Storytelling With Learner Data: Guiding Student Reflection on Multimodal Team Data

Gloria Fernández-Nieto, Vanessa Echeverria, Simon Buckingham Shum, Katerina Mangaroska, Kirsty Kitto, Evelyn Palominos, Carmen Axisa, and Roberto Martinez-Maldonado

***Abstract*—There is growing interest in creating learning analytics feedback interfaces that support students directly. While dashboards and other visualizations are proliferating, the evidence is that many fail to provide meaningful insights that help students reflect productively. The contribution of this paper is qualitative and quantitative evidence from two studies evaluating a multimodal teamwork analytics tool in authentic clinical teamwork simulations. Collocated activity data is rendered to help nursing students reflect on errors and stress–related incidents during simulations. The user interface explicitly guides student reflection using data storytelling principles, tuned to the intended learning outcomes. The results demonstrate the potential of interfaces that "tell one data story at a time", by helping students to identify misconceptions and errors, think about strategies they might use to address errors, and reflect on their arousal levels. The results also illuminate broader issues around automated formative assessment, and the intelligibility and accountability of learning analytics.**

***Index Terms*—Collocated spaces, feedback, guidance, multimodal learning analytics, reflection, visualization.**

I. Introduction

The provision of high–quality actionable feedback can have a strong positive effect on student reflection, performance, and achievement [1]. High–quality feedback has been defined as any piece of information that can help students take actions to reduce the discrepancy between teacher's pedagogical intentions, student's learning goals, and their actual performance [2]. What teachers commonly do when providing feedback to students is communicate insights obtained from formative or summative assessments (e.g., students' assignments or exams). Yet, delivering actionable feedback to students is challenging, especially in large classes [3] or in physical classrooms in which it is difficult to follow the progress of all the students at the same time [4].

Manuscript received January 21, 2020; revised June 04, 2021; accepted October 23, 2021. Date of publication TBD; date of current version October 25, 2021. *(Corresponding author: Gloria Fernandez-Nieto.)*

G. Fernandez-Nieto, V. Echeverria, S. Buckingham Shum, K. Kitto, E. Palominos, and C. Axisa are with the Connected Intelligence Centre, University of Technology Sydney, Ultimo NSW 2007, Australia (e-mail: gloria.m.fernandeznieto@student.uts.edu.au, vanessa.i.echeverriabarzola@student.uts.edu.au, simon.buckinghamshum@uts.edu.au, kirsty.kitto@uts.edu.au, evelyn.m.palominosletelier@student.uts.edu.au, carmen.axisa@uts.edu.au)

K. Mangaroska is with the Department of Computer Science, Norwegian University of Science and Technology, Trondheim, NO-7491, Norway (e-mail: katerina.mangaroska@ntnu.no)

R. Maldonado Martinez is with the Faculty of Information Technology, Monash University, Clayton VIC 3800, Australia (e-mail: roberto.martinezmaldonado@monash.edu)

There is a growing interest in creating Learning Analytics (LA) *student–facing* interfaces that can directly support students by delivering automated or semiautomated feedback. Typically, these take the form of dashboards, visualizations, or reports in which digital traces and analytics outputs are aggregated and presented [5]. However, recent literature reviews and empirical studies report that most student–facing LA interfaces have serious limitations, such as showing visualizations that are difficult to understand by nondata experts [6], lack of effectiveness in communicating insights [7], absence of meaningful information [8], and failing to make educationally meaningful impact [9],[10].

In order to improve the level of support for interpreting student–facing LA interfaces, this paper investigates how the application of a *data storytelling* (DS) approach can be useful in *guiding* students' reflection. We investigate how the *learning design* of a course can be leveraged to build interfaces that help students reconstruct stories about their behavior, that make sense in the context of the academic's learning intentions at a time. DS is a process of translating data analysis outputs into terms that can be understood by people without formal data science training, to influence a subsequent decision or action [11]. The implementation of DS in LA aims to *guide* students' reflection on evidence by reducing the complexity of data representations in order to reduce the chances of erroneous interpretation [12].

Our contribution is to present two empirical, qualitative studies that investigate the potential of guiding students' reflection on their activity data through learner *data stories*. More specifically, we investigate: i) if, and to what extent, students consider that data stories add some value to their reflections; ii) the envisaged potential uses of the design prototypes; iii) the perceived impact on students' accountability; and iv) the implications of revealing to students the algorithms used to generate the data stories. A total of 39 undergraduate nursing students, participated in two authentic simulation–based teamwork activities, in which students had to provide care to a simulated patient manikin (see Fig. 1, left). Two *DS prototypes* (visualizations enhanced with visual and narrative elements, e.g., Fig. 1, right) were provided for students to reflect upon insights semiautomatically identified from *multimodal data* captured during the team activity: logged actions and physiological data sources. The results show the potential of LA interfaces that *tell one story at a time* to guide and prompt student reflection, while also highlight challenges in terms of automated assessment and student accountability.

II. Background and Literature Review

Three research areas are relevant to the studies presented in this paper: i) foundations of guidance in visualization and DS; ii) empirical work in LA aimed at guiding students' data interpretation; iii) and the particular challenges of communicating insights from multimodal data.

A. Guidance and Data Storytelling

Students can generally be considered *nondata experts* (users who bring little or no data analysis expertise [13]). Empirical research has shown that even Science, Technology, Engineering, Arts and Mathematics (STEM) university students commonly face difficulties to interpret and understand charts [14]. Moreover, students can also be considered as casual users of LA systems [15] since firstly, LA tools commonly are a novel form of tools for most students, and secondly, LA tools will typically be used sporadically to reflect on work [16].

Research in information visualization (InfoVis) argues that *guidance* should be provided to support casual users, or those users with low analysis expertise, to interpret data visualizations [17]. *Guidance* can be defined as a computer–assisted processes aimed at narrowing the gap of data interpretation and exploration encountered by end users [18]. Schultz *et al.* [17] described different ways in which this concept can be materialized, such as by enhancing charts using visual cues, allowing users to select from various visualization techniques, and guiding users through prescriptive data exploration workflows or via *DS*.

DS is a suite of information design and "compression" techniques to help an audience effectively understand what is important in a visualization [11], communicating key messages clearly and effectively through the combination of data, visuals and narratives [19]. Ryan [11] and Knaflic [20] distilled the following DS principles:

DS1. DS is goal oriented. A data story should have a very specific goal which enables the identification of the data that should be visually emphasized. **DS2. A data story should rely on a suitable chart type**. Some charts work better for certain purposes. **DS3. A data story should be stripped down first**. Decluttering is a critical step to reduce the complexity of the visual representation. This involves removing headers, borders, grids and data points that are not central to the story. **DS4. A data story should guide attention**. Only data points that are critical to communicate a story should be emphasized using visual or narrative elements. This can be achieved by 1) *adding enhancements* such as arrows, lines, symbols, or enclosures; 2) changing *color*, *contrast* or *thickness*; and 3) *annotating salient data* features and *adding headline titles* that summarize the message of the story. These DS principles will be illustrated through the prototypes in Section IV, below.

B. Visualization in Learning Analytics

A core part of LA research has explored the opportunities and challenges of data visualizations and dashboards for educational purposes. In a recent review, Jivet *et al.* [21] reported that most dashboards lack mechanisms to support students to understand how data visualizations could help them *achieve specific learning goals*, emphasizing the need for designing LA dashboards aligned with the instructional design. In a similar way, Bodily *et al.* [22] found that only 17% (16 out of 93) of the systems reviewed included elements to *help students know what to do* based on the data shown to them. Likewise, Matcha *et al.* [9] argued that there exist some weaknesses in most of the current LA dashboards in *helping students interpret the data* and in communicating meaningful insights. In addition, Corrin [6] reported, through a qualitative study, that students found it hard to translate the "feedback" received from a dashboard into *strategies that could impact their performance*. That paper suggested that guidance is needed to give meaning to the data, and that teachers should have a key role in its configuration.

Various attempts have been made to provide such guidance. For example, visualizations have been enhanced with text narratives to further explain the meaning of datapoints [23] (for students); or by *highlighting critical pieces of information* such as scores [24] (intended for students) or patterns found in discourse data [25] (for students). These visualization enhancements are related to the concept of DS.

DS has only started to be recognized in LA. For instance, Chen *et al.* [26] proposed an approach for highlighting and annotating video elements and slideshows to present visual data for teachers to understand students' progress. Echeverria *et al.* [27] showed how visualizations enhanced with DS elements can drive the focus of attention of teachers and lead to deeper reflections on students' data. The same authors proposed that the teacher's instructional design should drive the visualization design [12], but to date have only reported lab–based trials with teachers in experimental settings. In sum, several authors have identified challenges in guiding students to interpret LA interfaces meaningfully. Our work builds on initial attempts to apply DS principles to guide interpretation of LA visualizations ([4], [12], [26], [27]). We build on the conceptual model proposed by Martinez–Maldonado *et al.* [28], who segmented the multimodal user interface into layers of information. While this and the previous works discussed above examined teachers engagement with the design, it is also critical to understand how students can benefit directly from multimodal data ([12], [27], [28]). We address this gap by reporting empirical findings from two authentic qualitative studies with nursing students using two high–fidelity prototypes, presenting multimodal learning stories for them to reflect on their errors and physiological arousal during medical ward simulations.

C. Multimodal Learning Analytic Interfaces

Data collection in physical learning spaces (such as classrooms and laboratories) is becoming feasible due to advancements in sensing and computer vision technologies. Evidence about different modalities of students' interaction (e.g., posture, positioning, and speech), and features that are commonly less visible (e.g., electrodermal activity and pulse) can be digitally captured via a combination of sensors, interactive devices, and observations. This is enabling the creation of

Fig. 1. *Left:* teamwork activity traces captured via a combination of sensor signals, system and human logs. *Right:* partial view of a storytelling prototype enhanced with (A) annotations, (B) highlighted data points, (C) colored regions, and (D) the rules triggering the feedback annotations.

new ways to study learning in physical contexts such as comprehending how collocated group behaviors connect to learning outcomes [29] or finding patterns that can be used to personalize instruction [30]. Multimodal learning analytics (MMLA) is an emerging subarea within LA focused on supporting learning beyond the clickstreams and keystrokes of conventional personal computing. However, analyzing multiple data streams brings further challenges related to data integration, interpretation, and visualization [29]. MMLA tools can very easily generate complex interfaces, which explains, in part, the dearth of MMLA user interfaces suitable for teachers and students. For example, Echeverria *et al.* [31] designed four visualizations, each related to one modality (speech, arousal, positioning, and logged actions), but did not fuse these into a single interface to facilitate reflection. This was attempted by Ochoa *et al.* [32], who visualized logs of students' activity around a tabletop. Multimodal data included logged actions, verbal participation, gaze direction, and emotional traits. Initial teacher feedback was positive, but the prototype was not evaluated in authentic contexts with students. Preliminary work by Vujović *et al.* [33] investigated how to compress arousal and noise information during meetings, but this was for interpretation by educational researchers, not teachers or students. In short, the current interpretation challenges found in many student–facing LA interfaces, in general, are even more profound for the case of MMLA innovations which commonly deal with complex, heterogeneous data streams. There is a small but growing interest in creating MMLA interfaces for nondata experts [31]. The work presented in the next sections addresses the lack of guidance for students to interpret and reflect upon their own multimodal data in the context of nursing simulation.

III. Learning Task Design

Healthcare simulation is a pedagogical approach that uses a constructivist learning model to provide students with opportunities to experience teamwork and patient situations without compromising the care of real patients [34]. Simulations often start with a description of learning goals, followed by the simulation itself, concluding with a debrief aimed at provoking students' reflection on performance and errors made. Although making errors in simulations can provoke negative feelings in some students, current studies suggest that addressing errors constructively can aid learning [35]. Although video–based products to support this reflection exist, they are commonly impractical for class use, resulting in students rarely using such evidence to inform reflection [36].

Various simulations are conducted as part of the curricula of the undergraduate Nursing program of the University of Technology Sydney. Simulation classrooms are equipped with 5–6 beds with a patient manikin on each. Students are commonly organized in teams of 4–5 members, to look after a patient each in a hypothetical scenario. The two studies (Studies 1 and 2) discussed in this paper focused on two simulations (Sims 1 and 2) conducted in regular classes of the course *Integrated Nursing Practice* in 2019 (semesters 1 and 2 respectively). The next subsections describe the learning goals of each simulation and the multimodal data that was collected.

A. Sim 1—Surgery Recovery

Sim 1 was run in 4 classes by the same teacher (the course coordinator). A total of 19 students in their third year (all females, one team in each class) volunteered to participate in the study and for their data to be recorded. The goal of the sim was to provide care to a patient after abdominal surgery. Students in each team played the roles of team leader, registered nurses (RN1, RN2), scribe (RN3), and the patient (not tracked, see Fig. 1). According to the assessment criteria set by the teacher, a highly effective team should have performed the following 5 *actions*: i) assess vital signs every 10 minutes; ii) check fluids, suction secretions, perform head tilt/chin, or add oxygen therapy after the patient presents breathing obstruction; iii) administer fentanyl within 10 minutes after the patient complains of abdominal pain; iv) administer a second bolus of fentanyl after the patient complains of severe abdominal pain; and v) administer ondansetron within 10 minutes after the patient experiences nausea.

B. Sim 2—Allergic Reaction

Sim 2 was run in 5 classes taught by 3 teachers (including the same course coordinator). A total of 25 students in their third year (21 females and 4 males) volunteered to participate. The aim of this simulation was to help nurses learn how to react when a patient is having an allergic reaction to some medication. Similar roles to those in Sim 1 were allocated to members of each team. These assessment criteria imply that a highly effective team should complete 6 specific actions throughout the simulation: i) perform an initial set of vital signs measurements, after the teacher reads the initial handover; ii) administer the intravenous (IV) antibiotics; iii) perform another set of vital signs measurements after the patient complains of chest tightness; iv) stop the IV antibiotic after the patient reacts with chest tightness; v) perform an electrocardiogram (ECG) after the patient complains of chest tightness; and (vi) call the doctor after stopping the IV antibiotic.

C. Data Collection

Students' *physiological data* was captured through (Empatica e4) wristbands. These record electrodermal activity (EDA) at 4 Hz. Some student *actions*, such as vital signs assessment, were detected by the mid–fidelity manikin (Laerdal Nursing Anne). Other actions performed by each student (e.g., stopping IV antibiotic, writing on charts, and calling the doctor) were manually logged by an observer (a researcher in studies reported in this paper), but it could also be a student using a web application, or a high–fidelity patient manikin not available in all the classrooms of the hosting university. Data streams were synchronized and down sampled at 1 Hz. Although additional data from each student was captured, such as indoor positioning, wrist acceleration and blood volume, only action logs, and physiological data were used in the studies reported in this paper, as requested by the teacher. All sessions were video recorded.

IV. System Design

This section describes the modelling techniques applied to map from low–level action logs and physiological data to data stories.

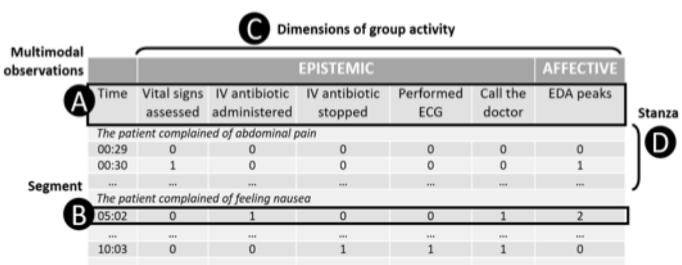

Fig. 2. Application of the Multimodal Matrix modelling technique on the data of one team member. (A) Multimodal observations, (B) Segments, (C) Dimensions of group activity, and (D) Stanza

A. Multimodal Data Modelling

The first step in crafting data stories was to convert the data into a meaningful multimodal data structure. For this, we used the *Multimodal Matrix (MM)* described by Echeverria *et al.* [31]. Fig. 2 shows a simplified representation of the modeling performed on the data of one student. The MM is a (m–by–n) data structure in which each data modality m is coded into n columns of the matrix which are called *multimodal observations*. For example, Fig. 2A shows the critical actions (e.g., vital signs assessment and administer IV antibiotic) and physiological arousal of one team member represented as columns in the matrix. *Segments* (m rows) are the smallest units of meaning considered for analysis and contain instances of group behaviors. For time series data, such as physiological data, each row can represent a time window (e.g., one second in our studies) of the team activity (Fig. 2B). This way, the content of each cell becomes an indicator of a particular aspect of a team member. In our study, we represented the absence (missing actions and slow responses) or presence of certain actions *(errors)* as binary flags (1, 0). To register the effective observations into the matrix two steps are needed:

- The recorded EDA, was passed through the EDA Explorer [37] algorithm to detect peaks in the whole dataset. A peak is a sudden change in EDA for a student, non baseline–based [38].
- The count of *arousal (skin conductance)* peaks per role detected by an increase of 0.03 μs was registered in each row of the MM per second.

For instance, the highlighted *row B* in Fig. 2 indicates that at the time 05:02 the action "IV antibiotic administered" was performed (see the number 1 in the column) and that 2 EDA peaks were detected in that time. That way, each data point is provided with meaning from the contextual data.

Columns can be conceptually grouped into dimensions of *group activity* (Fig. 2C). In our study, the actions performed by nurses are mostly associated to the *epistemic* dimension, and the physiological data with the *affective* dimension. The course coordinator was primarily interested in these dimensions but other aspects of collaboration (such as *social* and *physical*) have been investigated in other work [31]. Finally, segments can be grouped into stanzas to represent meaningful associations (Fig. 2D). In our study, the critical actions performed by the patient (such as experiencing pain or nausea) served to group the segments into *stanzas* in which students were expected to perform certain actions. Although the MM can be useful for mapping from multimodal data to more meaningful information, this structure cannot be directly transformed into data stories without further assessment of the data.

B. Automated Assessment of Data

Once the MM has been coded into meaningful discrete data, the next step is assessing such data according to the pedagogical intentions of the teacher. This corresponds to what feedback teachers would normally provide to students on comparing the intended with actual student behavior [2]. Automated feedback can be provided through rule–based algorithms derived from teacher's intentions or the assessment criteria. For this particular study, the learning intentions were extracted primarily based on the learning task design described in Sections III-A and III-B for simulations 1 and 2 respectively.

For instance, the nursing coordinator wanted to assess i) whether students performed the intended actions after the patient complained of specific symptoms, and ii) for students to reflect on how they experience arousal during the simulation. In this case, the coordinator set the requirement that *"the patient's vital signs should be checked at least every 10 minutes"*. A simple rule can be defined to assess the completion of this task by checking the timestamp difference of two logged *vital sign assessment* actions. If the *timestamp difference* is longer than 10 minutes, the system can detect that the team has made an error.

The next section describes how the DS principles presented in Section II-A were operationalized to map from outcomes of the rule–based assessment to data stories.

C. Crafting the Data Stories

The last modeling step is the visual representation of the data story. Based on the rules described above, two storytelling prototypes were created. Following principle **DS2** (selection of an appropriate chart type), both prototypes were based on a minimalistic visualization technique (**DS3**: strip out excess detail), namely, a *timeline of actions* by each nurse within a team during the simulation (captured by the manikin and the observer, presented as annotated blue circles on each nurse's line of actions—see Fig. 3). Actions performed by the patient (such as asking for help or complaining of severe pain) are represented by red vertical lines, which also divide the dataset into stanzas. Since line charts effectively show changes over time [11], this visualization without any visual enhancement was used as the background in both prototypes. For both prototypes, a user interface was generated to allow students to navigate through one data story at a time. Data stories were accessible via a set of buttons located at the bottom of the interface which add visual and narrative elements to the *timeline of actions* (see Fig. 5).

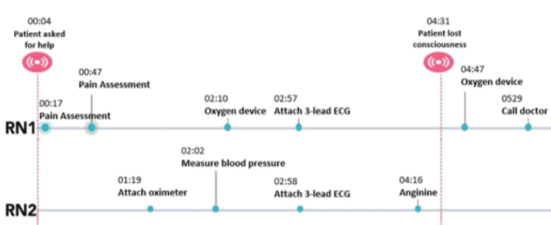

Fig. 3. Example timeline of actions for a team of two nurses, which served as the background of both prototypes.

Following principle **DS1**, each story focused on a very specific learning/reflection goal.

Prototype 1 presented six different stories, toggled by the six buttons in the dock. The first five stories presented the assessment of the critical actions that nurses were intended to perform in Sim 1 (i–v in Section III-A, an example focused on action iii shown Fig. 5).

Prototype 2 presented five different stories. The first four stories corresponded to the intended critical actions in sim 2 (i–iv in Section III-B, a story focused on action i shown in Fig. 6). The sixth story of each prototype showed all the

Learning Intention	Rule pseudocode
RULE 1 - used to generate a story about an **error made** shown in Figure 4, Example 1 Administer Fentanyl within 10 minutes after the patient complained of abdominal pain	IF time_between(Abdominal_pain, IV_medication) > 10: add_annotation("The patient needed pain relief.....") add_arrows(annotation, data_point) highlight_datapoints(data_point, "orange") add_rectangle(Abdominal_pain, IV_medication, "orange") add_title("The team administered Fentanyl X min late") ELSE add_annotation("You did it right! Well done!") highlight_datapoints(data_point_list, "blue")
RULE 2- used to generate a story on **arousal** shown in Figure 4, Example 4 How aroused was a nurse after the patient complained of experiencing chest pain	SPLIT (Max_arousal_ratio/4 quartiles) SWITCH (Arousal_ratio_in_a_stanza): > Quartile 4: add_annotation("Very High") in Quartile 3: add_annotation("High") in Quartile 2: add_annotation("Mild") in Quartile 1: "Low" do nothing in Quartile 1: "Very low" do nothing add_title("RN# presented several arousal peaks ...")

Fig. 4. Rules used to highlight visual elements in example 1 and 4 in Fig. 5 and 6.

arousal events detected per role (see Fig. 8 and Fig. 7). A sample interface exposing the rule–based algorithms driving each story was shown to the students along with the prototype (see Fig. 9).

Fig. 4 presents two exemplar rules used to create the data stories shown in Fig. 5 and Fig. 6 respectively. Endorsing principle *DS4*, selected elements of the visualization were emphasized by *adding enhancements* such as: a) icons (see Fig. 5–8, b) arrows , and c) enclosing areas; *changing color*, contrast, or thickness (also see elements marked as A, B, and C); and d) *annotating* salient data points, or e) *adding titles* that summarize the take–away message.

For example, in Fig. 5 an *if–then–else* rule (Rule 1 in Fig. 4) was applied to assess the presence and timing of a critical action in a certain stanza of the logged activity. The algorithm adds visual elements to the visualization, including: (a) orange icons to highlight the relevant data points; (c) a colored enclosure area to emphasize the stanza where the error was detected; (d) an annotation where the error occurred: *"The patient needed pain relief in less than 10 minutes"*; and (e) a prescriptive title : *"The team correctly administered fentanyl but 18 min late"*.

In Fig. 8 a more sophisticated *switch* algorithm (rule 2 in Fig. 4) assesses the ratio of arousal peaks for each nurse in each stanza and compares it to the highest ratio of arousal peaks experienced by a single student that we have detected in all of our nursing simulation studies (5 peaks/minute). This maximum ratio is divided into quintiles of equal size which are used to categorize the arousal experienced by a nurse (as very low, low, mild, high, or very high) in any given stanza. Only the last three categories are shown in the interface to provoke discussion on high levels of arousal.

V. Study and Analysis

The qualitative studies presented in this paper used a retrospective reflection technique [39] to investigate the opportunities and challenges of the DS prototypes, which were presented for the first time to the nursing students that are part of these studies.

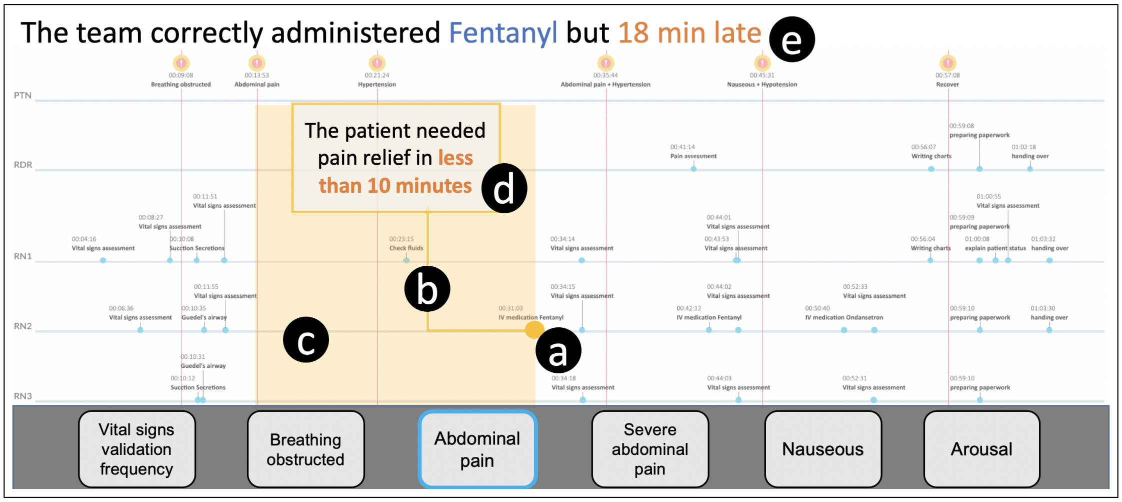

Fig. 5. Data story example 1. Prototype 1 showing a story on an error made in terms of time responsiveness.

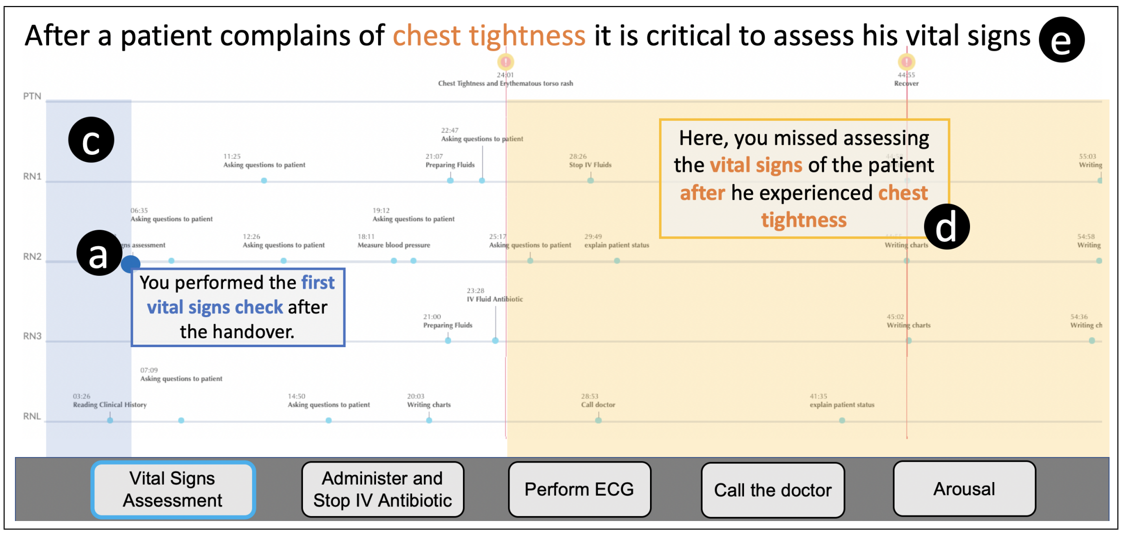

Fig. 6. Data story example 2. Prototype 2 showing a story on an error made in terms of actions omission.

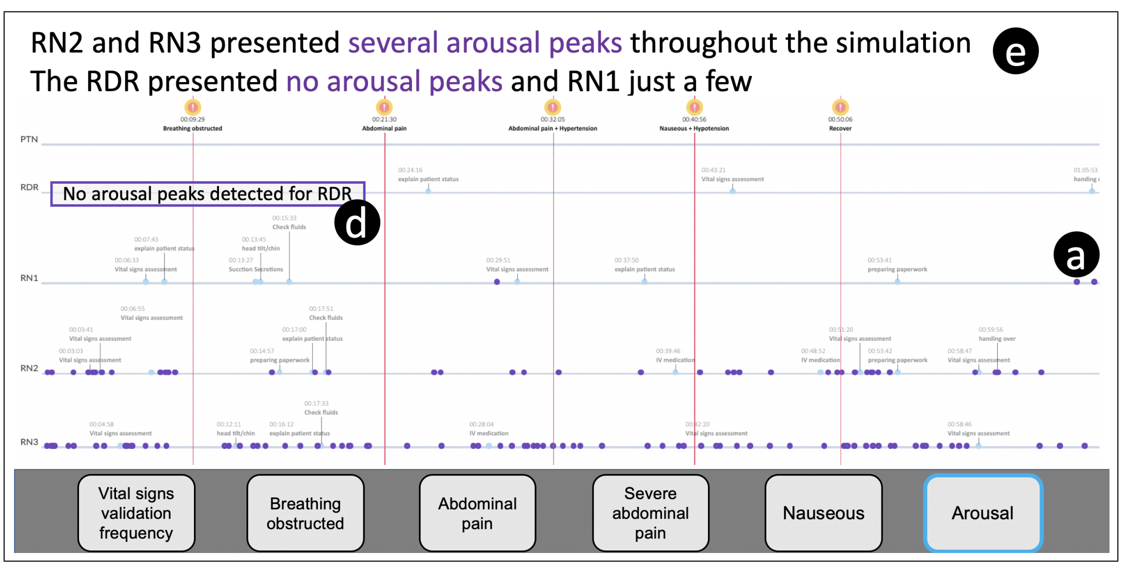

Fig. 7. Data story example 3. Prototype 1 showing a story on arousal with all relevant detected peaks

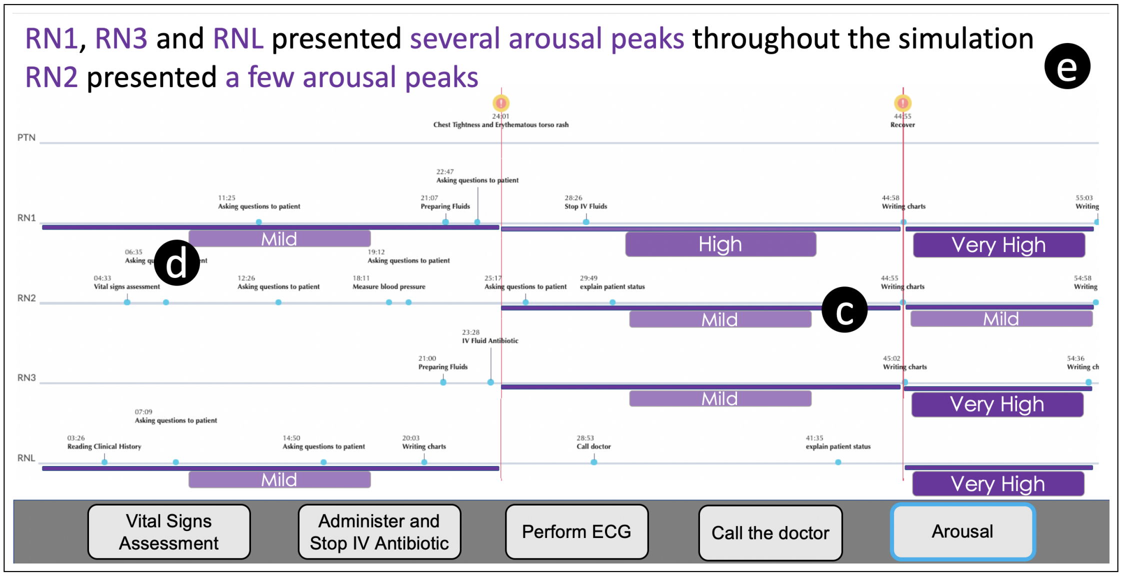

Fig. 8. Data story example 4. Prototype 2 showing a story on arousal showing annotations only.

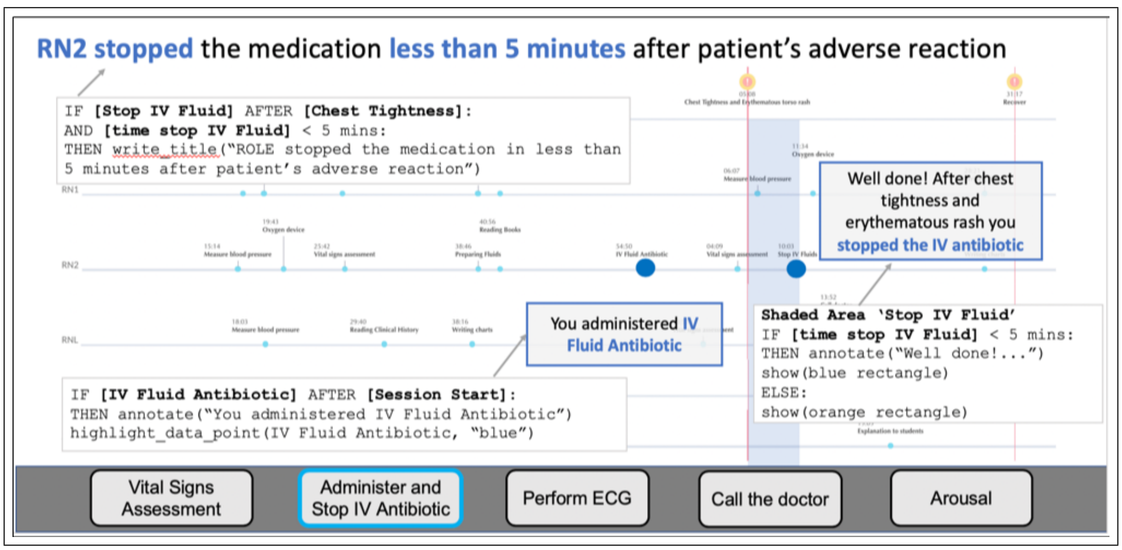

Fig. 9. Prototype 2 opening the algorithm used to enhance the timeline of events.

A. *Research Questions*

Both studies were conducted using LATEP (Learning Analytics Translucence Elicitation Process), an elicitation protocol for understanding how nondata experts envisage the use of LA systems [40]. Based on this, our studies investigated the following four research questions with a specific focus upon the student perspective:

1) Can students recognize any *added value* of interacting with multimodal data stories based on their own data? If so, how can this support their reflections?
2) What are the *anticipated potential uses* of DS tools?
3) What is the perceived potential impact on students' *accountability* in using data stories for their learning?

Inspired by the growing interest in explainable Artificial Intelligence (AI) [41] to support transparency and improve trust in systems, we also sought to understand: (4) What are the potential implications of *exposing the algorithms* used to create data stories to students using multimodal data?

B. *Participants*

We invited the 44 students who were recorded while enacting sim 1 and 2 to participate in optional team reflection sessions a week after each simulation. These were co–organized with their teachers as an optional extra activity at the end of the class time. A total of 16 out of the 19 female students in study 1 volunteered to participate in the post hoc reflections (aged 19–53 years, mean=27, std=10), and 23 out of the 25 students in study 2 (21 females and 2 males, aged 20–45 years, mean=23.5, std=5.4). Each of the 4 teams in study 1 (T1–T4) and study 2 (T5–T9) had from 2 to 5 students attending this reflection. All sessions were conducted in a meeting room next to the students' regular classroom.

C. Reflection Protocol

The reflections were conducted as a 30–minute focus group, which consisted of reflection sessions with students (S1–5) from each team (T1–9). This study design reflects the standard approach used in this class–based scenario, in which it is normal to conduct a clinical debrief after team simulations. The sessions were structured as follows:

1) *Pretest on procedural knowledge*. Before the group reflection, students were asked to individually list all the nursing actions that should have been performed at specific times during their simulation. A blank version of the timeline (Fig. 3), showing only the actions of the patient, was provided on paper for them to annotate *(aim 1)*.

2) *Think-aloud team reflection*. i) In teams students were asked to think aloud while jointly inspecting the timeline (without data stories) of their own team (e.g., Fig. 3). ii) Next, teams were asked to explore the *enhanced timelines* mocked up as interactive screens (e.g., Fig. 6), in any order they wished. Then, they were asked about their views on these versions *(aim 1)* as follows: a) Did the data stories add value to the timeline (without stories)? and b) how do you think your performance was based on the data stories?

3) *Usage and accountability*. Students were asked about usage and accountability opportunities and concerns *(aims 2, 3)*, as follows: i) how can the system be used in or outside the classroom? and ii) who should be able to see the interface, for which purpose and in what form?

4) *Explainability*. Students were asked to review the interface that exposes the algorithm (Fig. 9). Students were asked if they understood the rules used to craft the stories and whether they would like to see these rules in the interface *(aim 4)*.

5) *Posttest on procedural knowledge and student perceptions survey*. Finally, students were asked to go back to the paper version of the timeline annotated in step 1, and correct/add/remove any actions they wished *(aim 1)*, using red ink on the paper. A survey questionnaire was provided to the students to elicit their individual perceptions about the prototypes. The prototypes were presented to students using a 21–inch Liquid Crystal Display (LCD) display, connected to a laptop running macOS 10.14 with a wireless mouse for students to explore the stories.

D. Analysis

Tasks 1 and 5: Mean number of errors in the pre and posttests of procedural knowledge were calculated, per team and per study, as a measure to validate if students corrected some of the procedural misconceptions they had.

Tasks 3–4: Reflection sessions were audio-recorded, fully transcribed, and coded using NVivo. Two researchers were present in each session. We examined participants' statements and their actions exploring the prototypes. Following best practices of qualitative research ([42] p. 13), and given the direct alignment between the study protocol and the analysis themes, statements of interest were jointly coded [43] by two researchers according to the preset themes of the study protocol: a) added value of the data stories to support reflection; b) anticipated usage strategies; c) accountability and privacy; and d) explainability. Resulting coded statements were examined by authors who had several discussions to select instances that illustrate opportunities and concerns of the approach to create MMLA interfaces that communicate insights.

VI. Results

This section presents the results of the analysis organized around the four research questions presented above. The emerging topics are quantified based on the mentions of each topic per team (not per individual).

A. Added Value of the Data Stories

Compared to the test before exploring the data stories, in the posttest, all teams showed a decrease in the number of errors (defined as omitted actions and slow responses according to the learning design). Students made an average of 3.6 errors (Study 1, out of 5 actions) and 2.3 errors (Study 2, out of 6 actions), before inspecting the data stories, reduced to 2.6 and 1.8 errors respectively, after the reflection activity. This suggests that the reflection activity helped them to identify the errors they had not correctly identified. It is possible, however, that any activity that engaged them in reconstructing a week–old simulation might have helped refresh their understanding, so deeper insights are needed to claim that the data stories were adding specific value.

During the think-aloud protocol, students explained in detail their perspectives on how the data stories offered guidance for them to interpret the timeline.

Students in all nine teams appreciated the guidance offered by each data story to be able to **focus** on the expected learning goals. For example, a student explained: *"[the data stories] divide the screen into sections [critical incidents of patient]. These sections looked like the same thing in [the timeline of actions without enhancements]. It does not highlight the things that needed to be done and the timeframes"* (T2, S2).

Other students more explicitly explained the value of reflecting through the data stories, because they were able to clearly see the mistakes made during the sim for example: *"[data stories] highlight what you did not do [right]"* (T8, S4); and *"[the data stories] show both what we did and what we were meant to do"* (T8, S4).

Some DS elements and their usefulness were emphasized by students during their reflections. For example, two teams mentioned that **enclosing, colored regions** helped them to understand the expected timeframe in which they were meant to perform certain actions, and five teams agreed that the **annotations** informed them about how they could act upon the information: *"It is not just that we were late but we were late by that much time, so we can precisely see what we have to improve, and what we have done well. [The data stories] make a big difference"* (T1, S2).

Students also commented on the designs of the prototypes (V1, shown in Fig. 7 and V2, Fig. 8) and commented on how the **design decisions** impacted the interpretability. Three out of the five teams in Study 2 found that V2 was less complex than V1. Students explained their preference in terms of the simplicity of V2 to more quickly digest the arousal data, compared to looking at every arousal peak in V1. One student explained: *"We just want a quick snapshot to see how [our arousal] was. That is really interesting to see in [V2]"* (T6, S5). Another student compared both prototypes and explained: *"I tried for a long time to understand [V1]. I do not know how to understand this, but this other one [V2] is pretty straightforward"* (T9, S5).

Students in the other 2 teams suggested that both views could be combined. A student explained that *"because, the dots [in V1] can be more specific"* (T8, S2) students could first view V2 and then see the V1 details on demand. Another student suggested a third option, as follows: *"You can combine both [V1 and V2], using short and long lines to show arousal peaks"* (T9, S4).

Overall, students appreciated the benefits of augmenting the timeline of actions with annotations and visual elements that explained the errors they made, or the extent to which they experienced arousal during the reflection. Additionally, a survey was performed at the end of the interview to elicit students' perceptions of the added value of each layer, the results of which are described in Fig. 10. The survey results support the students' interview comments. In the next section, we report more in detail how students used the data stories to reflect on the simulation, and anticipated usage strategies.

B. Anticipated Usage Strategies

This subsection is divided into three parts, presenting anticipated usage opportunities of the data stories regarding i) errors made, ii) arousal, and iii) the storytelling tool as a whole.

1) Use of the tool for reflection on errors. Students in 7 of the 9 teams considered that the data stories on errors provided them with the **opportunity to improve** their clinical practice. For example, one student stated that the data stories *"showed [them] what [they] did well and what [they] should have done to improve"* (T3, S4). Some students mentioned that each data story helped the team to focus on improvements to be made. This was stated by one student as follows: *"[the data stories] help us to reflect on practical ways to improve and work as a team"* (T3, S5). They expressed that seeing their actual performance, with errors highlighted, made them think about **specific skills and knowledge** they need to master. One student said: *"[the data stories] highlighted areas where I need to strengthen my clinical thinking"* (T2, S4). Another student more specifically listed the kind of skills she had to strengthen as a result of her reflections: *"[the data stories] showed me that I need more knowledge about PAC [post anaesthetic care] nursing, I need more knowledge about gaining trust in postoperative, hypertension etc."* (T2, S2).

Students suggested some **strategies** for making use of the storytelling tool. For example, some of the students' reflections were quite individualized, suggesting potential uses of the tool to support individual reflection or reflection on particular roles (3 teams). For example, one student reflected as follows: *"I can use the timeline to see when and what action I did during the whole simulation, and it is very useful and helpful for me to reflect and analyze my work"* (T7, S2). In contrast, students in 7 teams suggested that the tool should be used to provoke group reflection, for instance: *"The timeline helps the group to recall what we did because sometimes we may forget about the details or the sequences of our own actions"* (T7, S3).

Reflections about **collaborative** skills and **teamwork** were shared by students while looking at both their own errors and those made by others in their team. For example, students mentioned that they could see the *"importance of teamwork in clinical scenarios"* (T3, S1), *"identify collaborative work as a team"* (T2, S4), that they needed to find *"ways to improve and work as a team"* (T3, S5) and work on improving their *"critical thinking"* (T2, S4).

2) Use of the tool for reflection on arousal. Students in 5 teams who exhibited very high arousal reflected on what they were doing at those points in the simulation **actions**. For example, one student explained: *"I was writing notes. Everyone else was talking and I was writing notes"* (T4, S2), or *"[my very high arousal] was more like during the airway"* (T4, S4). Another explained that not knowing which actions to perform during the simulation, made them feel under stress: *"I was nervous because I didn't know what to do in the situation. I tried my best to do it, but when I gave the handover [to the Doctor] and he did not accept it, I got more nervous. I had to come back and do it again"* (T3, S1).

Seven students associated their arousal peaks with their **roles**. For instance, a **team leaders** who displayed very high arousal peaks explained that this was due to their engagement with the team and the responsibility with the patient. For example, the team leader in T6 explained this as follows: *"I wanted to make sure I was doing the right thing for the patient. I had to constantly remind myself and the team members about the actions that had to be performed"*. Similarly, the team leader in T8 stated: *"I am very calm under pressure. I think I started mildly aroused because I thought: I am the team leader! I was chilled then because I knew what was going on and I was very confident with what was happening"* (S1). Another team leader reported: *"At first, I felt a bit of stress. Then, I did not feel panic or anything like that, but I was just more concentrated on the task than before"* (T9, S4).

Students who performed the **scribe role** had mixed perceptions about their high and low arousal. One scribe who exhibited a higher arousal mentioned that she was very engaged with the simulation: *"I think I was trying to work out how we should deal with the situation. I got excited about that"* (T6, S5). Another scribe who exhibited a lower arousal explained she was disengaged due to the irrelevant duties for her role: *"I am only the observer. I am not the nurse" (T9, S4)*.

Interestingly, students in 3 teams pointed out that **external factors** such as tiredness, fatigue, or sickness might have affected their arousal during simulations. One student explained this as follows: *"It was just later in the day, and you are tired, and want to go home"* (T3, S4). Another pair of students reflection on their arousal peaks reported that one felt very nervous as she had *"injured herself"* (T9, S3), while the other *"was sick"* (T3, S1).

In addition, students from 3 teams suggested that **wearing external devices and being recorded** could cause bias in arousal peaks. One student who exhibited higher arousal levels expressed how she felt during the recordings and data collec-

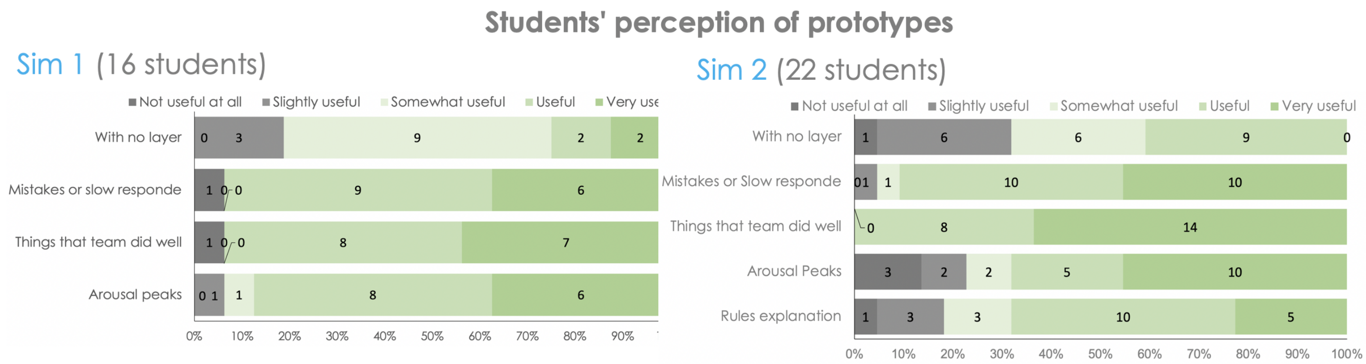

Fig. 10. Survey results of students' individual perceptions of the added value of the layered visualization: with no layer, timeline without enhancements (Fig. 3); mistakes or slow responses and things that team did well (Fig. 5–6); arousal peaks (Fig. 7– 8); and rule explanations presented to students of sim 2 (Fig. 9).

tion: *"I felt a bit stressed because [the group] was getting filmed"* (T7, S1). Similarly, one student in T8 mentioned that being recorded to some extent could cause spikes in arousal readings: *"As soon as anyone puts anything on me, like I am going to check your heart rate, I am like, no! So, my heart rate goes up in a way"* (S1). Nevertheless, the other 6 teams mentioned specifically that they felt comfortable using the devices, and felt that they behaved normally during the simulation.

3) Other uses of the tool in the classroom. In 6 teams, students described the value of the visualizations during the usual teacher–led **debriefs**. For example, a student envisaged how the tutor could help them to go through the data stories to reflect on further medical conditions and procedures as follows, which is not possible at present for logistical reasons: *"The tutor can actually explain things, if you have something to go off. Usually we do our sim and then we forget about it and never look at it again, or they [tutors] would not know what we have done anyway, because they are not watching each person"* (T7, S3).

Additionally, students in 5 teams mentioned the possibility of using the tool for the **assessment** of their performance and the actions they are intended to perform when enacting certain roles. However, most students (in 6 teams) preferred the tool to be used to enhance the provision of feedback (formative assessment) as opposed to perform automated summative assessments. Surprisingly, none of the students argued against the incompleteness of those summative assessments. However, they were mostly worried about the additional pressure this would put on them, as stated by one student as follows: *"I feel like an assessment would be so stressful"* (T6, S2).

To summarize, this section has illustrated students' broadly positive responses to the prototypes, and the different ways in which they envisaged their use. Students agreed that the tool would be very useful to enhance the debriefs with their teachers, which are currently performed without using any evidence, relying only on what teachers can see while dividing their attention between 5–6 teams.

C. Accountability and Privacy

Discussions about reflecting on the data stories and the mediating role of teachers was a matter of concern for some students. Students in six teams mentioned that debriefs guided by the data stories could provide a better team discussion, even without the teacher, and that they were willing to compare their performance with other teams (6 out of 9 teams) and *"see what each group member did"* (T3, S3). However, students in these teams agreed that this social comparison should be led by the tutor: *"Do not just give it to the student, get the tutor to sit down with the group and talk it through"* (T3, S3). Students were asked to explain their views about sharing their data with other students, tutors, or teams. Students in all teams agreed that it would be fine to **share** their timeline with other students, especially *"If it helps [other groups]"* (T3, S2) to reflect on their mistakes. One student (S4) added: *"If it is a great mistake, I want to share it with others"* (T9).

However, some concerns about **anonymity** were also raised in 2 teams who thought that the timeline should only expose activities performed by specific roles, without disclosing the names of the students. One of the students emphasized: *"you [researchers] do not have to put our [team members] names on it"* (T5, S1). Only one student suggested that their data should not be shown to other students. However, her rationale was that other students would find it boring to explore other team' timelines *"because it does not relate to what they have done. They cannot learn from it. I have to learn from what I have actually done"* (T1, S1).

Surprisingly, none of the students raised concerns about sharing their arousal traces (even when asked specifically about their perceptions about data sharing), which we attribute to them feeling comfortable in the classroom as an academic space. As mentioned by (S4): *"the lab is a safe environment for humans"* (T9) and they knew their data would be used only for learning purposes.

D. Explainability of the Rules

When students (only the 5 teams of Sim 2) were exposed to the rules used to craft the data stories, most students mentioned feeling confused, and hesitated to respond (e.g., *"You mean* *all those boxes? I am a bit confused"*, T9, S5). Students in three teams requested further explanations, for example: *"If you [interviewer] do not explain it, how long would it take to understand this stuff?"* (T9, S4). Other students judged that the rules needed some basic coding skills to be fully understood (e.g., *"If you do not know coding, it is hard to understand"*, T9, S4). Some students however, demonstrated that they could understand them, "translating" one rule: *"If time is less than five minutes, the box would be blue. Otherwise, it's going to be orange because it is going to be like a mistake"* (T6, S1).

When students were asked to comment on the **added value** of seeing the rules they had contrasting views. Two students from teams 6 and 8 asked to include the rules into the timeline. One of them suggested alternatives to present them, such as adding extra explanations in the existing stories on–demand: *"Can't you [researchers] just incorporate the rules into each [story] instead of doing it in an extra interface"* (T8, S1). However, another student (team 7) argued that by inspecting the rules, students could potentially solve emerging questions about their mistakes while exploring the timeline. The student argued that if she makes any mistake she will: *"ask a question about why? Why is that wrong? Like it is [data stories] feedback or whatever it is called, it should be constructive"* (T8, S1). The rest of the students (in teams 5 and 9) mentioned they would not find the rules useful and that they would not try to change the rules (*"I would like to keep [the parameters] like that"* T9, S1). Two students stated that teachers would be in a better position to *"understand the parameters"* (T5, S2) and that they would also need *"basic [concepts of] coding"* (T8, S1) to change them. Interestingly, in the survey (see Fig. 10) students rated the explanations in the rules layer as *mostly useful* (10 students) and *very useful* (5 students.) It is possible that they either misinterpreted what this question was asking about, or were more honest in the interviews than in the survey.

VII. Discussion and Future Work

In this section we summarize the key findings, share our critical reflections connecting to the broader literature, and note the limitations of the studies.

A. Storytelling for Learning Analytics

Until this study, the nursing simulation activity timelines we have been developing had been evaluated positively by academics/tutors, but not students. This paper's results show that the enhancements using DS principles helped students identify misconceptions, think about strategies to address errors they made, and reflect on the arousal levels they may have experienced during the simulations. Students from the nine nursing teams identified errors that they had missed prior to engaging with these enhanced designs, and reported, through both interviews and a survey, that overall, the DS enhancements assisted them in focusing their attention on the expected learning goals (Section VI-A). This evidence supports the proposal that DS principles can be designed to assist in communicating insights found in learning data, helping to address problems with visual LA products documented by other researchers [6], [9], [44].

Although the studies presented in this paper were conducted in the context of complex, multimodal learning situations, there is no reason why a storytelling approach could not be implemented to aid in the interpretation of more conventional LA visualizations supporting noncollocated teamwork. Guiding students while interpreting their data is a feature that is missing in most of the current LA dashboards (as per the most recent review in [9]). This paper illustrated the specific case of making evidence on *errors* and *arousal* available to computational analysis and scaffolding the interpretation of such data meaningfully.

Hence, our contribution should be seen as one instance of the extensive work that needs to be conducted on guiding students in interpreting their data, by aligning the visual representations with the learning goals and teacher's pedagogical intentions. The literature suggests that future work could involve extracting insights about other high-order features of learning besides errors and arousal, for instance, learning strategies (e.g., [9]), progress in achieving certain goals (e.g., [45]), or effectiveness of collaboration (e.g., [46]).

Moreover, while much has been said in the LA community about the need to align LA with the learning design, far less design work, or empirical work, has demonstrated the visual means to make this alignment explicit in the design of student–facing interfaces (see recent review in [47]). This paper contributes to addressing these gaps, expanding the work initiated by Echeverria *et al.* [27] who validated a DS approach, driven by the learning design, with educators under controlled conditions. This paper extends this work by implementing two DS solutions with real students, *in-the-wild*. The careful alignment of data stories created from the analysis of students' data, with the learning intentions of the activity, offers rich opportunities to support students and to identify potential changes needed in the learning designs. Future work should explore these opportunities in other educational contexts. Additionally, more experimental studies could be conducted to identify which specific elements of the LA interface drive students' visual attention (e.g., such as the study using eye–trackers conducted by Echeverria *et al.* [27]).

B. Limitations and Remaining Challenges

The evidence reported here should be considered in the context of the limitations of the studies. While the clinical simulations were authentic to how nurses are trained, the participants were from nine teams performing two particular exercises. It is possible that other students could have given different reactions, and that the design approach documented here may not transfer to other kinds of simulation exercises which require the tracking of student activity that is too complex to automate.

To realize the goal of fully automated, timely, intelligible feedback on collocated teamwork, a number of challenges remain, and motivate several avenues of future work.

Firstly, the data stories provide a prescriptive way to communicate insights from data. In doing so, the **risk of over**

interpretation is evident. Students did not explicitly mention potential issues related to bias in our prototypes. However, students could have questioned how their arousal levels were determined or whether the errors made were correctly detected. To mitigate this risk, in our studies each data story was carefully aligned to the teacher's learning intentions. Yet, further research should investigate the impact of differences in the way the data is modeled and how stories are presented. This involves, for example, investigating the impact of modeling arousal data using alternative heuristics to convert peaks into categories or presenting the insights using different combinations of text narrative and visual enhancements.

The most questioned feature in the designs was the exposure of the semiformal "pseudocode" *if–then* rules. The nursing students, who rarely have programming skills, found these difficult to interpret. In line with the growing interest in explainable AI solutions [41], this points to the need to find ways to communicate to nontechnical people how a particular insight was obtained from the data to encourage transparency and trust in LA systems. Based on the assumption that greater transparency encourages adoption, we argue that explainability can be considered at two levels: (1) explainability in the way feedback is communicated, for example, explainable visual analytics providing a clear narrative that explains what went well or wrong during a simulation experience, and why; and (2) explainability at the level of feedback generation, this is, how the machine picked a story. Current work in explainable AI tends to focus predominantly on the latter, as a way to provide transparency in automated decisions made by the machine [48], however, in education, this is an area that still needs work, since the interactions between students and automated processes can be operationalized into learning opportunities (e.g., for students and teachers to understand the reasons behind the feedback that has been generated). For example, Echeverria *et al.* [27] present an approach for explaining student data to drive teachers' interpretations of visual LA.

From a modelling perspective, we would like to emphasize that the expectations from educators are context and task–dependent, which means they will change depending on the simulation. Consequently, the design presented in Section IV might not be applicable to easily identify errors or to interpret arousal levels in new contexts.

The application of DS is in its infancy and more exploratory work is needed before conducting longitudinal studies. Besides this, the DS approach fit well in the learning context of this study, where simulations occur in a short period of time and in a complex, special learning space. We chose not to artificially change the learning design of the activity for the purpose of research. Instead, we valued in-the-wild exploration of DS (in an authentic classroom setting). This study is thus focused on this kind of authentic experience. This leaves open for future work, a possibility for longitudinally focused research to understand the sustained use of DS and narrative tools in LA dashboards and reports.

Finally, from an engineering perspective, we note two areas for improvement. Sensor data is susceptible to noise, and further work needs to investigate the implications of false positives or negatives in the automated assessment of data, and strategies to mitigate these. Secondly, while the evidence from these design prototypes builds our confidence that they effectively provoke students' reflections, and the DS techniques have been designed from the start to be automatable, the generation of data stories from multimodal data is not yet fully automated. The timeline of actions is automated and can be shown to students instantly after they complete the classroom task, but the integration of data captured from different modalities still requires some human intervention (e.g., running algorithms separately to model the data and semiautomatically crafting the stories). This is a challenge that forms an important part of the wider research agenda in MMLA research. [29].

VIII. Conclusion

Inspired by the way teachers provide feedback to students, and how humans communicate through stories, the prototypes described in this paper are aimed at telling *"one educational data story at a time"* (Section IV-C). As mentioned in the literature review, it is naïve to expect that students will be able to interpret data and make them actionable without further assistance. DS can be a promising approach to facilitate such assistance and to augment teacher–led reflection. This paper presented two qualitative studies conducted in authentic nursing simulation classrooms with the purpose of communicating insights to students through *data stories*. Given the limitations of current visual analytics, we anticipate that approaches such as DS will grow in importance to help students make the most of the new forms of feedback that are becoming possible.

Acknowledgment

Gloria Fernández-Nieto gratefully acknowledges the University of Technology Sydney for her PhD Scholarship support. Roberto Martinez-Maldonado's research is partly funded by Jacobs Foundation.

References

[1] J. Hattie and H. Timperley, "The power of feedback," *Rev. Educ. Res.*, vol. 77, no. 1, p. 81–112, Mar. 2007, doi: 10.3102/003465430298487.

[2] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Stud. Higher Educ.*, vol. 31, no. 2, p. 199–218, 2005, doi: 10.1080/03075070600572090.

[3] N. E. Winstone and D. Carless, *Designing Effective Feedback Processes in Higher Education : A Learning-focused Approach*, 1st ed. Oxford, U.K. and New York, NY, USA: Routledge (Peterson & Co.), Evanston, IL, USA, 2020.

[4] W. Chen, C.-K. Looi, and S. Tan, "What do students do in a F2F CSCL classroom? The optimization of multiple communications modes," *Comput. & Educ.*, vol. 55, no. 3, p. 1159–1170, Nov. 2010, doi: 10.1016/j.compedu.2010.05.013.

[5] K. Kitto, M. Lupton, K. Davis, and Z. Waters, "Designing for student-facing learning analytics," *Australasian J. Educ. Technol.*, vol. 33, no. 5, p. 152–168, 2017, doi:10.14742/ajet.3607.

[6] L. Corrin and P. De Barba, "Exploring students' interpretation of feedback delivered through learning analytics dashboards," in *Proc. 31st Annu. Conf. Australasian Society for Computers in Learning in Tertiary Education (ASCILITE'14)*, Dunedin, New Zealand, Nov. 23–36, 2014, p. 629–633. [Online]. Available: http://www.ascilite.org/conferences/dunedin2014/proceedings/

[7] R. Bodily and K. Verbert, "Trends and issues in student-facing learning analytics reporting systems research," in *Proc. 7th Learning Analytics and Knowledge Conf. (LAK'17)*. Vancouver, Canada: ACM Press, Mar. 13–17, 2017, p. 309–318, doi: 10.1145/3027385.3027403.
[8] W. Matcha, N. A. Uzir, D. Gašević, and A. Pardo, "A systematic review of empirical studies on Learning Analytics Dashboards: A Self-Regulated learning perspective," *IEEE Trans. on Learn. Tech.*, no. May, pp. 226–245, jun 2019, doi: 10.1109/TLT.2019.2916802.
[9] W. Matcha, D. Gašević, N. A. Uzir, J. Jovanović, and A. Pardo, "Analytics of learning strategies: Associations with academic performance and feedback," in *Proc. 9th Learning Analytics & Knowledge Conf. (LAK'19)*. Tempe, Arizona, USA: ACM, Mar. 4-8, 2019, pp. 461–470, doi: 10.1145/3303772.3303787.
[10] D. Gašević, S. Dawson, and G. Siemens, "Let's not forget: Learning analytics are about learning," *TechTrends*, vol. 59, no. 1, pp. 64–71, jan 2015, doi:10.1007/s11528-014-0822-x.
[11] L. Ryan, *The Visual Imperative: Creating a Visual Culture of Data Discovery*. Cambridge, MA, USA: Morgan Kaufmann, 2016.
[12] V. Echeverria, R. Martinez-Maldonado, R. Granda, K. Chiluiza, C. Conati, and S. Buckingham Shum, "Driving data storytelling from learning design," in *Proc. 8th Learning Analytics & Knowledge Conf. (LAK'18)*. New York, NY, USA: Association for Computing Machinery, Mar. 7-9, 2018, pp. 131–140, doi: 10.1145/3170358.3170380.
[13] M. Schield, "Information literacy, statistical literacy and data literacy," *IASSIST Quarterly*, vol. 28, no. 2, pp. 7–14, 2004. [Online]. Available: https://iassistquarterly.com/public/pdfs/iqvol282_3shields.pdf
[14] A. V. Maltese, J. A. Harsh, and D. Svetina, "Data visualization literacy: Investigating data interpretation along the novice–expert continuum," *J. College Sci. Teaching*, vol. 45, no. 1, pp. 84–90, 2015, doi: 10.2307/43631889. [Online]. Available: https://www.jstor.org/stable/pdf/43631889.pdf
[15] R. N. Cuff, "On casual users," *Int. J. Man-Mach. Stud.*, vol. 12, no. 2, pp. 163–187, feb 1980, doi: 10.1016/S0020-7373(80)80016-2.
[16] J. P.-L. Tan and E. Koh, "Situating learning analytics pedagogically: Towards an ecological lens," *Learn.: Res. & Pract.*, vol. 3, no. 1, pp. 1–11, 2017, doi: 10.1080/23735082.2017.1305661.
[17] H.-J. Schulz, M. Streit, T. May, and C. Tominski, "Towards a characterization of guidance in visualization," in *IEEE Inf. Visualization Conf.)*, Baltimore, MD, USA, Oct. 13-18, 2013, p. 1.
[18] D. Ceneda *et al.*, "Characterizing guidance in Visual Analytics," *IEEE Trans. Visualization & Comput. Graph.*, vol. 23, no. 1, pp. 111–120, 2017, doi: 10.1109/TVCG.2016.2598468.
[19] B. Dykes, "Data storytelling: What it is and how it can be used to effectively communicate analysis results," *Appl. Marketing Analytics*, vol. 1, no. 4, pp. 299–313, 2015.
[20] C. N. Knaflic, *Storytelling With Data: A Data Visualization Guide for Business Professionals*, 12th ed. New Jersey: John Wiley Sons, 2017.
[21] I. Jivet, M. Scheffel, H. Drachsler, and M. Specht, "Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice," in *Proc. 12th Eur. Conf. Technology Enhanced Learning (EC-TEL'17)*. Tallinn, Estonia: Springer International Publishing, Sep. 12–15, 2017, p. 82–96, doi: 10.1007/978-3-319-66610-5_7.
[22] R. Bodily and K. Verbert, "Review of research on student-facing learning analytics dashboards and educational recommender systems," *IEEE Trans. Learn. Tech.*, vol. 10, p. 405–418, 2017, doi:10.1109/TLT.2017.2740172.
[23] T. Broos, L. Peeters, K. Verbert, C. Van Soom, G. Langie, and T. De Laet, "Dashboard for actionable feedback on learning skills: Scalability and usefulness," in *Proc. 4th Int. Conf. Learning and Collaboration Technologies (LCT'17)*, vol. 10296 LNCS. Vancouver, Canada: Springer Verlag, Jul. 9-14, 2017, p. 229–241, doi:10.1007/978-3-319-58515-4_18.
[24] T. Broos, K. Verbert, G. Langie, C. Van Soom, and T. De Laet, "Small data as a conversation starter for learning analytics," *J. Res. Innov. Teaching & Learn.*, vol. 10, no. 2, p. 94–106, 3 July, 2017, doi: 10.1108/JRIT-05-2017-0010.
[25] T. D. Ullmann, A. De Liddo, and M. Bachler, "A visualisation dashboard for contested collective intelligence Learning Analytics to improve sensemaking of group discussion," *RIED. Revista Iberoamericana de Educación a Distancia*, vol. 22, no. 1, p. 41, jan 2019, doi:10.5944/ried.22.1.22294.
[26] Q. Chen, Z. Li, T.-C. Pong, and H. Qu, "Designing narrative slideshows for learning analytics," in *IEEE Pacific Visualization Symposium*. Bangkok, Thailand: IEEE, Apr. 23-26, 2019, p. 237–246, doi:10.1109/PacificVis.2019.00036.
[27] V. Echeverria, R. Martinez-Maldonaldo, S. Buckingham Shum, K. Chiluiza, R. Granda, and C. Conati, "Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling," *J. Learning Analytics*, vol. 5, no. 3, p. 72–97, dec 2018, doi: 10.18608/jla.2018.53.6.
[28] R. Martinez-Maldonado, V. Echeverria, G. Fernández-Nieto, and S. Shum Buckingham, "From data to insights: A layered storytelling approach for multimodal learning analytics," in *Proc. 38th Annu. ACM SIGCHI Conf. Human Factors in Computing Systems (CHI'20)*, Honolulu, HI, USA, Apr. 25-30, 2020, p. 15, doi: 10.1145/3313831.3376148.
[29] P. Blikstein and M. Worsley, "Multimodal learning analytics and assessment of open–ended artifacts," in *Learning Analytics in Education*, d. neimi, r. d. pea, b. saxberg, and r. e. clark, eds. ed. Charlotte, NC, USA: Information Age Publishing, 2018, ch. 4, p. 89–112.
[30] R. A. Sottilare, C. Shawn Burke, E. Salas, A. M. Sinatra, J. H. Johnston, and S. B. Gilbert, "Designing adaptive instruction for teams: A meta-analysis," *Int. J. Artificial Intell. Educ.*, vol. 28, no. 2, p. 225–264, Jun. 08, 2017, doi: 10.1007/s40593-017-0146-z.
[31] V. Echeverria, R. Martinez-Maldonado, and S. Buckingham Shum, "Towards collaboration translucence: Giving meaning to multimodal group data," in *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (CHI'19)*, no. 39, Glasgow Scotland, U.K., May 4-9, 2019, p. 1–16, doi: 10.1145/3290605.3300269.
[32] X. Ochoa, K. Chiluiza, R. Granda, G. Falcones, J. Castells, and B. Guamán, "Multimodal transcript of face-to-face group-work activity around interactive tabletops," in *Proc. 8th Learning Analytics & Knowledge Conf. (LAK' 18)*, Sydney, Australia, Mar. 7-9, 2018, p. 1–6. [Online]. Available: http://ceur-ws.org/Vol-2163/paper4.pdf
[33] M. Vujović and D. Hernández Leo, "Shall we learn together in loud spaces? Towards understanding the effects of sound in collaborative learning environments," in *Proc. 13th Int. Conf. on Computer Supported Collaborative Learning (CSCL' 19)*, Lyon, France, Jun. 17-21, 2019, p. 2. [Online]. Available: //hdl.handle.net/10230/37185
[34] L. Berragan, "Simulation: An effective pedagogical approach for nursing?" *Nurse Educ. Today*, vol. 31, no. 7, p. 660–663, October 2011, doi:10.1016/j.nedt.2011.01.019.
[35] E. Palominos, T. Levett-Jones, T. Power, and R. Martinez-Maldonado, "Healthcare students' perceptions and experiences of making errors in simulation: An integrative review," *Nurse Educ. Today*, vol. 77, p. 32–39, jun 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S026069171830577X
[36] B. Mariani and J. Doolen, "Nursing simulation research: What are the perceived gaps?" *Clinical Simul. Nursing*, vol. 12, no. 1, p. 30–36, Jan. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876139915000985
[37] "Eda explorer," https://github.com/MITMediaLabAffectiveComputing/eda-explorer, accessed: 2020-01-21.
[38] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, "Automatic identification of artifacts in electrodermal activity data," in *37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Milan, Italy, Aug. 25-29, 2015, p. 10. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26736662/
[39] M. Hassenzahl and D. Ullrich, "To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals," *Interact. with Comput.*, vol. 19, no. 4, p. 429–437, Jul. 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095354380700029X
[40] R. Martinez-Maldonado, V. Echeverria, D. Elliott, C. Axisa, T. Power, and S. Buckingham Shum, "Designing translucent learning analytics with teachers: An elicitation process." *Interactive Learn. Environ.*, in press, doi: 10.1080/10494820.2019.1710541.
[41] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim, "Designing theory-driven user-centric explainable AI," in *Conf. Human Factors in Computing Systems (CHI'19)*. Glasgow Scotland, U.K.: ACM Press, May 4-9, 2019, p. 1–15, doi: 10.1145/3290605.3300831.
[42] N. Mcdonald, S. Schoenebeck, and N. McDonald, "Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice," in *Proc. ACM Hum.-Comput. Interact.*, vol. 3, Austin, Texas, USA, Nov., 2019, doi:10.1145/3359174.
[43] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Res. Physiol.*, vol. 3, no. 2, p. 77–101, Jan. 2006. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa
[44] B. A. Schwendimann *et al.*, "Perceiving learning at a glance: A systematic literature review of learning dashboard research," *IEEE Trans Learn. Technol.*, vol. 10, no. 1, p. 30–41, Oct. 2016, doi: 10.1109/TLT.2016.2599522.
[45] D. Ifenthaler and J. Y.-K. Yau, "Utilising learning analytics to support study success in higher education: A systematic review," *Educ. Technol.*

Res. Develop., vol. 68, no. 4, pp. 1961–1990, aug 2020. [Online]. Available: http://link.springer.com/10.1007/s11423-020-09788-z

[46] E. Er, Y. Dimitriadis, and D. Gašević, "Collaborative peer feedback and learning analytics: Theory-oriented design for supporting class-wide interventions," *Assessment & Eval. Higher Educ.*, vol. 46, no. 2, pp. 169–190, feb 2021, doi: 10.1080/02602938.2020.1764490.

[47] K. Mangaroska and M. N. Giannakos, "Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning," *IEEE Trans. Learn. Technol.*, vol. 12, no. 4, pp. 516–534, Sept. 2018, doi:10.1109/TLT.2018.2868673.

[48] H. Hagras, "Toward human-understandable, explainable AI," *Comput. J.*, vol. 51, no. 9, pp. 28–36, sep 2018, doi: 10.1109/MC.2018.3620965.

Gloria Fernandez-Nieto received the B.S. degree in telematics engineering from Universidad Distrital Francisco Jose de Caldas, Bogotá, Colombia, in 2011 and the M.S. degree in computer science from Universidad de Los Andes, Bogotá, Colombia, in 2015. She is currently a Ph.D. candidate at the Connected Intelligence Centre, University of Technology Sydney, Sydney, Australia. Her current research focuses on exploring alternatives of feedback to understand traces from data collected in the computer-supported collaborative learning classroom, with the goal of prompting reflection in teaching and learning practices.

Vanessa Echeverria holds a Ph.D. in Learning Analytics from the University of Technology Sydney, Australia. Her research interests are on designing effective feedback for collocated teams using multimodal learning analytics evidence Her research focuses on the exploration of intelligent technologies to support the development of students' skills in collaborative spaces in digital and physical environments

Simon Buckingham Shum received his B.Sc. degree in psychology from the University of York, York, U.K., in 1987; M.Sc. degree in ergonomics from University College London, London, U.K., in 1988; and a Ph.D. in psychology from the University of York, in 1992. He is currently professor of learning analytics at the University of Technology Sydney, Sydney, Australia, where he serves as director of the Connected Intelligence Centre. His research focuses on making thinking visible through the use of computer-supported analytics and visualization.

Katerina Mangaroska has recently graduated from the Department of Computer Science at the Norwegian University of Science and Technology. Her primary research area is learning analytics and multimodal learning analytics. She is also working with sensing technology to gain insights about learners' cognitive-affective states and their behavior in programming environments. Her other research interests center around learning design, intelligent tutoring systems, and human-computer interaction. Mangaroska is currently an associate professor at the University of South-Eastern Norway and a Fulbright scholar.

Kirsty Kitto is Associate Professor at the University of Technology Sydney. She models the many ways in which humans interact with information, and how this can change as a result of the different contexts in which they find themselves. Her work covers both theoretical models of context, and more applied domains in Learning Analytics and Educational Technology.

Evelyn Palominos has a background in nursing and received the Master degree in Learning Sciences and Technology from The University of Sydney, Sydney, Australia, in 2014. She is currently a Ph.D. candidate at the Faculty of Health, University of Technology Sydney, Sydney Australia. Her current research areas are learning technologies, educational psychology, learning from errors, and productive-failure designs in healthcare education.

Carmen Axisa received the master of nursing degree majoring in cardiothoracic from the Australian Catholic University, Sydney, Australia, in 2007. She received the Ph.D. degree from Sydney Medical School, The University of Sydney, Sydney, Australia, in 2021, where the focus of her doctoral work was physician mental health and wellbeing. She is currently a lecturer with the Faculty of Health, University of Technology Sydney, Sydney, Australia, where she teaches undergraduate and postgraduate nursing students.

Roberto Martinez-Maldonado is a Senior Lecturer at Monash University, in Melbourne, Australia. He has a background in Computing Engineering. His areas of research include Human-Computer Interaction, Learning Analytics, Artificial Intelligence in Education, and Collaborative Learning (CSCL). In the past years, his research has focused on applying artificial intelligence and visualization techniques to help understand how people learn and collaborate in physical spaces.