

ESSAY

Personalizing simulation-based medical education: the case for novel learning management systems

Scott Pappada^{1,2,3}, Mohammad Hamza Owais³,
Shaza Aouthmany⁴, Paul Rega^{4,5,*}, Jeffrey Schneiderman⁶,
Serkan Toy⁷, Adam Schiavi⁸, Christina Miller⁸,
Rodrigo Daly Guris^{9,10,11}, Thomas Papadimos¹

¹Department of Anesthesiology, College of Medicine and Life Sciences, University of Toledo, Toledo, OH, USA

²Department of Bioengineering, College of Engineering, University of Toledo, Toledo, OH, USA

³Department of Electrical Engineering and Computer Science, College of Engineering, University of Toledo, Toledo, OH, USA

⁴Department of Emergency Medicine, College of Medicine and Life Sciences, University of Toledo, Toledo, OH, USA

⁵Department of Public Health and Preventive Medicine, School of Population Health, University of Toledo, Toledo, OH, USA

⁶College of Medicine and Life Sciences, EMS Education, University of Toledo, Toledo, OH, USA

⁷Department of Basic Science Education, Virginia Tech Carilion School of Medicine, Roanoke, VA, USA

⁸Department of Anesthesiology and Critical Care Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA

⁹Department of Anesthesiology and Critical Care Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA

¹⁰Center for Simulation, Advanced Education and Innovation, Children's Hospital of Philadelphia, Philadelphia, PA, USA

¹¹Department of Anesthesiology and Critical Care, University of Pennsylvania, Philadelphia, PA, USA

Corresponding author: Scott Pappada, scott.pappada@utoledo.edu

<https://ijohs.com/article/doi/10.54531/MNGY8113>

ABSTRACT

Simulation-based medical education (SBME) is often delivered as one-size-fits-all, with no clear guidelines for personalization to achieve optimal performance. This essay is intended to introduce a novel approach, facilitated by a home-grown learning management system (LMS), designed to streamline simulation program evaluation and curricular improvement by aligning learning objectives, scenarios, assessment metrics and data collection, as well as integrate standardized sets of multimodal data (self-report, observational and neurophysiological). Results from a pilot feasibility study are presented. Standardization is important to future LMS applications and could promote development of machine learning-based approaches to predict knowledge and skill acquisition, maintenance and decay, for personalizing SBME across healthcare professionals.

What this essay adds

- This essay introduces a novel learning management platform that standardizes both the creation and assessment of simulation-based medical education.
- The essay discusses the necessity of multimodal data acquisition to evaluate and skill acquisition, maintenance and decay during simulation.
- This essay highlights the potential of neurophysiological data monitoring to evaluate proficiency, performance and future training requirements.
- This essay documents the necessity of standardized data and measures to future implementation of machine learning and artificial intelligence-based approaches for personalized training and education.

Introduction

Simulation-based medical education (SBME) is an established and valuable component of health professions education, as it allows repeated practice with immediate feedback in a low-risk environment. SBME has demonstrated value in teaching technical skills, communication and teamwork [1,2], and compared to other modalities is associated with improved, cognitive, behavioural and psychomotor skills, which may translate to improved patient care and safety, as well as increased learner satisfaction [3–6]. A well-resourced simulation program affords the ability to observe inter- and intra-learner variations and alter the situational complexity to which participants are exposed [7].

Despite the accumulating evidence of the utility and efficacy of SBME, the modality's full potential is far from realized. In particular, robust, standardized and generalizable assessment frameworks are currently lacking. Insofar as multimodal assessment frameworks do exist, they are limited to defined surgical procedures and have procedure-specific metrics that are not generalizable beyond narrow use cases [8].

A simulation participant's proficiency (and presumed operational readiness) is commonly determined using observational tools such as checklists [9]. Such tools tend to be binary in nature (task was or was not accomplished) and may lack both granularity and nuance. Therefore, the generation of individualized, specific and actionable feedback for performance improvement, and the customization of simulation complexity to derive maximal educational impact are largely left to individual facilitators, with little guidance other than their own experience and preferences. While personalization is crucial to SBME effectiveness [10], it is not yet standard (or standardized) practice. In short, obtaining optimal results from SBME curricula remains a key-person-dependent endeavour, and the definition of optimal (or ability to reproduce optimal) may be wildly variable between practitioners.

To further complicate matters, learning and forgetting curves are dynamic between individuals, and studies suggest that skill decay can begin as early as three months after training [11,12]. The intensity and amount of initial training necessary to achieve proficiency in a particular skill and also the content and frequency of continued training to maintain proficiency remain somewhat of a mystery, but

they are likely to be variable between individuals, due to both intrinsic characteristics as well as lived experience. That said, much of medical education, including SBME, is delivered in a one-size-fits-all approach, with identical content and exposure frequency for every learner. It stands to reason that this probably results in underexposure in some learners and overexposure in others – an inefficient use of limited educational resources. As such a proportion of learners will not be prepared to perform at their peak capacity because they did not achieve the predetermined learning objectives (underexposure). In other learners, there is likely overexposure to training that occurs when learners do not require retraining due to maintenance of knowledge and skills due factors such as recent clinical experiences.

Finally, there is a wealth of untapped knowledge that can be generated by consistently studying individual and team performance in a controlled simulated setting. It is possible to objectively stratify top performers from standard and lower performers [13]; understanding what differentiates top performers could inform efforts to achieve peak team performance.

In an ideal world, SBME design would incorporate objective individual-level feedback to allow personalized and finely granular manipulation of training content and frequency. This is no small feat and requires not only the collection of numerous subjective and objective data points, but their deliberate interpretation and application. While there are commercially available systems that offer the ability to define custom performance measures or learning event assessments (e.g. SimCapture by Laerdal and Learning Space by CAE Healthcare), these do not offer standardization of performance measures (i.e. there is freedom to define only custom events or performance measure which promotes lack of standardization), compromising generalizability. As such, the landscape is wide open for innovation.

In this essay, we describe the pilot development, implementation, and testing of a home-grown learning management system (LMS) dubbed PREPARE (*PRE*diction of Healthcare Provider Skill Acquisition and Future Training *RE*quirements). This is but one example of what is possible; our hope is that this description demonstrates the value of such an approach and encourages further research, development and collaboration in training with the ultimate goal being peak performing healthcare teams.

Platform description and development

PREPARE was designed as a comprehensive measurement and assessment platform with the explicit purpose of integrating multiple data streams at the learner, instructor, and training environment levels. To achieve this, PREPARE affords the following capabilities:

- (i) Learning scenario design (standardized methodology)
- (ii) User-generated data entry
- (iii) Automated data capture
- (iv) Data analysis and interpretation

Learning scenario design is broadly organized around *goals* (i.e. what the learner will accomplish, e.g. the management of a difficult airway in an acutely decompensating patient), and specific *learning objectives* around which training and evaluation revolve. During scenario design, simulation scenarios are populated with *learning events*, which are tactical-level instances during the simulation and are what allow for finely granular assessment with millisecond-level resolution if so desired. Each learning event is tagged or mapped to one or more *learning objectives*. This approach allows for intentionality in scenario design and learner evaluation. Figure 1 includes a user interface where facilitators assess various *learning events* across scenario timeline (Figure 1 Item #3) that are preprogrammed to a learning scenario. For future generalizability, *learning events* are also mapped to broad and generalized skill categories, and discipline-specific skills (Figure 2).

Established and standardized measurement frameworks are core to LMS platforms serving as multimodal assessment tools. Three types of performance measures are defined within PREPARE, two of which are defined at the instructor level and one of which is defined at the learner level. All measures and assessments defined or derived by the platform are mapped to predefined *learning events* such that they can be associated with specific knowledge, skills, and learning objectives.

Subjective and objective data entry is possible at both the learner and the facilitator levels. This allows for the collection of information such as demographic and lived experience data, baseline knowledge and confidence levels and cognitive load assessments, among others. Within this iteration of PREPARE, facilitators grade performance of each individual learning event at the learner level via two approaches (see Figure 1 Item #4): on a continuous colour-coded scale which is translated to a 0–100 quantitative measure, and via a subjective evaluation of each learning event, including overall performance on the scenario as *novice*, *intermediate*, or *expert*. This allows for categorization of learners into finite classes that can be mapped or evaluated with respect to other system measures.

The platform facilitates automated objective data capture at the environmental level by allowing synchronization of learning events with audio and video recording of simulation events (if time-stamped data streams are available), simulator data (or other data from the training environment), as well as capture of time-stamped

Figure 1: PREPARE user interface providing system measurement and assessment capabilities

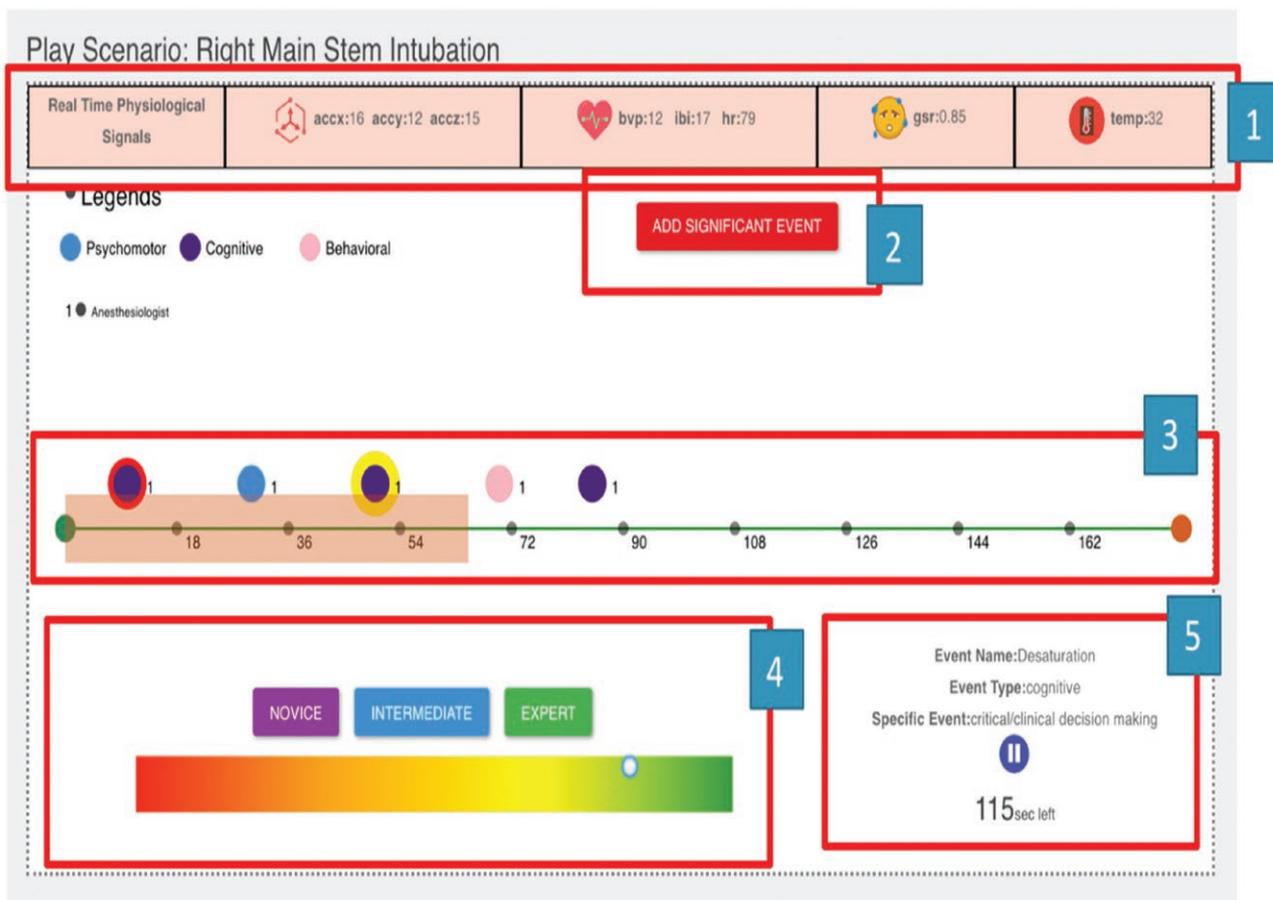
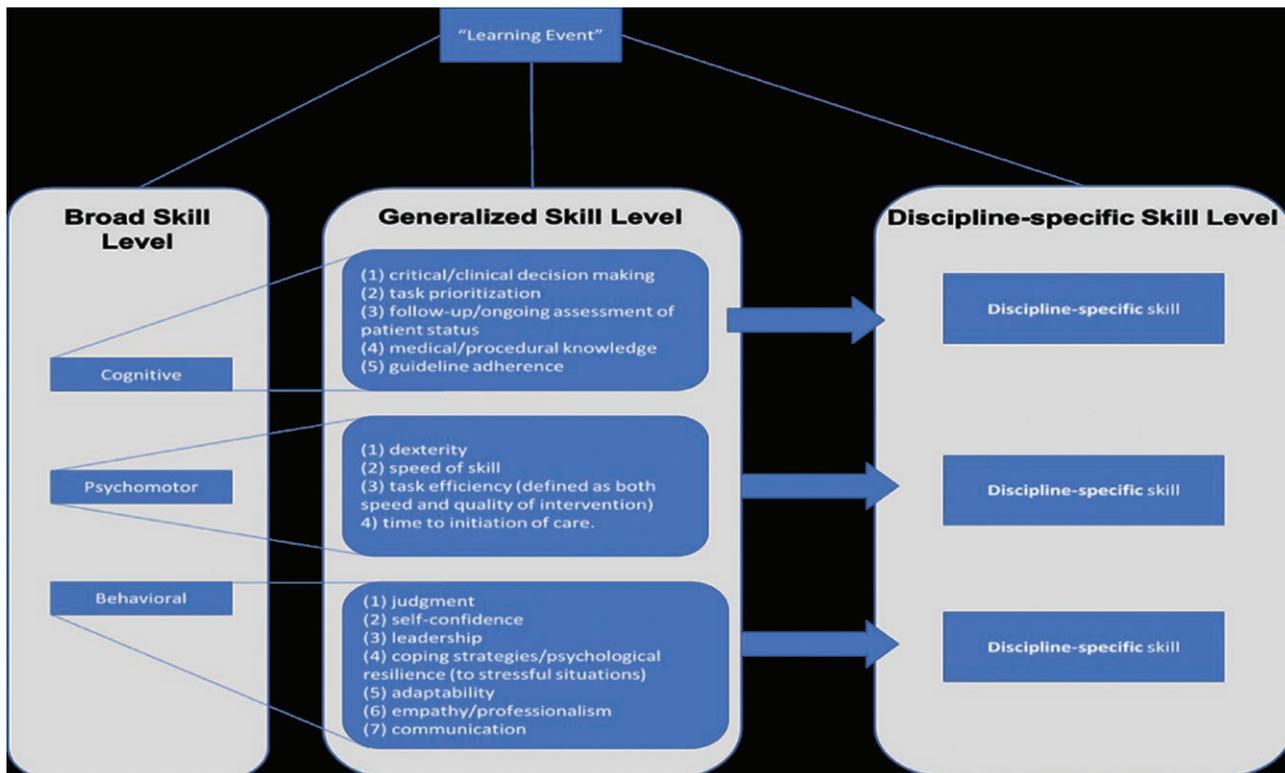


Figure 2: Hierarchy of measures collected and derived by PREPARE which promotes standardization



learner physiologic measures from any custom built or commercially available wearable devices and monitors. The incorporation of neurophysiological data as a core performance measure is intended to provide more objective performance measures that support and are complementary to observer-based/subjective assessments made with the platform. Neurophysiological data is standardized by its very nature. For example, everyone has a heart rate and increases in heart rate reflect a state of arousal, stress, or excitement. Although the presence of a set of neurophysiological measures is standard across humans, values of heart rate, cortical activity, sweat and others vary tremendously across individuals. Individuals have different resting physiological values and/or stress responses. There is a necessity to standardize neurophysiological measures to provide generalizability across data collected and derived by the platform. To promote standardization of neurophysiological measures, PREPARE is designed with algorithms that automatically detect baseline physiological measures and calculate a percentage deviation relative to baseline measures.

To facilitate debriefing, a visual representation of evaluation data is automatically generated at the conclusion of each scenario. In current state, these data are interpreted by the facilitator and raw data can be exported for further analysis. With the generation of sufficient data in the future, there will be a role for application of machine learning techniques and other statistical methods to classify learner expertise and predict future training requirements to optimize acquisition and maintenance of knowledge and skills.

To maximize usability, PREPARE was built as a platform-agnostic web-based software application compatible with

most browser-enabled devices, including smartphones and tablets.

Initial implementation and pilot study design

A pilot study was developed to test and evaluate the PREPARE platform and to demonstrate its potential. This pilot study was conducted with the approval of the University of Toledo's Institutional Review Board (Protocol #202907) and was executed at a medium-sized suburban academic institution's simulation centre. Written consent was obtained by all Emergency Medicine (EM) residents prior to their enrolment in the study. The study was conducted during regularly scheduled monthly simulation sessions for EM trainee physicians (PGY 1-3 residents). Three EM faculty members participated in learner evaluations. Seven of the institution's 25 EM residents participated in this pilot study, which was conducted during regularly scheduled monthly simulation sessions. Residents enrolled to the study participated in three simulation scenarios while monitored by the PREPARE platform. The scenarios included: (1) treatment of paediatric abuse, (2) treatment of a trauma patient with pelvic fracture and (3) treatment of patient with a Le Fort fracture. Each simulation was approximately 15–20 minutes in duration. Goals, objectives and learning events for each scenario were identified (*a priori*) and evaluated (during simulations) by the three participating EM faculty members (one assigned to generate each scenario). The defined goals, objectives and learning events were all mapped to specific Accreditation Council of Graduate Medical Education (ACGME) milestones.

The intent of this initial platform roll out and pilot study was threefold. The first aim was to evaluate the features and

functionality of the platform during intended use to ensure that they met the requirements for curriculum and data standardization. The second aim of the pilot study was to gather feedback on utility and perception of the platform compared to conventional simulation-based training and debriefing, from both facilitators and learners. The third aim was to identify and evaluate relationships between physiological responses (e.g. heart rate, electrodermal activity) and observed performance recorded by domain experts (i.e. clinical faculty). Data analysis efforts were limited to visual inspection due to the small number of subjects enrolled.

Usability was evaluated through user-generated feedback from both facilitators and learners (after post-scenario debriefs), focussed on utility and initial impression compared to traditional simulation evaluation and debriefing.

To test neurophysiological data acquisition and processing capabilities of the platform, all learners wore a commercially available wrist-worn wearable device (Empatica E4 Wristband [Boston, MA]). Due to the fact that many commercially available devices such as the Empatica device produce heavily averaged estimates of heart rate, PREPARE software derives a set of instantaneous physiological measures. For the Empatica device, a photoplethysmogram signal was processed via PREPARE software algorithms to derive instantaneous estimates of heart rate rather than the values reported by the device. In addition to measures of heart rate, the Empatica device also provided measurement of EDA (i.e. sweat), skin temperature, and motion via a 3-axis accelerometer. Our hypothesis

was that there would be an inverse relationship between physiological responses (indicating stress) and performance. As such the expectation was that there would be stress responses associated with poorer performance as evaluated by EM faculty.

To minimize the occurrence of faulty inferences, physiologic data were only analysed in relation to learning event assessments if the facilitator's evaluation of the event was completed within 15 seconds of the event occurrence, as identified by video review. Evaluations made outside of this 15 second window were defined as delayed and removed from analysis.

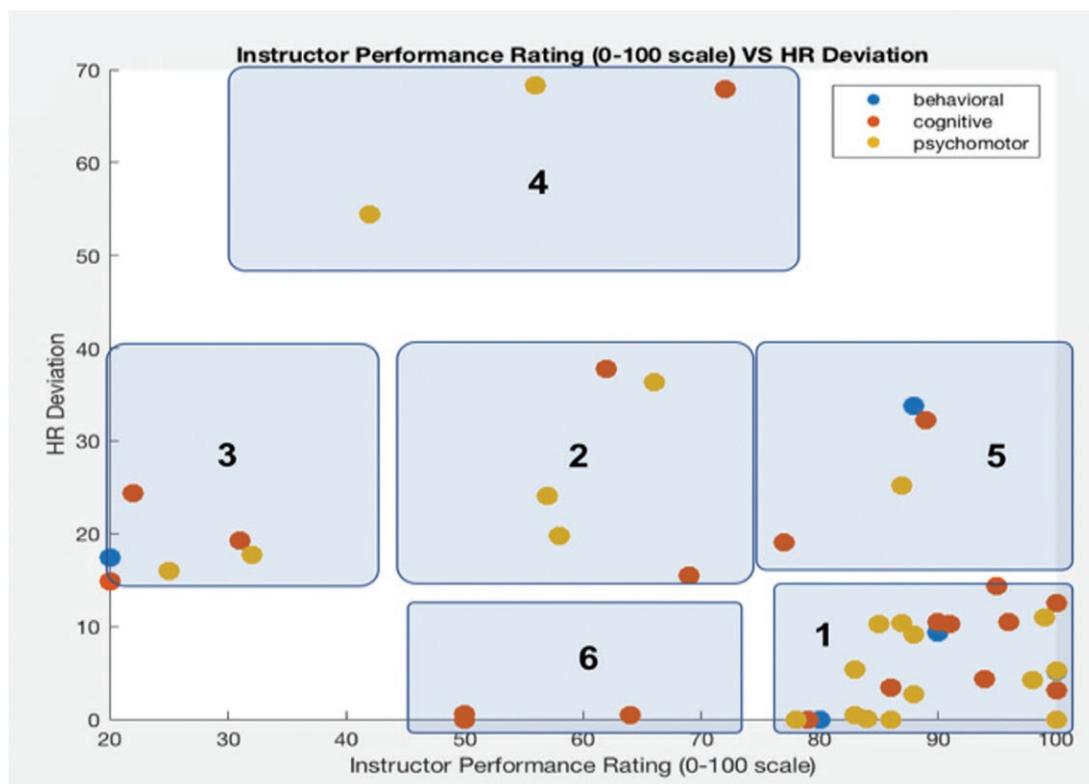
Preliminary data and results

Comprehensive multimodal dataset of performance measures and physiological data were collected from seven (two PGY1, three PGY2 and two PGY3) EM residents. Figure 3 shows potential correlations with performance (x-axis) and percentage deviation from algorithm-derived baseline heart rate (y-axis).

This figure demonstrates there are potential clusters which could be visually identified as consistent with our initial expectation. A smaller deviation from baseline heart rate appears to be associated with better facilitator-documented performance ratings. There were a total of 45 facilitator documented performance ratings across all subjects enrolled in this study. Each of these ratings is colour coded as a cognitive, behavioural or psychomotor skill.

Cluster group 1 (Figure 3) includes the intersection of heart rate values with the smallest deviation (<15%) from baseline

Figure 3: Data collected using the Empatica E4 wristband and PREPARE indicating potential relationship between heart rate and learner proficiency/performance



and the highest performance ratings (>75 out of 100). This group also contains the highest number of facilitator-documented performance ratings (24 out of 45 ratings or 53.3% of all ratings) collected across the study population.

Heart rate had the most consistent relationship with performance as detailed above. Other measures from the E4 wearable device such as EDA had some visible relationships with performance however because changes in EDA were more gradual and sustained than heart rate measures, it was more difficult to associate changes in sweat with changes in performance over smaller time frames. Skin temperature did not change significantly across any study participant.

Faculty and learners appreciated the data rich debriefing capability provided by the platform and its intelligent data visualizations. Learners valued the degree of quantifiable feedback available compared to conventional debriefing sessions. Instructors believed that the data represented in the platform provided a more focussed and targeted debrief and that core learning objectives could be easily referenced and discussed with respect to learner expertise and comprehension.

Figure 4 includes examples of the PREPARE platform's data visualizations. These user interfaces (UIs) display content which allows learners and facilitators to review performance assessments and physiological responses for each training activity documented in the platform. Learning events are organized based on the type of skill (based on measurement hierarchy, Figure 2) and have a 0–100 measure associated with them as evaluated by a facilitator. The sequence of events as they were completed is also represented. Events that are solid blue were completed in the correct sequence and those that are grey were not done in the anticipated sequence. Using this, a facilitator can easily discuss sequence of actions, if relevant, during a debrief. The mean of all skills evaluated by faculty at each broad skill level (cognitive,

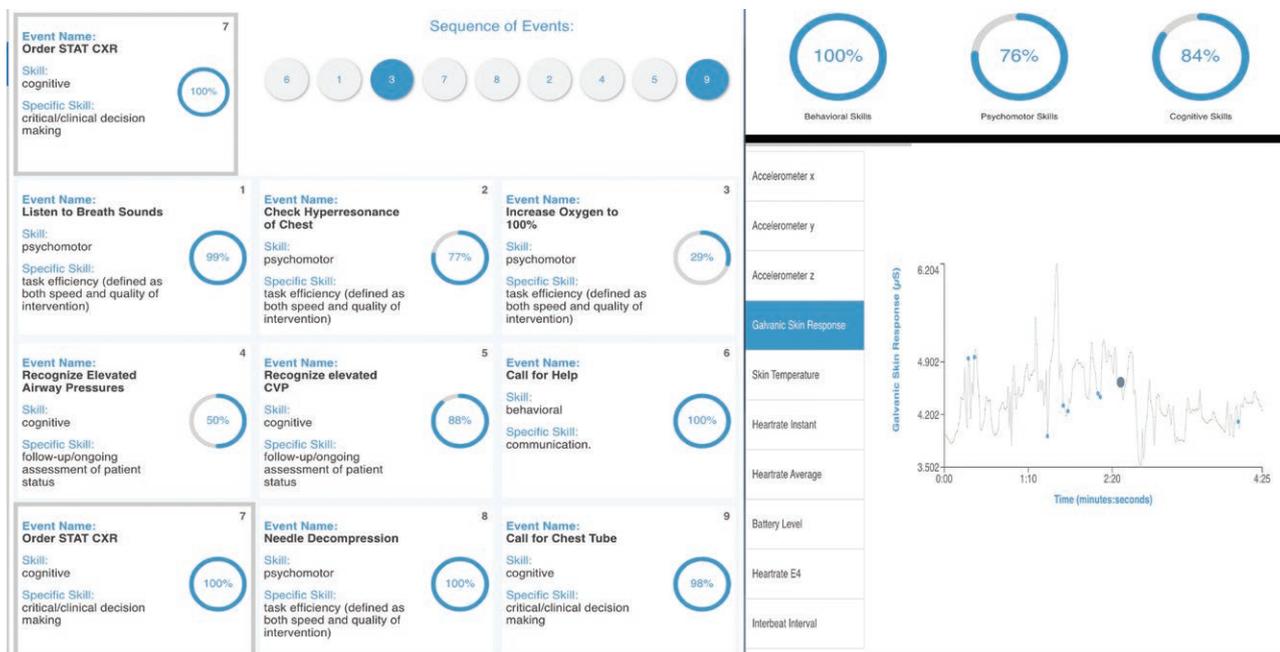
behavioural and psychomotor) are represented in the upper right quadrant of the screen. Data collected from the E4 wristband are represented below the broad skills breakdown and which provides a visualization of trends in physiological measures around learning events (circles on plot) evaluated by faculty.

Interpretation of initial results

This pilot study suggests that it is feasible to simultaneously collect subjective and objective observed performance and device-recorded physiologic data, mapped to specific learning events, in a standardizable fashion during SBME for emergency medicine trainees. These data suggest an association may exist between heart rate, EDA and observed performance, though the strength of this association likely varies between individuals and may also be affected by other extrinsic influences. Findings are consistent with prior work demonstrating an inverse relationship between physiologic signatures of stress and performance [14–16].

Feedback on the platform was overwhelmingly positive from both facilitators and learners. Further rigorous study with large datasets is required to identify reliable physiologic patterns that may be predictive of performance both between individuals and, perhaps more reliably, within a single participant. Although faculty were trained on how to evaluate learners with the platform, and instructed to do so in a timely fashion, this was not always achieved. A limitation of this study was delayed or missing evaluations. This was a somewhat common occurrence with an average of two delayed evaluations per scenario across the faculty. This was not a surprising result, as staffing limitations present at our institution required faculty to serve as a both a confederate and an evaluator for this study. As such, faculty had two (at times competing) responsibilities which included driving or reporting changes in patient state

Figure 4: Platform user interface and data visualizations with interactive capabilities which allow data-driven exploration during instructor/learner debriefing sessions



(dynamically adapting the scenario), and evaluating learner performance. Future development efforts seek to potentially automate the cueing of instructor assessments based on the automated detection of key events, action or activities that occur in real-time during a SBME session.

Future directions for multimodal data-driven simulation

To ensure that clinicians and clinical teams perform at the peak of their potential, it is crucial to employ optimized teaching and training approaches. This necessarily includes personalized training, targeted to individual needs, abilities, and preferences. Achieving this in a consequential way depends on the ability to access and analyse comprehensive multimodal data. Standard terminology and architecture will likely prove useful in interpreting large amounts of multi-source data across individuals, disciplines and institutions.

The pilot study described in this essay barely scratches the surface of what is possible; following the results of this pilot, the PREPARE platform is now being rolled out across all graduate medical training programs and undergraduate medical clerkships at the parent institution. This should immediately enhance existing simulation education curricula by facilitating objective, data-driven debriefing around specific learning events. In the long term, datasets generated by use of the platform will allow for the application of machine learning, cluster analysis and other statistical techniques to better understand factors that facilitate or inhibit peak performance. With a standardized portable evaluation platform such as PREPARE, it may also be possible to evaluate real-world clinical performance, which could allow for correlative analysis between performance in training and performance *in vivo*, further informing the optimization of training modalities and improving the understanding of human performance in the clinical setting.

Research in the areas of wearable physiological monitoring for performance assessment during training and real-world operational settings is not a new pursuit and has been the subject of many prior and ongoing research efforts [17,18], but these have yet to be fully integrated into a comprehensive standardized performance evaluation system. As commercially available physiologic monitors become less intrusive and of higher fidelity, efforts in the near future should be able to incorporate additional data streams such as portable electroencephalography and eye tracking, and thus provide objective insights into physical, mental and temporal workloads and distractions [19,20]. Eye tracking and accelerometry may be able to inform human factors and ergonomics driven approaches to optimize workflows.

One potential downside to current commercial wearable monitors is that their out-of-the-box algorithms are heavily processed to minimize noise. While this may be acceptable in certain applications, during relatively short and intense simulation sessions signals may be masked by averaging algorithms. Systems should be designed around the ability to access and analyse high sampling frequency raw data to maximize the ability of signal detection. To this end,

PREPARE contains algorithms to capture instantaneous changes in physiological data (e.g. heart rate) which helps capture learner experiences and changes in stress as they dynamically evolve during patient care scenarios.

Taking a lesson from Safety II [21,22], cluster analysis using adaptive learning systems such as PREPARE may help identify characteristics that top performers have in common, perhaps informing subsequent efforts to upskill other individuals in a personalized fashion. Learner-level analytics, informed by tracking performance over time and compared to large datasets, may be applied to personalize training, informing aspects such as frequency, content, intensity and modality.

Ultimately, it is possible that such endeavours could help identify impending performance degradation using physiologic monitoring of clinical teams *in vivo* (e.g. by monitoring for signs of engagement, distraction, stress, cognitive load, fatigue), allowing for proactive strategic intervention.

Declarations

Authors' contributions

None declared.

Funding

None declared.

Availability of data and materials

None declared.

Ethics approval and consent to participate

None declared.

Competing interests

None declared.

References

1. Cook DA, Brydges R, Hamstra SJ, et al. Comparative effectiveness of technology-enhanced simulation versus other instructional methods: a systematic review and meta-analysis. *Simulation in Healthcare*. 2012;7(5):308–320.
2. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*. 2011;306(9):978–988.
3. Ford DG, Seybert AL, Smithburger PL, Kobulinsky LR, Samosky JT, Kane-Gill SL. Impact of simulation-based learning on medication error rates in critically ill patients. *Intensive Care Medicine*. 2010;36(9):1526–1531.
4. Pascual JL, Holena DN, Vella MA, et al. Short simulation training improves objective skills in established advanced practitioners managing emergencies on the ward and surgical intensive care unit. *Journal of Trauma and Acute Care Surgery*. 2011;71(2):330–338.
5. Jansson MM, Syrjälä HP, Ohtonen PP, Meriläinen MH, Kyngäs HA, Ala-Kokko TI. Randomized, controlled trial of the effectiveness of simulation education: a 24-month follow-up study in a clinical setting. *American Journal of Infection Control*. 2016;44(4):387–393.

6. Eyikara E, Baykara ZG. The importance of simulation in nursing education. *World Journal on Educational Technology: Current Issues*. 2017;9(1):2–7.
7. Okuda Y, Bryson EO, DeMaria Jr S, et al. The utility of simulation in medical education: what is the evidence? *Mount Sinai Journal of Medicine*. 2009;76(4):330–343.
8. Hermsen JL, Mohamadipanah H, Yang S, et al. Multimodal cardiopulmonary bypass skills assessment within a high-fidelity simulation environment. *The Annals of Thoracic Surgery*. 2021;112(2):652–660.
9. Hardy J-B, Gouin A, Damm C, Compère V, Veber B, Dureuil B. The use of a checklist improves anaesthesiologists' technical and non-technical performance for simulated malignant hyperthermia management. *Anaesthesia Critical Care & Pain Medicine*. 2018;37(1):17–23.
10. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher*. 2005;27(1):10–28.
11. Arthur Jr W, Bennett Jr W, Stanush PL, McNelly TL. Factors that influence skill decay and retention: a quantitative review and analysis. *Human Performance*. 1998;11(1):57–101.
12. Yang C-W, Yen Z-S, McGowan JE, et al. A systematic review of retention of adult advanced life support knowledge and skills in healthcare providers. *Resuscitation*. 2012;83(9):1055–1060.
13. Guris RJD, Miller CR, Schiavi A, Toy S. Examining novice anaesthesia trainee simulation performance: a tale of two clusters. *BMJ Simulation & Technology Enhanced Learning*. 2021;7(6):548.
14. Harvey A, Bandiera G, Nathens AB, LeBlanc VR. Impact of stress on resident performance in simulated trauma scenarios. *Journal of Trauma and Acute Care Surgery*. 2012;72(2):497–503.
15. Wetzel CM, Black SA, Hanna GB, et al. The effects of stress and coping on surgical performance during simulations. *Annals of surgery*. 2010;251(1):171–176.
16. McGraw LK, Out D, Hammermeister JJ, Ohlson CJ, Pickering MA, Granger DA. Nature, correlates, and consequences of stress-related biological reactivity and regulation in Army nurses during combat casualty simulation. *Psychoneuroendocrinology*. 2013;38(1):135–144.
17. Pappada SM, Papadimos TJ, Lipps JA, et al. Establishing an instrumented training environment for simulation-based training of health care providers: an initial proof of concept. *International Journal of Academic Medicine*. 2016;2(1):32.
18. Pappada S, Geyer A, Durkee K, Freeman J, Cohn J, editors. Modeling operational workload for adaptive aiding in Unmanned Aerial Systems (UAS) operations. *Proceedings of the 84th Annual Aerospace Medical Association (AsMA) Scientific Meeting*; May 15, 2013.
19. Roy RN, Charbonnier S, Campagne A, Bonnet S. Efficient mental workload estimation using task-independent EEG features. *Journal of Neural Engineering*. 2016;13(2):026019.
20. So WK, Wong SW, Mak JN, Chan RH. An evaluation of mental workload with frontal EEG. *PLoS One*. 2017;12(4):e0174949.
21. Bentley SK, McNamara S, Meguerdichian M, Walker K, Patterson M, Bajaj K. Debrief it all: a tool for inclusion of Safety-II. *Advances in Simulation*. 2021;6(1):1–6.
22. Staender S. Safety-II and resilience: the way ahead in patient safety in anaesthesiology. *Current Opinion in Anesthesiology*. 2015;28(6):735–739.