McRobert, AP, Mercer, SJ, Raw, D, Goulding, J and Williams, MA

Effect of expertise on diagnosis accuracy, non-technical skills and thought processes during simulated high-fidelity anaesthetist scenarios

Article

The effect of expertise on diagnosis accuracy, non-technical skills and thought processes during simulated high-fidelity anaesthetist scenarios

Allistair P McRobert

Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, Merseyside, UK

Simon J Mercer, David Raw & Jeff Goulding

Centre for Simulation and Patient Safety, Aintree University Hospitals NHS Foundation Trust, Merseyside, UK

A. Mark Williams

Department of Health, Kinesiology, and Recreation, The University of Utah, Salt Lake City, Utah, USA

Please address correspondence to:

Allistair McRobert

Liverpool John Moores University

Tom Reilly Building

Byrom Street

Liverpool, L3 3AF

United Kingdom

Tel - +44 151 231 4320

Email: a.p.mcrobert@ljmu.ac.uk

Keywords: Decision-making, Anaesthesia, Verbal reports, Medical simulation.

Word count (excluding title page, abstract, references, figures and tables): 3472

The effect of expertise on diagnosis accuracy, non-technical skills and thought processes during simulated high-fidelity anaesthetist scenarios

RUNNING HEAD: Decision-making in anaesthetists

ABSTRACT

We examined the thought processes and performance of anaesthetists during simulated environments. Verbal reports of thinking and the Anaesthetists Non-Technical Skills (ANTS) were recorded to examine cognitive processes, non-technical behaviours and diagnosis accuracy during fully immersive, high-fidelity medical scenarios. Skilled (n = 6) and less skilled (n = 9) anaesthetists were instructed to respond to medical scenarios experienced in theatre. Skilled participants demonstrated higher diagnosis accuracy and ANTS scores compared to less skilled participants. Furthermore, skilled participants engaged in deeper thinking and verbalised more evaluation, prediction and deep planning statements. The ability to employ an effective cognitive processing strategy, more efficient non-technical behaviours and superior diagnosis is associated with superior performance in skilled participants.

WHAT THIS PAPER ADDS

- High-fidelity medical simulation is an ideal tool to examine differences in knowledge and performance of anaesthetists because scenarios can be standardised and replicated.
- Diagnosis accuracy, ANTS and concurrent verbal reports can be used as performance measures during simulation to compare anaesthetists of different levels of expertise.
- Concurrent verbal reports are a technique that offers important inferences about thought processes.
- Skilled anaesthetists are characterised by their ability to make more evaluation statements about the current situation, predict future events and outcomes, and forward plan actions.
- Simulation training may present an ideal context for developing higher levels of expertise and the associated cognitive and technical skills

INTRODUCTION

Anaesthesia in the modern operating environment requires a range of attributes and skills to ensure patient care and safety. Furthermore, each patient poses a unique set of challenges and constraints that make it difficult to directly compare performance levels between and across individuals. For example, some patients present with complex and difficult problems, compared to more straightforward diagnosis and treatment pathways. The challenge for educators and researchers is to capture and examine technical skills, non-technical skills, thought processes and performance using standardised (i.e., valid, objective and reliable), representative tasks.(1) Recent developments in high-fidelity medical simulators provide more realistic, fully immersive environments for examining cognitive (decision-making), behavioural (technical and non-technical skills) and performance (diagnosis accuracy) differences between participants.(2-4) Furthermore, simulators allow control and replication of scenarios while maintaining ecological validity. In this study, the expert performance approach was used to develop anaesthetist, high-fidelity medical scenarios to examine differences in performance, cognition and behaviour.(5)

The expert performance approach has been used in sport, music, education and medicine to examine expertise.(6-11) First, experts are observed *in situ* so that the essential skills and competencies underpinning expertise can be captured. Based on these observations, simulated representative laboratory or field-based task(s) are created so that skill based differences in performance can be reliably and objectively measured under more controlled conditions. Second, to promote understanding of expertise, process-tracing measures such as think-aloud verbal reports and /or eye-movement recordings to examine perceptual and cognitive mechanisms that support performance. Finally, development of expert skills and mechanisms are traced to identify how and when they were acquired, while considering the practical implications for guiding practice and instruction.

In medicine, high-fidelity simulators are used as a valid and reliable tool for assessment of competency and training because scenarios can be standardised, repeated and require the participant to interact with the patient to provide care.(3,12-14) Performance is assessed using diagnosis accuracy, response time, rating scales, non-technical behaviours (e.g., Anaesthesia Non-Technical Skill, ANTS) and technical skills with little focus on examining the thought processes and decision-making (i.e., cognition) simultaneously.(3,4,15,16)

Cognitions are recorded and analysed using think-aloud verbal reports and protocol analysis, providing a quantifiable measure that represents the thought processes underpinning participant actions.(17) Fox et al. (18) conducted a meta-analysis of 94 studies and found no difference in task performance when comparing think-aloud verbal reports to a matched condition without verbal reports. Therefore, they suggested that the cognitive processes mediating task performance do not alter, allowing researchers to make inferences about performance based on verbalised cognitions (see Ericsson and Simon (17) for a more detailed theoretical rationale). Furthermore, Whtye et al. (19) reported that concurrent think-aloud verbal reports provide a more complete representations of nurse’s cognitions when admitting a patient suffering from an acute exacerbation of congestive heart failure to a simulated intensive care unit.

Think-aloud protocols have been used by researchers to examine cognitions and make inferences on performance in law enforcement, sport and medicine (20-22). For example, Cormier et al. (2) examined cognitions and performance of low and high performing student nurses during high-fidelity medical simulations. They measured verbal statements coded as ‘observations’ (e.g., context, patient and monitor) or ‘actions’ (e.g., based on orders or patient condition) and physiological data of the patient’s respiratory status. High performing student nurses were better at observing relevant cues suggesting patient deterioration and forward plan responses that positively altered the patient’s physiological trajectory. In comparison, low performing student nurses verbalised more irrelevant cues and failed to perform action directly related to the patient’s condition.

In the current study, we use high-fidelity scenarios so that the cognitive processes associated with expertise can be fully understood. Diagnosis accuracy and non-technical behaviours of skilled and less skilled anaesthetists were examined while they provided think-aloud concurrent verbal reports during the scenarios. It was hypothesised that skilled anaesthetists would demonstrate superior diagnosis accuracy compared to less skilled anaesthetists. Second, the Anaesthetist Non-Technical Skills (ANTS) was used as a valid and reliable tool for assessing non-technical behaviours.(23) It was hypothesised that skilled participants would score a higher global rating compared to less skilled anaesthetists. Finally, skilled anaesthetists’ superior performance would be supported by more evaluation, prediction and deep planning statements when compared to less skilled counterparts.

METHOD

Participants

Nine less skilled (mean age = 29.4 years, SD = 2.3) and 6 skilled (mean age = 37.2, SD = 4.5) anaesthetists were recruited. Less skilled participants were pre-FRCA trainees in their first 2 years of anaesthesia with a mean of 4.2 years (SD = 1.1) experience as a doctor. Skilled participants were at consultant or experience staff grade level with a mean of 13.3 years (SD = 3.1) experience. The study was approved by the lead institute's Research Ethics Committee and carried out under its ethical guidelines.

Simulated Test Environment (STE)

A simulated task environment (STE) was developed. The STE consisted of a high-fidelity adult METI human patient simulator (CAE Healthcare, Florida, USA) and fully equipped operating theatre, that included an anaesthetic machine, monitoring familiar to the anaesthetists, realistic functioning medical gases, intravenous pumps (IV), a real crash cart with defibrillator, medication and medical supplies. The STE allowed real-life scenarios to be dynamically presented to participants with the onset and choice of patient response under full experimenter control. METI human patient simulator responded to treatment in a realistic manner exhibiting applicable changes in cardiovascular, pulmonary, physiological, pharmacological, metabolic, genitourinary and neurological states.

Scenarios

Three tightly scripted and highly standardised simulator scenarios were developed. Each scenario lasted approximately 15 minutes, involved an emergency developing during the course of an operation and required the anaesthetist to intervene and administer treatment.

In scenario one (anaphylaxis), the patient is a 68-year-old man, non-smoker, no medication, allergies or airway problems, drinks alcohol occasionally, no previous history of general anaesthetic (GA) and suffers from active reflux so requires an endotracheal tube (ETT). In addition, he has requested a GA, as he 'didn't like the sound of a spinal'. A confederate Surgeon, Scrub nurse and Operating Department Practitioner (ODP) are on hand to provide assistance when requested. The participant enters the room and is guided over to the anaesthetic machine were they receive a handover from a colleague who is being relieved for lunch. "The patient has just been anaesthetised for a left knee replacement under GA. Skin is prepared for surgery and the surgeon and scrub nurse are ready. Patient-controlled analgesia (PCA) is planned for post-operative pain relief". The Surgeon requests 1.5g of Cefuroxime prior to tourniquet inflation. Administration is rapidly followed by cardiovascular system (CVS) collapse and bronchospasm. The patient will require 100% oxygen, termination of the suspected agent, administration of adrenaline and referral to the intensive treatment unit (ITU).

In the second scenario (Malignant Hyperpyrexia), the patient is a 69-year-old man, non-smoker, no medication, allergies or airway problems, drinks alcohol occasionally, no previous history of GA or anaesthetic problems and has active reflux so was intubated for the procedure. A confederate Surgeon and Scrub nurse are busy operating with an ODP on hand to provide assistance if requested. The participant receives a handover in the theatre from a colleague who is being relieved for a meeting. "The patient was anaesthetised 20 minutes ago for removal of metalwork from the left ankle. PCA is planned for post-operative pain relief". Malignant hyperpyrexia gradually starts once the participant is settled. The participant will be required to stop the volatile, hyperventilate the patient to reduce end tidal CO_2, give patient 100% oxygen, change breathing circuit/anaesthetic machine, administer Dantrolene and actively cool patient.

In the third scenario (Supraventricular Tachycardia post induction), the patient is a 59-year-old man, who is fit and well, no previous GA problems, no allergies or airway problems, ex-smoker and drinks alcohol occasionally. The patent takes an oral dose of Lansporazola (30mg) and has well-controlled reflux. A confederate ODP is on hand to provide assistance when requested. The participant receives drugs ready drawn up in a kidney bowl and a handover, before they enter the theatre. “The patient is due to have an elective laparoscopic cholecystectomy for gallstones. PCA for post-operative pain relief is planned if an open procedure is required”. The participant will induce anaesthesia. Once intubated, the patient develops narrow complex tachycardia and his cardiovascular function becomes compromised. The participant will be required to give DC shock cardio version to the unstable patient.

Procedure

Participants completed a biographical (i.e., age, number of years training and details of training activity) information sheet and provided informed consent. Participants were instructed in a think-aloud verbal reporting method by Ericsson and Simon (17) that has been used previously in sport, law enforcement and medicine .(19,20,24) Participants practised giving verbal reports while solving non-clinical alphanumerical problems until the criterion for providing concurrent verbal reports were met. For example, if the researcher asked, “What the third letter after A is?”, you might respond “D, I counted three letters forward” but your actual thoughts might have been “A, B, C, D”. The former is an example of a summary, whereas the latter are your actual thoughts. During the training, participants accuracy on alphanumerical tasks were not recorded, as the objective was to provide feedback so that they could understand the difference between verbalising a summary of their thoughts compared to think-aloud statements about their thoughts. Following their verbal report training, participants were familiarised with the STE using a structured protocol and then completed a simple practice trial while providing a concurrent verbal report. The practice trail required them to check the equipment and conduct an objective assessment of the patient. Concurrent verbal reports were selected as they provide a more complete cognitive representation of performance within the STE, compared to retrospective reports.(19) Following their practice trial, participants were given an opportunity to ask further questions prior to the experimental scenarios and further feedback were provided.

Participants were fitted with a lapel microphone and radio transmitter and in conjunction with 2-angle video cameras, visual and audio observations were captured. The orders of scenarios presented to each participant were counterbalanced across the participants to reduce order and learning effects. Participants interacted with the patient and confederate staff involved. In order to standardise the scenarios, confederates were equipped with a two-way radio transceiver so that the experimenters could control the information provided to the participant. If the participant remained silent for an extended period of time, the confederate ODP was instructed to remind them to concurrently verbalise their thoughts using the prompt "think-aloud". Each scenario terminated when the 15 minutes elapsed.

Data analysis

Diagnosis data

A diagnosis response was marked as correct and awarded a score of '1' if the participants verbalised the correct scenario diagnosis. If participants failed to provide or provided an incorrect diagnosis response, a score of '0' was recorded for that scenario. Diagnosis accuracy was defined as the total number of correct responses (i.e., frequency score) divided by the total number of scenarios (N = 3) and multiplied by 100 to create a percentage score. An independent samples *t*-test was conducted to compare diagnosis accuracy between skilled and less skilled participants.

Anaesthetists' Non-Technical Skills (ANTS) data

The Anaesthetists' Non-Technical Skills (ANTS) system was used to assess behavioural markers during the medical scenarios. The ANTS system's validity and reliability has previously been established when assessing non-technical behaviours.(23) A consultant anaesthetist trained and qualified in the use of the ANTS system reviewed videotapes of the participants' non-technical skills during the scenarios. ANTS describes the main observable non-technical skills associated with good anaesthetic practice.

The ANTS system comprises of four skill categories and beneath these are fifteen skill elements (see Table 1). Each element has a definition and examples of good and poor behaviours. The reviewer rated the 15 skill elements based on a four-point scale (i.e., 4 = good; 3 = acceptable, 2 = marginal; 1- poor; 0 – not observed). The rating score for each element was added together to calculate the participant's global performance score for each scenario. A mean ANTS score was calculated based on the participant's global scores from the three scenarios. A higher score on the ANTS represented better non-technical behaviours during the scenarios. An independent samples *t*-test was conducted to compare differences in ANTS scores between skilled and less skilled participants.

Table 1: ANTS four skill categories and fifteen skill elements

Category	**Elements**
Task Management	Planning and preparing
	Prioritising
	Providing and maintaining standards
	Identifying and utilising resources
Team Working	Co-ordinating activities with team members
	Exchanging information
	Using authority and assertiveness
	Assessing capabilities
	Supporting others
Situation Awareness	Gathering information
	Recognising and understanding
	Anticipating
Decision Making	Identifying options
	Balancing risks and selecting options
	Re-evaluating

Verbal Report data

Scenario videos were captured and analysed using Studiocode, version 3.5 (Sportstec Ltd, Australia), an analytical tool that allows the researcher to mark and code segments of video and audio into categories. Following the procedures outlined in Ericsson and Simon (17) verbal reports were initially segmented using natural speech and syntactical markers.

Verbal statements were coded based on categories developed and previously used by McRobert et al. (22) to monitor thought processes of participants during emergency medicine simulated scenarios. To ensure logical validity and remove the individual bias of the researcher a panel of anaesthetists (N = 3), medical educators (N

= 2) and researchers (N = 2) reviewed the sample of statements from the simulated scenarios that had been assigned to the coding categories. (25,26) The panel provided written feedback and agreed that the categories and operational definitions were appropriate for analysing concurrent verbal statements.

Table 2: Verbal report statement categories and definitions

Category	Definition
Monitor	Statements representing information that was present or previously present in the current environment.
Evaluation	Some form of positive or negative assessment of a prior action, event or statement.
Prediction	Statements about what could, would and should occur next in the environment.
Deep Planning	Statements are about future actions and options in a future situation.

The primary experimenter analysed all participant data and assigned each concurrent verbal statement to the monitor, evaluation, prediction or deep planning category (Table 2). Fifty verbal statements were selected at random to establish objectivity and reliability using the inter- and intra-observer percentage agreement formulas.(27) The primary experimenter coded the 50 statements on four separate occasions (i.e., prior to coding, after 6, 12 and 18 participants) with an intra-observer agreement range from 95% to 98%. Inter-observer agreement was assessed prior to coding (92%) the statements, and after 12 participants (93%) by an independent experimenter.

Two-way analysis of variance (ANOVA) was conducted with group (less skilled/skilled) as the between participant factor and verbal statement type (monitor/evaluation/prediction/deep planning) as the within-participant factors. Partial eta squared ($\eta p2$) values are provided as a measure of effect size. When making comparisons between two means, Cohen's *d* measures are reported. Posthoc Bonferroni corrected pairwise comparisons are reported as follow-ups where appropriate.

RESULTS

Diagnosis data

An independent samples *t*-test was conducted to compare differences in diagnosis accuracy between skilled and less skilled participants. Levene's test for homogeneity of variance was violated ($p < .05$); therefore the test statistic for equal variance not assumed was reported. There was a significant difference in diagnosis accuracy scores for less skilled and skilled participants, $t13 = -2.688$, $p = .02$. Skilled participants (M = 94%, SD = 14) diagnosis accuracy scores were significantly higher than less skilled participants (M = 63%, SD = 31).

Anaesthetists' Non-Technical Skills (ANTS) data

The independent samples *t*-test was conducted to compare ANTS score differences between skilled and less skilled participants. There was a significant difference in ANTS scores between skilled and less skilled participants, $t13 = -3.215$, $p = .02$. Skilled participants (M = 22, SD = 4) ANTS scores were significantly higher than less skilled participants (M = 13, SD = 6).

Verbal Report data

ANOVA showed a significant Group × Verbal Statement Type interaction (F3,39 = 20.33, $p < .001$, $\eta p^2 = .52$). As predicted, less skilled participants made more monitoring statements ($d = 1.19$) compared to their skilled counterparts (see Figure 1). In contrast, skilled participants made more evaluation ($d = 2.2$) and deep planning ($d = 1.76$) statements compared to less skilled participants. Skilled participants did make more prediction statements compared to less skilled participants, however this was only a medium effect size ($d = 0.52$). The Cohen's *d* effect size score quantifies the magnitude of the difference between the skilled and less skilled groups. There is a larger difference between the skilled and less skilled groups for evaluation and deep planning, compared to prediction. It is worth noting that skilled and less skilled standard deviations for the monitor, evaluation and deep planning are relative to the mean. In contrast, the standard deviation for prediction is large relative to the mean score. The large variance could be due to the low number of prediction statements compared to the other categories.

Significant main effects for group (F3,39 = 10.83, $p < .05$, $\eta p^2 = .45$) and verbal statement type (F3,39 = 67.12, $p < .001$, $\eta p^2 = .83$) were observed. Skilled participants (M = 12, SD = 8) made more verbal statements, compared to less skilled participants (M = 10, SD = 5). Pairwise comparisons demonstrated that all participants made more evaluation and deep planning than monitor and prediction statements ($p < .05$). Second, a higher number of monitor statements were coded compared to prediction statements ($p < .05$).

INSERT FIGURE 1 HERE.

DISCUSSION

We examined the cognitive processes, diagnosis accuracy and non-technical behaviours in anaesthetists during simulated medical scenarios. Skilled and less skilled participants were required to diagnose the patient's condition and engage in effective treatment during fully immersive, high-fidelity, medical scenarios. Verbal reports were collected concurrently during the simulation as an index of thought processes. In addition, the Anaesthetists' Non-Technical Skills (ANTS) were assessed retrospectively to rate the effectiveness of the participant's behaviours. It was hypothesised that skilled participants would demonstrate superior diagnosis accuracy and a higher global rating score on the ANTS, compared to less skilled participants. Moreover, skilled participants' would be characterised by a higher total number of statements, specifically more evaluation, prediction and deep planning statements.

As predicted, skilled participants outperformed less skilled participants on diagnosis accuracy. Findings support differences in skilled and less skilled performance reported previously by researchers in medicine and other domains such as aviation, law enforcement and sport using dynamically evolving simulated tasks.(3,20,28,29) In addition, skilled participants had higher ANTS scores suggesting that as a group their task management, team working, situation awareness and decision-making skills when managing the scenarios are better.

Think-aloud concurrent verbal reports and protocol analysis was conducted to examine if there were any differences between less skilled and skilled anaesthetists thought processes. Skilled participants made more verbal statements in total, compared to less skilled participants. In addition, skilled participants' made more evaluation, prediction and deep planning and fewer monitoring statements, compared to less skilled participants'. Moreover, these differences are larger for evaluation and deep planning statements between the groups compared to prediction. Skilled participants appear to be characterised by their ability to make more evaluation statements about the current situation, predict future events and outcomes, and forward plan actions. In addition, the verbal report data in this study support previous work demonstrating skilled participants thought processing during dynamically evolving tasks in sport and law enforcement.(20,21)

To provide further context, monitoring statements often refer to an observation of oxygen saturation, core temperature or actions such as administering oxygen, adrenaline and reducing end tidal CO2 that have taken place during the scenario. Monitoring statements are just observations, whereas evaluation statements involve an assessment of the observed information or action. Evaluation statements would involve the observation of oxygen saturation or core temperature followed by an assessment of the value compared to the norm value expected. Prediction statements are assumptions or forecasts about what might or should happen next based on evaluation of observed information (e.g., change in patient vitals) or most recent action (e.g., patient intervention). Deep planning statements provide information on future actions that the participant will to undertake in order to mange potential future situations that might occur. This action plan profile is often based on a primary diagnosis and differential diagnosis for so that they can plan a series of future actions to deal with anaphylaxis (e.g., provide 100% oxygen, terminate suspected agent, administer adrenalin and referral to eth intensive care unit), malignant hyperpyrexia (e.g., stop the volatile, hyperventilate the patient to reduce end tidal CO2, give 100% oxygen, change breathing circuit/anaesthetic machine, administer Dantrolene and actively cool) and supraventricular tachycardia (e.g., give DC shock cardio version).

The current findings have implications for those in medicine that have used the recent developments in high-fidelity simulation for the training and assessment of competency.(2-4) Our data provides construct validity for the use of high-fidelity medical simulation as a vehicle for assessment of skill competency. Moreover, such simulations may provide an ideal vehicle for training because it presents a safe environment, where scenarios can be replicated so that the examiner is confident that each individual can experience and be evaluated on the same factors. In sum, it presents a valid, reliable and objective environment for testing and training expertise in medicine. Furthermore, simulators can conduct scenarios that occur on a rare basis. This ability to present novel or rare scenarios will allow participants to develop domain-specific knowledge that can facilitate diagnosis and patient care if the scenario was to occur during their normal working hours. However, it is worth noting that the size, expense and staff support required to manage a simulation environment are often a barrier to widespread use, particular outside of large hospitals and major cities. Finally, the recording of verbal reports provides an opportunity to design educational programmes that specifically target clinical reasoning based on an expert-model approach (30,31). Immersive simulation techniques often involve uninterrupted real-time cases, followed by an individual or group debriefing. Introducing think-aloud verbal reporting during a scenario, allows participants to retrospectively reflect on their thought processes in conjunction with the educator to identify knowledge gaps or reframe clinical reasoning (i.e., cognitive processing).

In conclusion, we examined the cognitive processes and non-technical skills used by skilled and less skilled anaesthetists during high-fidelity medical scenarios. Skilled participants demonstrated superior diagnosis accuracy, better non-technical behaviours and verbalised more evaluation, prediction and deep planning statements, and, compared to less skilled participants. In comparison to less skilled participants, skilled participants were characterised by the ability to evaluate the current situation and engage in prediction of future events and to forward plan actions rather than thinking about immediately available information. Our findings suggest that simulations present a valid method for quantifying and exploring expertise in medical contexts, whereas at a practical level the significant potential to use such environments for training health care practitioners is highlighted.

Contributions

MAW, DR and APM were responsible for the conception and design of the current study. Second, DR, SM, JG and APM collected the data and engaged in many hours of subsequent analysis and interpretation of this data. Third, ARM was responsible for the initial draft of the article, revisions and approval of the final version submitted. Finally, MAW, DR and SM were responsible for critically revising drafts for intellectual content and quality of submission.

Funding

There are no funders to report for this submission.

Competing interests

There are no competing interests.

REFERENCES

1. Ericsson KA, Ward P. Capturing the Naturally Occurring Superior Performance of Experts in the Laboratory: Toward a Science of Expert and Exceptional Performance. *Curr Dir Psychol Sci*. 2007;16(6):346–50.

2. Cormier EM, Pickett-Hauber R, Whyte J. Cognitions and clinical performance: a comparison of high and low performing baccalaureate nursing students. *Int J Nurs Educ Scholarsh*. 2010;7(1):Article27.

3. Weller J, Bloch M, Young S, Maze M, Oyesola S, Wyner J, et al. Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. *Br J Anaesth*. 2003;90(1):43–7.

4. Zheng B, Tien G, Atkins SM, Swindells C, Tanin H, Meneghetti A, et al. Surgeon's vigilance in the operating room. Am J Surg. Elsevier Inc. 2011 May;201(5):673–7.

5. Ericsson KA, Smith J. *Toward a general theory of expertise: Prospects and limits*. Cambridge: Cambridge University Press; 1991.

6. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100(3):363.

7. Ford PR, Low J, McRobert AP, Williams AM. Developmental activities that contribute to high or low performance by elite cricket batters when recognizing type of delivery from bowlers' advanced postural cues. *J Sport Exer Psychol*. 2010;32(5):638–54.

8. Plant EA, Ericsson KA, Hill L, Asberg K. Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemp Educ Psychol*. 2005;30(1):96–116.

9. Whyte J, Ward P, Eccles DW. The relationship between knowledge and clinical performance in novice and experienced critical care nurses. *Heart Lung*. 2009 Nov;38(6):517–25.

10. Williams A, WARD P. Perceptual Expertise: Development in Sport. In: Starke J, Ericsson K, editors. *Expert Performance in Sport: Recent Advances in Research on Sport Expertise*. Human Kinetics, Champaign, IL; 2003. pp. 220–49.

11. Causer J, Barach P, Williams AM. Expertise in medicine: using the expert performance approach to improve simulation training. *Med Educ*. 2014;48(2):115–23.

12. Lorraway PG, Savoldelli GL, Joo HS, Chandra DB, Chow R, Naik VN. Management of simulated oxygen supply failure: is there a gap in the curriculum? *Anesth Analg*. 2006;102(3):865–7.

13. Nehring WM. U.S. boards of nursing and the use of high-fidelity patient

simulators in nursing education. *J Prof Nurs*. 2008;24(2):109–17.

14. Weller J, Robinson B, Jolly B. Psychometric characteristics of simulation-based assessment in anaesthesia and accuracy of self- assessed scores*. *Anaesth*. 2005;60:245–50.

15. Fletcher GCL, Mcgeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth*. 2002;88(3):418–29.

16. Schwid HA, Rooke GA, Carline J, Steadman RH, Murray WB, Olympio M, et al. Evaluation of anesthesia residents using mannequin-based simulation: a multiinstitutional study. *Anesthesiol*. 2002;97(6):1434.

17. Ericsson KA, Simon HA. Protocol Analysis: Verbal Reports as Data, revised edition. revised edition. A Bradford Book; 1993.

18. Fox MC, Ericsson KA, Best R. Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychol Bull*. 2011;137(2):316–44.

19. Whyte J IV, Cormier E, Pickett-Hauber R. Cognitions associated with nurse performance: A comparison of concurrent and retrospective verbal reports of nurse performance in a simulated task environment. *Int J Nurs Stud*. 2010;47(4):446–51.

20. Ward P, Suss J, Eccles DW, Williams AM, Harris KR. Skill-based differences in option generation in a complex task: a verbal protocol analysis. *Cog Process*. 2011;12(3):289–300.

21. Roca A, Ford PR, McRobert AP, Mark Williams A. Identifying the processes underpinning anticipation and decision-making in a dynamic time-constrained task. *Cog Process*. 2011;12(3):301–10.

22. McRobert AP, Causer J, Vassiliadis J, Watterson L, Kwan J, Williams MA. Contextual information influences diagnosis accuracy and decision making in simulated emergency medicine emergencies. *BMJ Qual Saf*. 2013; 22(6):478–84.

23. Fletcher G, Flin R, Mcgeorge P, Glavin R, Maran N, Patey R. Rating non-technical skills: developing a behavioural marker system for use in anaesthesia. *Cog Tech Work*. 2004;6(3):1–8.

24. North JS, Ward P, Ericsson A, Williams AM. Mechanisms underlying skilled anticipation and recognition in a dynamic and temporally constrained domain. *Memory*. 2011;19(2):155–68.

25. George K. Validity in clinical research: a review of basic concepts and definitions. *Phys Ther Sport*. 2003;4(3):115–21.

26. Cushion C, Harvey S, Muir B, Nelson L. Developing the Coach Analysis and Intervention System (CAIS): Establishing validity and reliability of a

computerised systematic observation instrument. *J Sports Sci*. 2012;30(2):201–16.

27. Thomas J, Nelson J, Silverman S. Research Methods in Physical Activity. Champaign, IL: Human Kinetics; 2010.

28. Doane SM, Sohn YW, Jodlowski MT. Pilot ability to anticipate the consequences of flight actions as a function of expertise. *Hum Factors*. 2004;46(1):92–103.

29. Mann DTY, Williams AM, Ward P, Janelle CM. Perceptual-cognitive expertise in sport: a meta-analysis. *J Sport Exerc Psychol*. 2007;29(4):457–78.

30. Ericsson KA, Whyte J, Ward P. Expert performance in nursing: reviewing research on expertise in nursing within the framework of the expert-performance approach. *Adv Nurs Sci*. 2007;30(1):58–71.

31. Kneebone RL. Simulation reframed. *Adv Simul*. 2016;1(27):1–8.

Figure legend

Figure 1: Number of verbal statement types (with *SD* bars) for skilled and less skilled participants.

1