THEME ARTICLE: HUMAN-CENTERED AI

HuCETA: A Framework for Human-Centered Embodied Teamwork Analytics

Vanessa Echeverria, Roberto Martinez-Maldonado, Lixiang Yan, Linxuan Zhao, Gloria Fernandez-Nieto, and Dragan Gašević, *Monash University, Clayton, VIC, 3800, Australia*

Simon Buckingham Shum, *University of Technology Sydney, Sydney, NSW, 2007, Australia*

Collocated teamwork remains a pervasive practice across all professional sectors. Even though live observations and video analysis have been utilized for understanding embodied interaction of team members, these approaches are impractical for scaling up the provision of feedback that can promote developing high-performance teamwork skills. Enriching spaces with sensors capable of automatically capturing team activity data can improve learning and reflection. Yet, connecting the enormous amounts of data such sensors can generate with constructs related to teamwork remains challenging. This article presents a framework to support the development of human-centered embodied teamwork analytics by 1) enabling hybrid human–machine multimodal sensing; 2) embedding educators' and experts' knowledge into computational team models; and 3) generating human-driven data storytelling interfaces for reflection and decision making. This is illustrated through an in-the-wild study in the context of healthcare simulation, where predictive modeling, epistemic network analysis, and data storytelling are used to support educators and nursing teams.

Individuals who demonstrate effective teamwork and communication skills are in high demand across all professional sectors.[1] High-stakes sectors, such as healthcare and emergency response (e.g., see Figure 1(top), are particularly susceptible to underdeveloped teamwork skills, where if teamwork issues are not addressed, wrong decisions may lead to people getting injured or even death.[2] It is thus of great importance for education providers to produce graduates who are equipped with the skills that members of high-performing teams require.

Making evidence of teamwork activity available for learning and assessment has been proposed as an effective way to encourage team members to reflect on their own experiences to learn from them.[3] Yet, in practice, educators, coaches, and team members commonly have sparse evidence to reflect upon.[4] Creating spaces equipped with sensors and input devices capable of automatically capturing multimodal behavioral traces of collocated team activity can potentially be used to assess it for the purposes of learning, reflection, and coaching.[5] These traces can range from low-level logs, such as click-streams, to nonmediated human actions, such as body positioning, manipulation of objects, and gestures. These automatically captured activity logs, along with the observation by team experts or coaches, can be analyzed using data science and artificial intelligence (AI) techniques to cluster team behaviors, identify archetypal teams, and identify highly effective practices.[6]

A growing body of contemporary research has aimed at modeling salient aspects of teamwork using sensor-based multimodal innovations in learning contexts[5] and for research purposes.[7,8] Other works have proposed conceptual frameworks, such as the input-process-outputs, to align teamwork salient aspects with multimodal and sensor data.[9] Yet, little work has moved from conceptualization to actual solutions focused on both supporting and involving the *end-users* (i.e., team members and their coaches) in the coconfiguration of the analytics they will ultimately use. Connecting the enormous amounts of data,

1536-1268 © 2022 IEEE
Digital Object Identifier 10.1109/MPRV.2022.3217454

FIGURE 1. Top: A training hospital ward for running high-fidelity team simulations enriched with wearable sensors capable of automatically capturing team activity data. Bottom: The Human-Centered Embodied Teamwork Analytics HuCETA) Framework: a teamwork analytics infrastructure enables human agents to contribute their expertise to the modeling of team performance, generating exploratory and explanatory visual analytics for researchers, educators/coaches, and team members.

which sensors can generate with meaningful constructs related to high-performance teamwork (e.g., effective communication, coordination, and leadership) remains a key challenge. The critical underexplored challenge is how to assist educators and learners, who may be considered nondata science savvy users, in the formative assessment and improvement of collocated teamwork, by making multimodal activity traces visible and available for computational analysis, and rendering these in meaningful, actionable user interfaces.

This article presents the Human-Centered Embodied Teamwork Analytics (HuCETA) framework, which is designed to address the aforementioned challenge by enabling people without formal data analysis training to directly contribute their expertise on teamwork or the team task to key analytics tasks, namely 1) multimodal sensing, 2) team modeling, and 3) the automated generation of team data visualizations. We illustrate this through an in-the-wild study in the context of healthcare simulation; and briefly discuss implications of automatically generating meaningful representations from sensor data.

THE HUCETA FRAMEWORK

An overview of the conceptual framework is shown in Figure 1(bottom). The framework incorporates a network of wireless sensors that captures multimodal team activity (e.g. voice, body position, physiological signals, and torso rotation), and three computational components: 1) hybrid human–machine sensing; 2) quantitative ethnographic team modeling, and 3) an expert-driven storytelling engine. Each of these three components can provide user interfaces for various stakeholders to provide input to the modeling and AI algorithms or to receive outputs in the form of exploratory and explanatory user interfaces. Each component is further described below.

Hybrid Human–Machine Sensing

Sensors can capture vast amounts of interaction data from physical spaces at scale without observers influencing the activity. Although such data can directly be mined using data science and AI techniques, data are commonly and easily stripped from their higher order meanings.[4] In contrast, compared to sensors, humans are able to observe a much broader spectrum of physical and social events and give meaning to certain team actions. Yet, humans are not as reliable as a tested sensing infrastructure in terms of the correctness of certain reported observations.[10]

Inspired by the notion of "humans as sensors,"[10] the first component aims at integrating the advantages of both people and sensors in capturing activity traces by enabling hybrid human–machine unobtrusive observations. A network of wireless sensors can be used to monitor team behaviors that can be hard for humans to track (e.g., exact positions and angular body orientation of team members) or be completely invisible to a regular observer (e.g., sudden changes in physiological signals or subtle prosodic features of speech). Humans (e.g., coaches, researchers, or team members), equipped with suitable user interfaces, can log critical events as activity unfolds, or afterward for the purpose of contextualizing the various streams of data. It can also serve to estimate the reliability of observations when multiple observers are coding behaviors for further classification purposes.

Quantitative Ethnographic Team Modeling

While logs can illuminate what users *do*, they often say much less about *why*.[11] We address this challenge of enriching quantitative data streams with the qualitative insights needed to make sense of them by embracing a quantitative ethnography (QE) approach.[12] QE views Big Data as evidence about the discourse of a particular culture. To make meaning from this evidence, and thus gain some understanding of the culture, the aim is to achieve a qualitatively "thick" description of such data, that is, the low-level data needs to be *coded* into meaningful higher order constructs. For example, low-level x-y positioning coordinates can be automatically coded into spatial and orientation relationships among team members in the physical space (i.e., *f-formations*), which can be indicative of effective team practices.[13] A combination of particular team members' formations and actions can be indicative of performance.

Inspired by this idea, this component builds a model, termed the multimodal matrix,[14] which takes as inputs: 1) multiple streams of team data from both sensors and human observers coming from the prior component Figure 1(bottom); and 2) the semantics (i.e., *codes*), derived from a qualitative interpretation of the context in which the team activity occurs. These codes can be provided to the computational systems by researchers (e.g., team experts), or educators, coaches, and task designers (i.e., task experts) using an input user interface. These can take the form of rules, parameters, or templates used by the system to convert *low-level data* into *meaningful information*. For example, task experts or researchers can create rules to assess whether certain combinations of logged events are indicative of correct team procedures or set threshold parameters to assess whether the physiological signals are indicative of higher stress levels. Similarly, AI models could be applied to automatically label the discourse of the team to identify instances of high-performance team dialog.

Outputs from this component can be in the form of *exploratory* user interfaces designed for experts in data analysis or researchers in team science and educational technologies in search of insights from unfamiliar datasets.[15] Exploratory interfaces invite people to explore the data without a strong educational or coaching goal in mind. However, to support less technical team members and educators/coaches, we provide the next component in the framework.

Expert-Driven Storytelling Engine

Significant progress has been made in recent years to integrate and analyze multimodal data.[16] Integration of data streams from various modalities and actors can result in complex interfaces and difficult interpretations. Yet, little research has examined the challenge of visualizing and supporting sensemaking of these data to inform team coaching and learning.

Data storytelling is an emerging suite of information design and "compression" techniques to help an audience effectively understand what is important in a visualization by communicating key messages clearly through a combination of data, visuals, and narratives.[15] This component is inspired by human-centered AI notion of *supervisory control*, which promotes the human operation and oversight of highly automated systems.[17] This component takes as an input a set of parameters from team performance experts (i.e., educators and task designers who have expertise in what *highly effective teamwork* looks like for a particular scenario). It then automatically renders visible the modeling outputs in the form of data stories in ways that end-users (e.g., team members and their coaches) can readily understand.

A data story can, for example, highlight key data points and add text descriptions to a chart to communicate the assessment of a team's procedure. It could also inform team members about errors they made by highlighting such errors in a visualization. It can also show members their recorded stress levels to spark conversations about emotions experienced in a reflective debrief. As a result, this approach generates *explanatory* learning analytics visualizations that communicate insights rather than data points.[15] We illustrate this and the other components of the framework through an authentic case study presented below.

CASE STUDY

Based on the framework described above, we are developing an open-source embodied teamwork analytics platform[a] distilling progress from a five-year learning analytics program in healthcare teamwork simulation. This section presents two studies illustrating how the framework supports the development of human-centered teamwork analytics for in-the-wild nursing education.

Participants, Educational Context, and Task

The simulations presented in this case study were conducted in an authentic learning setting as part of the regular classes of a nursing degree between August and September 2021. The case study focuses on 57 team simulations (228 students, aged between 20 and 23) in which all four team members consented to participate. Three educators led these simulations, taking turns to evaluate student (task and collaboration) performance from a one-way mirror control room. Each simulation started with an educator's briefing introducing the scenario. Students performed the simulation with four patient manikins controlled by educators from the control room, and a patient's relative played by an educator (see "Actor" in Figure 1 (top). The average duration of simulations was 20.25 min (std = 8.13 min). After each simulation, an educator evaluated the performance of each team in terms of the collaboration and task-related learning objectives using a seven-point likert scale. We identified a team as high performing in terms of collaboration or task performance if their score was within the top three out of seven scale items. This binary classification rule was validated by the educators.

Apparatus and Data

All consenting students were asked to wear a waist bag containing an *indoor positioning sensor*[b] to collect spatial data such as body orientation and x-y coordinates (measured in degrees and millimeters, respectively). Each student also wore a *wireless headset microphone* connected to an audio interface to enable real-time synchronization, and a *physiological wristband*[c] to record their physiological data (e.g., electrodermal activity and heart rate). All sensors can be controlled using a central synchronization platform recording data locally and/or to a cloud repository. In this study, all team simulations were video recorded locally using a 180° camera.

Study 1: Spatial Analytics for Team Assessment

This study illustrates how the framework was applied to support teachers' teamwork assessment by modeling students' positioning data.

Input: Quantitative Ethnographic Codes

We first captured the combined expertise about teamwork and task performance from the simulation task designer, two experienced nursing educators, and two members of our data science team. These experts designed a rule-based classification template that maps from students' indoor positioning data to meaningful *spatial constructs* related to team behaviors that are likely to occur in certain areas of the learning space Figure 2 (left). The *primary task* spaces were those in which students worked with the medically deteriorating patient [see red encircled areas in Figure 2 (left)]. The meaning of team members' presence in spaces encircled in blue related to procedural tasks that were less critical *secondary tasks*.

The social aspects of students' spatial behaviors were also embedded into the mapping template. The template encodes a potential instance of collaboration if the proximity between two students was within 1 m for more than 10 consecutive seconds (to prevent misidentifying unintended collocation, for example, when students passed by each other). Four different types of spatial constructs were classified based on

[a]Yarn-Sense https://teamwork-analytics.github.io/yarn-sense/

[b]Pozyx https://www.pozyx.io

[c]Empatica-E4 http://lp.empatica.com/e4-wristband

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

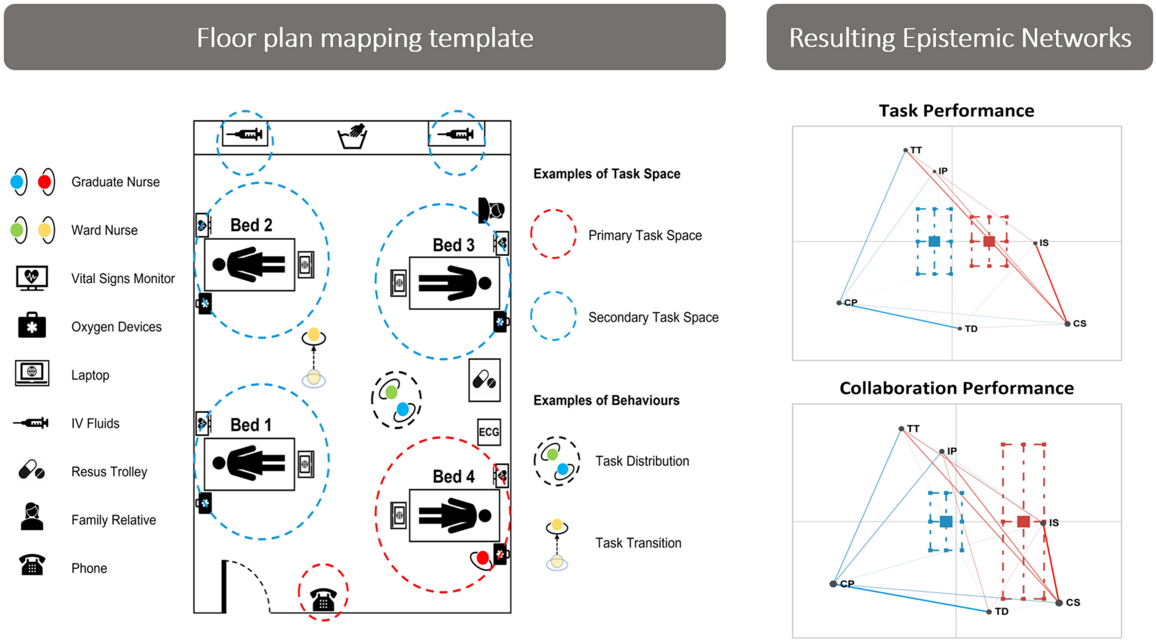

FIGURE 2. Left: Floor plan of the simulation space as a template for mapping from x-y coordinates of team members to spatial constructs. Right: Epistemic network (EN) comparison plots of all the low-performing (red edges) and high-performing (blue edges) teams that participated in the study, regarding their task (top) and collaboration (bottom) performance. Nodes represent spatial constructs: CP/CS = presence of two or more nurses in areas associated with the primary or secondary tasks, respectively; IP/IS = physical presence of a single nurse in areas associated with the primary or secondary task, respectively; TD = two or more nurses distributing tasks; TT = a single nurse transitioning from a task area to another.

the combination of task spaces and collaboration behaviors. *Collaborate_Primary (CP)* and *Independent_Primary (IP)* represent two or more students working collaboratively and each student working individually on the primary task, respectively. In contrast, *Collaborate_Secondary (CS)* and *Independent_Secondary (IS)* represent two or more students working collaboratively and each student working individually on the secondary task, respectively.

Another two spatial constructs were identified through the combined efforts of researchers and teachers. Based on teamwork theory,[2] an essential stage of effective teamwork involves team members discussing and distributing the responsibilities of different tasks. The teachers suggested that this *Task_Distribution (TD)* behavior could be captured if two or more students were collocated outside of the task spaces, where they were more likely to distribute responsibilities instead of collaborating on specific tasks [see an example in Figure 2(left)]. Teachers also suggested the importance of monitoring students' *Task_Transition (TT)* behaviors, explaining that when team members move from one task to another frequently it commonly indicates feeling overwhelmed and insecure.

Expert-Driven Team Modelling

We built five ML models, namely, logistic regression, support vector machine, random forests, *k*-nearest neighbors, and multilayer perceptron artificial neural networks, to predict teachers' task and collaboration performance assessment scores, differentiating low-performing from high-performing teams. The aggregates of the spatial constructs over the whole team activity were used as the input data, e.g., the percentage of time each team demonstrated in each of the aforementioned six spatial constructs. Further details about the dataset sampling and the predictive modeling can be found in Yan et al.[18]

Behavioral differences between low-performing and high-performing teams were revealed through ENs.[12] These calculate the sequential occurrence of the spatial constructs for each team during the whole simulation. For example, two students moving together from the primary to the secondary task would be marked as a co-occurrence of the codes CP and CS, and this would contribute to a thicker edge between such nodes in the EN. Visualizations in Figure 2 (right) represent differential networks between high-performing (blue) and low-performing (red) teams (i.e., red links show where low-performing teams had more co-occurrences than the high-performing teams, and vice versa).

Outputs: Exploratory Interfaces

Concerning the predictive models, these demonstrated an acceptable level of accuracy (best AUC $= 0.74$) in distinguishing between low-performing and high-performing teams regarding their collaboration performance based on their spatial behaviors. Although such accuracy still requires improvements before being suitable for practical adoption, in a post-hoc focus group, teachers already envisioned its potential role as a confirmatory tool to boost the confidence of beginner and inexperienced teachers regarding their assessment on *collaboration performance*, a common highly subjective evaluation.

Concerning the ENs, the teachers endorsed these as supportive tools for post-hoc reflective debriefs. The resulting visualizations Figure 2(right) showed clear differences in the strategies of low-performing and high-performing teams. In particular, low-performing teams prioritized the secondary task, working either collaborative or individually and rarely engaged in task distribution (see thicker red lines linking various constructs, especially the nodes related to the secondary task—IS and CS). In contrast, the high-performing teams focused mainly on working collaboratively on the primary task and frequently engaged in task distribution (see blue lines linking various constructs mainly with the nodes CP and TD).

We do not consider the ENA examples to be polished user interfaces suitable for most teachers, and indeed, our other work has shown the potential for confusion in interpreting them.[13] The fact these teachers were able to comprehend the ENs, despite their inherent complexity, potentially reflects the benefits of their early involvement, contributing their first-hand expertise to give meaning to data. Such involvement could empower them with the knowledge to comprehend the ENs and enhance their confidence in using these visualizations. The teachers explained that they would like to use these visualizations to encourage reflective practices in both low-performing and high-performing teams. For the former, this would involve contrasting their incorrect prioritization strategy with the correct ones and guiding them through the underlying reasons behind their strategies. For the latter, the ENs can also act as visual evidence to encourage them to reflect on the elements that helped them succeed.

Study 2: Data Stories for Students

This study illustrates the framework's application to automatically generate data stories to support students' reflection on their team activity.

Input: Human Logging

To provide context for the analysis, task designers and educators defined five *critical actions* that an effective team should accomplish promptly, namely: 1) administering oxygen after the patient had respiratory depression; 2) assessing vital signs every 5 min; 3) ceasing PCA (patient-controlled analgesia) after the patient had altered conscious state; 4) activating Medical Emergency Team (MET) calls after the patient started deteriorating; and 5) administering Naloxone afterwards. While these actions can be performed by individual team members, it is the effective coordination, communication, and timeliness of such actions that enable the differentiation between high and low performing teams. These actions, which were too complex for automated logging, were logged by a researcher either in real time or post-hoc using a web-based logging interface that injected the logs as a new column per team member into the multimodal matrix.

Expert-Driven Team Modeling

Figure 3(a) shows a storytelling editor that an educator or coach can use to translate their pedagogical intentions into rule-based algorithms that the system can use to render visual data stories [see Figure 3(b)]. The user can select the type of story from a predefined set of options (e.g., feedback based on interpersonal proximity or on the timeliness of actions). Then, the user can select the actions, phases in the team activity or constructs that need to be considered in the modeling. For example, Figure 3(b) shows an example of a data story focused on interpersonal proximity (i) while the team of nurses were engaged on the primary task (being near bed 4, the patient that required more attention) during the handover (ii). The users can also select the team member(s) that they want to provide feedback to and define a personalized feedback message that would appear in the output of the modeling [see Figure 3(iii) and (iv)].

The storytelling editor is used as input for the team modeling infrastructure. Team data can be imbued with contextual meaning using such inputs from team experts, educators, or coaches. For this study, we instantiated a multimodal matrix using the critical actions logged by the human observers and the positioning data captured by the positioning sensors. Part of the matrix is used to identify team successes and errors based on the logged critical actions. Three types of errors were automatically identified for this simulation. *Sequence errors* are flagged if the team performed critical actions in the wrong order, for example, if they forgot to perform a vital signs assessment. *Timeliness errors* are identified if team members

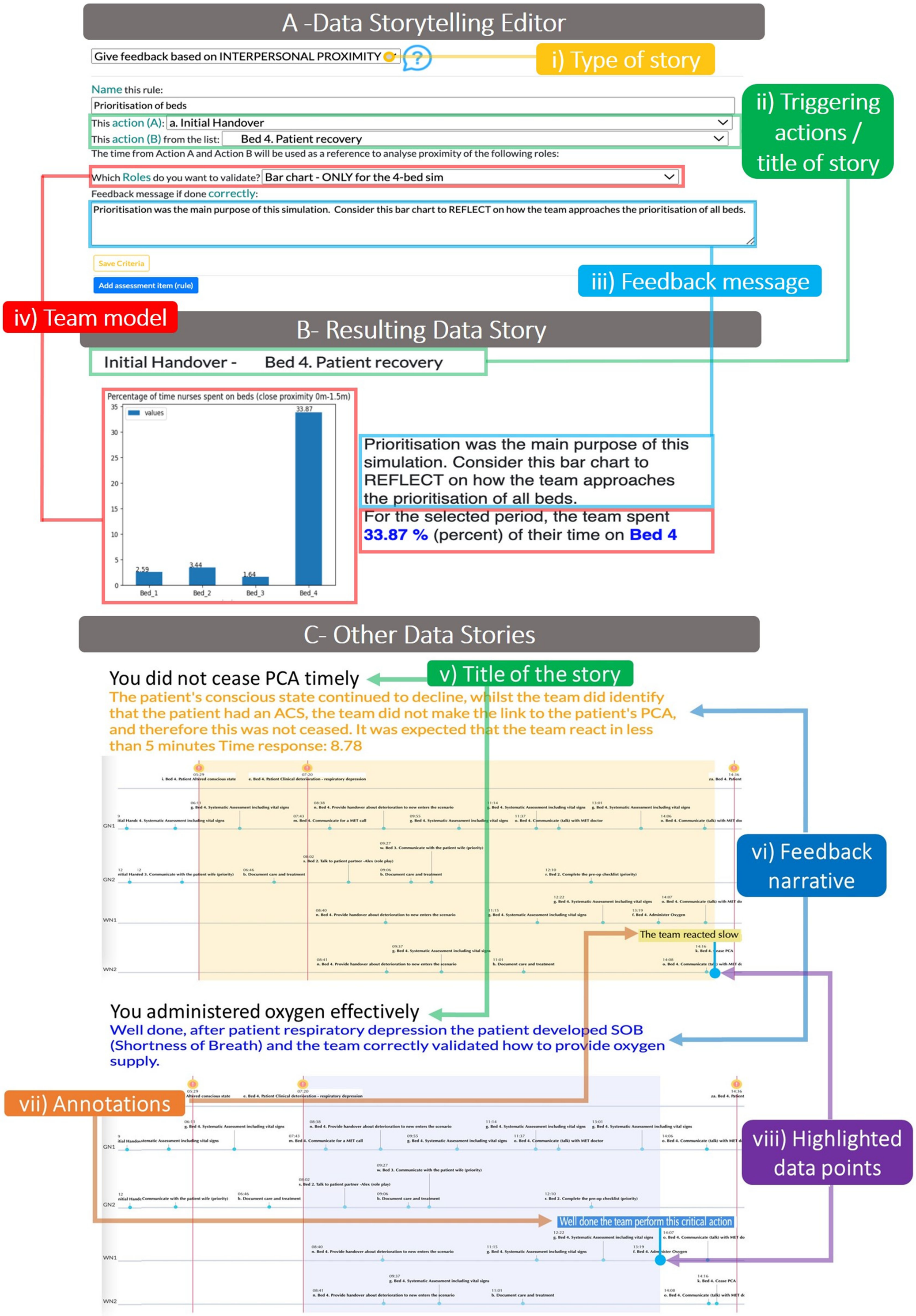

FIGURE 3. A: Web-based storytelling editor that an educator or coach can use to translate their pedagogical intentions into rule-based algorithms that the system can use to render web-based, visual data stories. B: Data story example associated with the copresence of team members near the four patient beds. C: Other example data stories that communicate an error (top) and a correct action by the same nursing team (bottom).

reacted slowly according to healthcare guidelines and the educator's experience, for example, if they took more than 5 min to stop the PCA. An error related to *frequency of actions* is defined by calculating the timestamp difference between two critical actions that are meant to be repeated, for example, assessing the patient's vital signs every 5 min. In this case, the columns in the matrix correspond to all the actions that team members are expected to perform, and rows indicate the time (one row per second) when each action was performed by a particular team member. Thus, the rules created by the educator using the storytelling editor assess the order, timeliness, and/or frequency of the selected actions.

The remaining columns in the matrix were defined to model low-level positioning data into higher order proximity constructs. Considering the construct of *copresence in interactional spaces*, it was possible to validate the proportion of time that nurses spent in close proximity (within a distance of 0m–1.5m[13]) to 1) the four patients; 2) the patient in the primary bed; and 3) other team members. Here, the columns of the matrix corresponded to the detected proximity among team members and patients; and rows captured the time when such co-presence was detected. Educators can configure the kind of feedback to be generated for each case as illustrated in Figure 3, and detailed next.

Outputs: Explanatory Interfaces

Figure 3(b) presents an example of a data story emphasizing that a team was highly effective in prioritizing the patient in bed 4. Figure 3(c) presents two additional data stories, one highlighting an error that the team made (top) and another showing a critical action performed correctly (bottom). Both interfaces have a self-explanatory title headlining the primary message (Figure 3(v); elaborated by the feedback narrative crafted by the educator to explain strengths/weaknesses (vi). Moreover, the data stories both *highlight* (vii) and *directly annotate* (viii) relevant data points, in this case, the critical actions that were used by the rule-based algorithm configured by the educator to create each story. In prior works, we have evaluated high-fidelity prototypes of these explanatory visual analytics, and gained evidence that students were positive about the potential of these data stories.[19] In our most recent evaluation with 41 nursing students who engaged with a fully automated version described here, we corroborated that students used the data stories as evidence to reflect on their prior performance. For example, some students realized that the team did not react as promptly as they recalled. The data stories also supported them to formulate strategies to address weaknesses in their communication and performance.

In addition to Studies 1 and 2, empirical evaluations of user interface prototypes show that data storytelling can drive visual attention to the key details in a visualization;[15] that visualizations of nurses' positioning provided educators with insights that helped them differentiate good from poor teams[13]; and that the data stories [Figure 3(c)] helped students to identify misconceptions, strategize improvements to address errors, and reflect on their arousal levels.[19]

DISCUSSION

Our experiences in designing and deploying the two studies in our case study have brought to light three overarching themes related to improving future human-centered teamwork analytics.

Synergy Between Human Intelligence and AI

In Ben Shneiderman's[17] view, *"computer scientists should build devices to enhance and empower—not replace humans."* From a human-centered AI perspective, our HuCETA framework blends the strengths of humans, sensing technologies, and AI. This can be seen as an example of the emerging research area called hybrid human-artificial intelligence, which acknowledges that both human intelligence and AI each have their own strengths and limitations.[20] In our approach, educational stakeholders retain supervisory control of the analytics to imbue data with meaning, in order to create data stories that end-users can actually understand, and connect meaningfully to the coaching or learning purposes for the teams. Hence, removing full autonomy of the system is deliberate.

Through our QE and data storytelling approaches, we make the most of the key capabilities of computers, which are efficient in continuously gathering large amounts of data via sensors, and effective in discovering patterns from large datasets. These are harnessed to amplify the capabilities of educators or coaches by providing ways for them to 1) make their assumptions, inferences and interpretations about teamwork explicit and trackable as streams of multimodal data (first HuCETA component); 2) indicate to the computer how to interpret sensor data by applying QE principles (second component); and 3) assist the computer to present analytics/AI-enabled feedback as intelligible data stories (third component).

The Potential of Data Storytelling

We received consistent positive feedback from both educators and students, who recognized the value of capturing digital traces of collocated activity and rendering it visible as explanatory, data stories for the purpose of provoking reflection on team simulation experiences.[19] They also recognized the value of capturing both objective (via sensors) and subjective (via human observers) evidence about teamwork, in contrast to the current educational situation, where reflective team debriefs dependent on the expert observations by coaches or educators (often split over multiple teams), and the (often stressed, and always biased) memories of students.

While some participants in our studies were interested in the potential of the approach to assess team performance *summatively*, others expressed concerns about the risk of false negatives in the data stories. Multimodal data stories can certainly introduce bias in the way narrative is added to the visualizations [component (iii) in the framework], how rules to map low-level data into higher order constructs are created [component (ii)], and how accurately sensors or humans capture data [component (i)]. Although false negatives were not identified in any of our prototypes (partly because of the close involvement of the research team), we concur with a more cautious view focused on formative feedback to guide deeper reflection and dialog, in which humans determine the ultimate meaning and actions to perform. Fully automated grading is significantly beyond the current maturity of this infrastructure.

Advancing Team Science Versus Supporting Users

The proposed framework (instantiated in many explanatory interfaces, such as the data stories presented in Study 2) can help educators and team members test and reflect on teamwork constructs (defined as hypotheses, assumptions, or beliefs). As we have emphasized, however, more complex exploratory interfaces that invite users to explore the data can offer the potential to assist researchers in testing theories about what makes teamwork effective, and can support educators' and coaches' inquiry into their own practices, as illustrated in Study 1. The HuCETA framework, and the open source, modular architecture enabling its implementation, offer a new form of 'team science research platform', which we hope may accelerate research in embodied teamwork, whether for exploratory, descriptive analyses, or for the testing of hypotheses.

CONCLUSION

Both classic and current research on artificial and human intelligence has envisioned the 'expansion' of human capabilities as a coevolving process of people, computers, and networks to approach complex tasks. The HuCETA framework contributes to this vision by enabling educators and researchers to 1) log key team events that exceed the automated sensing infrastructure; 2) convert from complex multimodal data to meaningful higher order team constructs; and 3) drive the explanatory power of end-user interfaces that can communicate team insights. We illustrate how this can be operationalized in the context of undergraduate nursing simulation. Moreover, our analytics infrastructure, tuned to the needs of different contexts, can be transferable to other learning/training environments and industries. Our research therefore has the potential to strongly impact on vocational and higher education communities. The framework crafts the building blocks for facilitating services in physical learning spaces once thought to be confined to the realm of online learning. These include the provision of personalized feedback, and supporting decision-making processes that can enrich collocated training experiences and make them more effective.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council under Grant DP210100060. The work of Roberto Martinez-Maldonado's was supported by Jacobs Foundation. This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Monash University Human Research Ethics Committee (MUHREC) under Application No. 28026, and performed in line with The Australian Code for the Responsible Conduct of Research.

REFERENCES

1. S. M. Nisha and V. Rajasekaran, "Employability skills: A review," *IUP J. Soft Skills*, vol. 12, no. 1, pp. 29–37, 2018.
2. E. Salas, D. E. Sims, and C. S. Burke, "Is there a "big five" in teamwork?," *Small Group Res.*, vol. 36, no. 5, pp. 555–599, 2005, doi: 10.1177/1046496405277134.
3. K. Knipfer, M. Prilla, T. Herrmann, and U. Cress, "Computer support for collaborative reflection on captured teamwork data," in *Proc. Comput. Supported Collaborative Learn.*, 2011, pp. 938–939.

4. A. F. Wise, S. Knight, and S. Buckingham Shum, *Collaborative Learning Analytics*. Berlin, Germany: Springer, 2021, pp. 425–443.
5. A. Mohan, H. Sun, O. Lederman, K. Full, and A. Pentland, "Measurement and feedback of group activity using wearables for face-to-face collaborative learning," in *Proc. IEEE Adv. Learn. Technol.*, 2018, pp. 163–167.
6. Á. Fidalgo-Blanco, M. L. Sein-Echaluce, F. J. García-Peñalvo, and M. Á. Conde, "Using learning analytics to improve teamwork assessment," *Comput. Hum. Behav.*, vol. 47, pp. 149–156, 2015, doi: 10.1016/j.chb.2014.11.050.
7. S. C. E. Salas, P. R. Stevens, J. Gorman, N. J. Cooke, S. Guastello, and A. A. von Davier, "What will quantitative measures of teamwork look like in 10 years?," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 59, no. 1, pp. 235–239, 2015.
8. A. Penzkofer, P. Müller, F. Bühler, S. Mayer, and A. Bulling, "Conan: A usable tool for multimodal conversation analysis," in *Proc. ACM Int. Conf. Multimodal Interact.*, 2021, pp. 341–351.
9. J. T. Paige, C. L. Rogers, K. E. Kerdolff, D. D. Garbee, L. S. Bonanno, and Q. Yu, "Conceptualizing a quantitative measurement suite to evaluate healthcare teams," *Simul. Gaming*, vol. 53, no. 1, pp. 75–92, 2022.
10. D. Wang et al., "Using humans as sensors: An estimation-theoretic perspective," in *Proc. IEEE Symp. Inf. Process. Sensor Netw.*, 2014, pp. 35–46, doi: 10.1109/IPSN.2014.6846739.
11. S. Dumais, R. Jeffries, D. M. Russell, D. Tang, and J. Teevan, "Understanding user behavior through log data and analysis," in *Ways of Knowing HCI*. New York, NY, USA: Springer, 2014, pp. 349–372, doi: 10.1007/978-1-4939-0378-8_14.
12. D. W. Shaffer, *Quantitative Ethnography*. Charlottesville, VA, USA: Cathcart Press, 2017.
13. G. Fernandez-Nieto et al., "What can analytics for teamwork proxemics reveal about positioning dynamics in clinical simulations?," *Proc. the ACM Hum.-Comput. Interact.*, vol. 5, no. 185, pp. 1–24, 2021, doi: 10.1145/3449284.
14. V. Echeverria, R. Martinez-Maldonado, and S. Buckingham Shum, "Towards collaboration translucence: Giving meaning to multimodal group data," in *Proc. ACM Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–16, doi: 10.1145/3290605.3300269.
15. V. Echeverria et al., "Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling," *J. Learn. Anal.*, vol. 5, no. 3, pp. 72–97, 2018, doi: 10.18608/jla.2018.53.6.
16. M. N. Giannakos, K. Sharma, I. O. Pappas, V. Kostakos, and E. Velloso, "Multimodal data as a means to understand the learning experience," *Intl. J. Inf. Manage.*, vol. 48, pp. 108–119, 2019, doi: 10.1016/j.ijinfomgt.2019.02.003.
17. B. Shneiderman, "Human-centered AI," *Sci. Technol.*, vol. 37, no. 2, pp. 56–61, 2021.
18. L. Yan, R. Martinez-Maldonado, B. Gallo Cordoba, J. Deppeler, D. Corrigan, and D. Gašević, "Mapping from proximity traces to socio-spatial behaviours and student progression at the school," *Brit. J. Educ. Technol.*, in press, Aug. 2022, doi: 10.1111/bjet.13203.
19. G. M. Fernandez-Nieto et al., "Storytelling with learner data: Guiding student reflection on multimodal team data," *IEEE Trans. Learn. Technol.*, vol. 14, no. 5, pp. 695–708, Oct. 2021, doi: 10.1109/TLT.2021.3131842.
20. L. Chen, H. Ning, C. D. Nugent, and Z. Yu, "Hybrid human-artificial intelligence," *IEEE Comput.*, vol. 53, no. 8, pp. 14–17, Aug. 2020, doi: 10.1109/MC.2020.2997573.

VANESSA ECHEVERRIA is a research fellow at Monash University, Clayton, VIC, 3800, Australia, and an associate professor at Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador. Her research focuses on multimodal learning analytics. She is the corresponding author of this article. Contact her at vanessa.echeverria@monash.edu.

ROBERTO MARTINEZ-MALDONADO is a senior lecturer at Monash University, Clayton, VIC, 3800, Australia, and a Jacobs Foundation Research Fellow. His research encompasses creating data-intensive tools to support people to collaborate effectively. Contact him at Roberto.MartinezMaldonado@monash.edu.

LIXIANG YAN is currently working toward the Ph.D. degree at the Monash University, Clayton, VIC, 3800, Australia. His research encompasses applying machine learning to extract meaningful educational insights from physical and physiological traces. Contact him at Lixiang.Yan@monash.edu.

LINXUAN ZHAO is currently working toward the Ph.D. degree at Monash University, Clayton, VIC, 3800, Australia. His research encompasses analyzing audio data to generate supportive analytics automatically. Contact him at linxuan.zhao@monash.edu.

GLORIA FERNANDEZ-NIETO is a research fellow at Monash University, Clayton, VIC, 3800, Australia. Her research investigates the generation of automated data storytelling interfaces. Contact her at gloria.m.fernandeznieto@student.uts.edu.au.

DRAGAN GAŠEVIĆ is a professor in the Faculty of Information Technology and the director of the Centre for Learning Analytics at Monash University, Clayton, VIC, 3800, Australia. His research is focused on analytics for self-regulated and collaborative learning. Contact him at dragan.gasevic@monash.edu.

SIMON BUCKINGHAM SHUM is a professor with University of Technology Sydney, Ultimo, NSW 2007, Australia, where he is the director of the Connected Intelligence Centre. His research investigates how software makes thinking visible, and analytics to improve feedback for learning. Contact him at Simon.BuckinghamShum@uts.edu.au