx-2018-0084.
7. **Halloun I**, **Hestenes D.** The initial knowledge state of college physics students. *Am J Phys* 53: 1043–1055, 1985. doi:10.1119/1.14030.
8. **Michael JA.** Students' misconceptions about perceived physiological responses. *Am J Physiol* 274: S90–S98, 1998. doi:10.1152/advances.1998.274.6.S90.
9. **Fiore S**, **Vogel-Walcutt J.** Making metacognition explicit: developing a theoretical foundation of metacognitive prompting during scenario-based training. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, 2010.
10. **Flavell JH.** Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *Am Psychol* 34: 906–911, 1979. doi:10.1037/0003-066X.34.10.906.
11. **Zimmerman BJ.** Self-regulation involves more than metacognition: a social cognitive perspective. *Educ Psychol* 30: 217–221, 1995. doi:10.1207/s15326985ep3004_8.
12. **Zimmerman BJ.** Self-regulated learning and academic achievement: an overview. *Educ Psychol* 25: 3–17, 1990. doi:10.1207/s15326985ep2501_2.
13. **White CB**, **Gruppen LD**, **Fantone JC.** Self-regulated learning in medical education. In: *Understanding Medical Education: Evidence, Theory and Practice*, edited by Swanwick T. Chichester, UK: Wiley Blackwell, 2014, p. 201–211.
14. **Kohn L**, **Corrigan J**, **Donaldson M** (editors). *To Err Is Human: Building a Safer Health System*. Washington, DC: Committee on Quality of HealthCare in America, Institute of Medicine, National Academy Press, 2000.
15. **World Health Organization & World Alliance for Patient Safety Research Priority Setting Working Group.** Summary of the evidence on patient safety: implications for research (Online), 2008. Geneva: World Health Organization. http://www.who.int/iris/handle/10665/43874 [2018 Apr 4].
16. **Jha AK**, **Larizgoitia I**, **Audera-Lopez C**, **Prasopa-Plaizier N**, **Waters H**, **Bates DW.** The global burden of unsafe medical care: analytic modelling of observational studies. *BMJ Qual Saf* 22: 809–815, 2013. doi:10.1136/bmjqs-2012-001748.
17. **Makary MA**, **Daniel M.** Medical error—the third leading cause of death in the US. *BMJ* 353: i2139, 2016. doi:10.1136/bmj.i2139.
18. **Liaison Committee for Medical Education.** Functions and Structure of a Medical School: Standards for Accreditation of Medical Education Programs Leading to the MD Degree (Online). http://lcme.org/publications/ [2018 Apr 4].
19. **Rajkomar A**, **Dhaliwal G.** Improving diagnostic reasoning to improve patient safety. *Perm J* 15: 68–73, 2011. doi:10.7812/tpp/11-098.
20. **Bereiter C**, **Scardamalia M.** *Surpassing Ourselves: An Inquiry into the Nature and Implications of Expertise*. Chicago, IL: Open Court, 1993.
21. **Stark M**, **Fins J.** The ethical imperative to think about thinking: diagnostics metacognition, and medical professionalism. *Camb Q Healthc Ethics* 23: 386–396, 2014. doi:10.1017/S0963180114000061.
22. **Kassirer JP.** Diagnostic reasoning. *Ann Intern Med* 110: 893–900, 1989. doi:10.7326/0003-4819-110-11-893.
23. **Kassirer JP**, **Kopelman R.** *Learning Clinical Reasoning*. Baltimore, MD: Williams and Wilkins, 1991.
24. **Croskerry P.** The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 78: 775–780, 2003. doi:10.1097/00001888-200308000-00003.
25. **Croskerry P.** Commentary: the affective imperative: coming to terms with our emotions. *Acad Emerg Med* 14: 184–186, 2007. doi:10.1197/j.aem.2006.08.016.
26. **Meyer AN**, **Payne VL**, **Meeks DW**, **Rao R**, **Singh H.** Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med* 173: 1952–1958, 2013. doi:10.1001/jamainternmed.2013.10081.
27. **Cassam Q.** Diagnostic error, overconfidence and self-knowledge. *Palgrave Commun* 3: 17025, 2017. doi:10.1057/palcomms.2017.25.
28. **Garbayo L.** *A Naturalized Bridge to Virtue Epistemology: Contributions of Experimentally Shifted Contexts of Justification to the Virtuous Self-Articulation of de se and de re Beliefs*. Presidential Address, Southwest Philosophical Studies, 2017.
29. **Richie M**, **Josephson SA.** Quantifying heuristic bias: anchoring, availability, and representativeness. *Teach Learn Med* 30: 67–75, 2018. doi:10.1080/10401334.2017.1332631.
30. **Harris JR**, **Helyer RJ**, **Lloyd E.** Using high-fidelity human patient simulators to teach physiology. *Med Educ* 45: 1159–1160, 2011. doi:10.1111/j.1365-2923.2011.04105.x.
31. **Harris DM**, **Ryan K**, **Rabuck C.** Using a high-fidelity patient simulator with first-year medical students to facilitate learning of cardiovascular function curves. *Adv Physiol Educ* 36: 213–219, 2012. doi:10.1152/advan.00058.2012.
32. **Harris DM**, **Bellew C**, **Cheng ZJ**, **Cendán JC**, **Kibble JD.** High-fidelity patient simulators to expose undergraduate students to the clinical relevance of physiology concepts. *Adv Physiol Educ* 38: 372–375, 2014. doi:10.1152/advan.00063.2014.
33. **Helyer R**, **Dickens P.** Progress in the utilization of high-fidelity simulation in basic science education. *Adv Physiol Educ* 40: 143–144, 2016. doi:10.1152/advan.00020.2016.
34. **Nguyen K**, **Ben Khallouq B**, **Schuster A**, **Beevers C**, **Dil N**, **Kay D**, **Kibble JD**, **Harris DM.** Developing a tool for observing group critical thinking skills in first-year medical students: a pilot study using physiology-based, high-fidelity patient simulations. *Adv Physiol Educ* 41: 604–611, 2017. doi:10.1152/advan.00126.2017.
35. **Michael J.** Misconceptions—what students think they know. *Adv Physiol Educ* 26: 5–6, 2002. doi:10.1152/advan.00047.2001.
36. **Modell H**, **Michael J**, **Wenderoth MP.** Helping the learner to learn: the role of uncovering misconceptions. *Am Biol Teach* 67: 20–26, 2005. doi:10.2307/4451776.
37. **Garik P**, **Garbayo L**, **Benétreau-Dupin Y**, **Winrich C**, **Duffy A**, **Gross N**, **Jariwala M.** Teaching the Conceptual History of Physics to Physics Teachers. *Sci Educ* 24: 387–408, 2015. doi:10.1007/s11191-014-9731-9.

38. **Mata A**, **Ferreira MB**, **Sherman SJ.** The metacognitive advantage of deliberative thinkers: a dual-process perspective on overconfidence. *J Pers Soc Psychol* 105: 353–373, 2013. doi:10.1037/a0033640.
39. **Trowbridge RL.** Twelve tips for teaching avoidance of diagnostic errors. *Med Teach* 30: 496–500, 2008. doi:10.1080/01421590801965137.
40. **Association of American Medical Colleges.** Core Entrustable Professional Activities for Entering Residency. Curriculum Developers Guide Association of American Medical Colleges (Online). https://store.aamc.org/downloadable/download/sample/sample_id/63/%20 [2018 Apr 16].
41. **Kibble JD**, **Bellew C**, **Asmar A**, **Barkley L.** Team-based learning in large enrollment classes. *Adv Physiol Educ* 40: 435–442, 2016. doi:10.1152/advan.00095.2016.
42. **Agency for Healthcare Research and Quality.** TeamSTEPPS: Team Strategies & Tools to Enhance Performance & Patient Safety (Online). https://www.ahrq.gov/teamstepps/index.html [2018 Apr 18].
43. **Garbayo LS.** Epistemic considerations on expert disagreement, normative justification, and inconsistency regarding multi-criteria decision making. In: *Constraint Programming and Decision Making, Studies in Computational Intelligence*, edited by Ceberio M, Kreinovich V. Cham, Switzerland: Springer, 2014, vol. 539.