From Data to Insights: A Layered Storytelling Approach for Multimodal Learning Analytics

Roberto Martinez-Maldonado[1], Vanessa Echeverria[2,3], Gloria Fernandez Nieto[3] and Simon Buckingham Shum[3]

[1]Faculty of Information Technology, Monash University, Clayton, VIC, 3800 Australia.
[2]Escuela Superior Politécnica del Litoral, ESPOL, Gustavo Galindo Km 30.5 Vía Perimetral, Guayaquil, Ecuador.
[3]Connected Intelligence Centre, University of Technology Sydney, Ultimo, NSW, 2007, Australia.
[1]FirstName.LastName@monash.edu; [3]FirstName.LastName@uts.edu.au

ABSTRACT

Significant progress to integrate and analyse multimodal data has been carried out in the last years. Yet, little research has tackled the challenge of visualising and supporting the sensemaking of multimodal data to inform teaching and learning. It is naïve to expect that simply by rendering multiple data streams visually, a teacher or learner will be able to make sense of them. This paper introduces an approach to unravel the complexity of multimodal data by organising it into meaningful layers that *explain* critical insights to teachers and students. The approach is illustrated through the design of two data storytelling prototypes in the context of nursing simulation. Two authentic studies with educators and students identified the potential of the approach to create learning analytics interfaces that communicate insights on team performance, as well as concerns in terms of accountability and automated insights discovery.

Author Keywords

CSCW; teamwork; visualization; data storytelling;

CSS Concepts

• **Human-centered computing~Information visualization;** Applied computing~Collaborative learning

INTRODUCTION

Analytics and artificial intelligence (AI) are changing the nature of work in multiple sectors, and although predictions vary, this trend will continue for the foreseeable future [2]. This will make human-AI interaction commonplace [41], and a particularly important form that this takes is the need for rapid, evidence-based decision-making. This places a premium on the design of user interfaces that can facilitate the effective use of data by people who are experts in their domain but novices in working with data [37]. Visualisation researchers have made the distinction between *exploratory* and *explanatory* data visualisation on the basis of the target audience and their expected data and visualisation expertise [31, 40, 48, 92, 100]. *Exploratory* refers to those visualisations aimed at people with (at least some) data analysis expertise, in search of insights from unfamiliar datasets [99], while *explanatory* refers to the challenging task of providing insights to *frequent* users who bring little or no data analysis expertise [84] or *casual* users who use a system occasionally [19]. This distinction is critical in many sectors in which data and evidence are becoming an essential part of decision-making [95].

In the educational sector, educators and students are already being challenged by a growing number of prototypes and commercial products in the form of 'learning dashboards' [5, 6, 81]. Yet, as the number of learning analytics (LA) interfaces rise, their limitations are coming under critical scrutiny. Recent reviews identify the difficulties students have in interpreting and acting on data to improve learning [5, 44, 60], and the same applies to teachers [55].

A variety of strategies to tackle this problem has emerged. We can provide students with a level of personalisation, enabling them to configure widgets presented in dashboards [54]; or with charts enhanced with textual prompts [21]. An explainable AI approach [96] would involve helping people understand what the system knows about them [6]. However, these approaches still seem to frame the problem as one of enabling *non-data experts* gain insights from *exploratory* visualisations, and do not make the transition to *explanatory* visualisations that help to close the interpretive gap with intended users.

An example of such an approach is to improve the design of LA tools to more effectively *communicate insights* [52]. This paper focuses on a particularly challenging case: designing a feedback tool for collocated teamwork based on

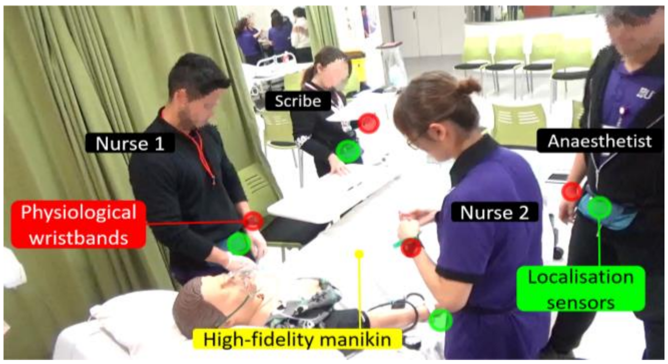

Figure 1: **Multimodal learning analytics in a nursing classroom.**

multiple sources of data captured via a combination of sensor signals (e.g. positioning and physiological markers), system logs and human logs. Making sense of data captured from multiple channels during group situations has been one of the goals of multimodal learning analytics (MMLA) [3] and teamwork science [46], but although there exist some MMLA interfaces tailored to *researchers or experts* [23, 66], there is to our knowledge no work investigating how to provide teamwork insights to *students and teachers*.

The contribution of this paper is an approach to unravel the complexity of multimodal data by visualising it into explanatory layers. The approach is illustrated through the design and validation of two prototypes, which incorporate *data storytelling* principles, in the context of nursing team simulation (Figure 1). We present two qualitative studies investigating uses of the *multimodal layered approach* to support reflection. The first study asked eight educators to make sense of students' data, and the second was an *in-the-wild* study conducted with eight teams of students reflecting on their own activity. Our findings show the potential of the prototypes to support reflection on team performance, stress management, and errors made. Our analysis also considers concerns around algorithmic and human accountability.

RELATED WORK

Two areas are relevant to this paper: the fundamentals of explanatory visualisation; and current work in MMLA.

Explanatory Visualisation and Data Storytelling

Explanatory visualisation is related to the emerging concept of *data storytelling* (DS) that builds on classic InfoVis [34, 94, 97] and narrative storytelling foundations [50]. DS is an information compression technique for communicating insights to an audience through the combination of data, visuals, and narrative [29, 76]. Data stories are commonly crafted manually by an analyst (with exceptions discussed in next section) who identifies insights from the data and creates charts with enhancements (e.g. changes in size, colour or saturation, and adding text and markers) highlighting only the data points that are most relevant to make a claim in a consumable form [48].

Explanatory visualisation is also related to *prescriptive guidance* [13, 80]. Providing guidance in visualisation involves developing approaches to facilitate sensemaking, with minimal knowledge required from casual users. Prescriptive guidance has been mostly tailored to support data analysts to find insights from multidimensional datasets [1, 89]. This shows that even expert data analysts require guidance to explore heterogeneous data.

Most empirical work concerning explanatory visualisation has focused on helping presenters *to manually craft* data stories [48], particularly in areas such as journalism [74, 82] and sports [88]. Indeed, a recent review [93] identified that most research on DS has focused on creating authoring tools to help designers to graphically craft their stories or annotate charts. However, a growing interest in generating annotations programmatically to facilitate the interpretation of charts by both analysts [9, 45, 87] and non-data experts [39, 62, 75] is being observed. Examples of the former include Bryan et al.'s [9] approach to automatically annotate data points in line charts and flow charts based on generic attributes such as value extremes, stable regions or sudden changes. Kandogan [45] presented a similar approach to annotate scatterplots. Srinivasan et al.'s [87] system extracts insights from a dataset to then, based on templates, suggest to the user ways in which those can be visualised via specific chart types and embellishments (e.g. opacity changes or added correlation lines). There has also been interest in targeting non-data experts. Ruchikachorn's tool [75] inserts short descriptions below scatterplots and parallel coordinates charts generated by analysing journalistic sources. Similarly, Hullman et al.'s [39] and Metoyer et al.'s [62] systems extract key text insights from data sources (e.g. news entries) and attach them to salient data points of visualisations to enrich their explanatory power. Notably, in the above examples, the annotations and highlighting are automatically generated based on either rules [9, 45, 75, 87] or text sources [39, 62].

This body of work highlights the relevance of explanatory visualisation beyond education, but it does not shed light on the key educational problem of how to communicate *pedagogical* constructs using multimodal data. Only a small amount of work has investigated the potential of DS in learning contexts. This includes Echeverria et al. [31] who demonstrated, using eye tracking, how explanatory visualisations can drive the focus of attention of teachers and lead to deeper reflections on students' performance. The same authors proposed that the teacher's instructional design should drive the visualisation design [30]. Chen et al. [16] proposed a guided tour of visualisations annotated automatically to facilitate the exploration of MOOC data.

Our work builds on work reported in this section, mainly on recent educational applications [16, 30, 31]. However, none of these works has been tested in authentic settings and only one [30] has considered the particularities of the learning design for crafting data stories. Our work goes beyond this work by 1) proposing an approach for dynamically generating explanatory MMLA tools that communicate insights based on the assessment criteria; 2) investigating the implications of educators and students using these interfaces; and 3) illustrating the approach in two authentic deployments in the context of nursing team simulations.

Multimodal Learning Analytics User Interfaces

MMLA is emerging as a promising, and increasingly affordable, way to capture and analyse human activity that until now, has remained ephemeral and invisible to computational analysis. Collocated teamwork in the classroom and the workplace is one such example [59]. Evidence about different modalities of students' interaction (e.g. posture, positioning and speech), and features invisible to the naked eye (e.g. electrodermal activity and pulse) can be captured via sensors, interactive devices or observations. With these data, inroads have been made into understanding how collocated group behaviours are connected to performance and learning outcomes [77, 78]; and finding patterns that can be used to personalise instruction [4, 57, 86]. Yet, the integration of data streams from multiple modalities, and people, can result in rather complex interfaces, hard to interpret [13]. This explains the dearth of MMLA user interfaces tailored to teachers and students.

Most current interfaces that provide automated feedback on teamwork are limited to mirroring simple information, such as amount of speech [53], or work well only in controlled conditions [15]. More work still needs to be done to tackle the complexity of showing insights from multiple data streams to non-data experts. Only two recent works have addressed this problem. Echeverria et al. [32] presented four visualisations, each presenting information related to one modality, namely speech, arousal, positioning and logged actions. The challenge is how to fuse the multimodal data into a single interface to facilitate reflection. This was attempted by Ochoa et al. [67] who visualised logs of students activity around a tabletop. Data shown included logged actions, verbal participation, gaze direction and emotional traits. However, more work needs to be done to assess the complexity of this interface and investigate how multimodal data can be visualised for explanatory purposes.

In sum, the few MMLA interfaces complex and exploratory, and are mostly targeted at data analysts (see review in [23]). To our knowledge, this is the first attempt in providing explanatory guidance to teachers and students to gain insights on team activity from multimodal data.

THE LEARNING CONTEXT

Nursing simulation is a constructivist learning model that provides student and registered nurses with the opportunity to experience communication, teamwork and patient situations while minimising risk of injury [26]. Simulations often start with a description of learning goals, followed by the simulation itself, concluding with a teacher-led *debrief* aimed at provoking students' reflection on performance and errors made [68]. Video-based products to support this reflection already exist. However, it is demanding for teachers to record and review videos from multiple teams in authentic classroom situations [35]. Thus, teachers and students rarely use evidence to reflect upon, which has been identified as a shortcoming that needs to be addressed [56].

Learning situations

This paper focuses on two authentic simulation situations:

Sim 1-Resuscitation. Nine students enrolled in the Bachelor of Nursing at University of Technology Sydney, aged from 20 to 53 years (avg=34, std=10), volunteered to participate in a simulation (sim) organised by a teacher. They were organised into three teams, of four (2 females, 2 males), three (2 females, 1 male) and two (female) students. The goal of this sim was providing care to a patient requiring cardiopulmonary resuscitation (CPR). A patient manikin (see Figure 1) was programmed to deteriorate over time, dividing the task into two phases. Phase 1 involved the patient's assessment, including four sub-tasks: i) give oxygen therapy; ii) assess chest pain; iii) give medication; and iv) connect an electrocardiography (ECG) device. In Phase 2 students had to perform CPR after the patient stopped breathing. Each student enacted one of four roles as registered nurses (RN1-4). RN1 was the team leader, with RNs 2-4 responsible for subtasks ii, iii and iv, respectively. Each sim lasted an average of 9.5 ±0.7 minutes. Phase 1 lasted 5 ±0.8 and Phase 2 4.5 ± 0.4 minutes.

Sim 2-Surgery Recovery. This sim was run in eight classes of the undergraduate subject Integrated Nursing Practice at the same university. Around 25 students attend each class, who are organised in teams of 4-6, each performing the sim around a patient bed. One team in each class (44 students in total, 40 females, 4 males, aged from 19 to 53 years, avg=25 ±7.8) volunteered to participate. Students in each team played the roles of anaesthetist doctor, recovery nurses (RN1, RN2), scribe (RN3) (see Figure 1), observer and patient. The student playing the role of the patient enacted what a real patient would do while recovering from an abdominal surgery. According to the *assessment criteria* set by the teacher, an effective team should perform the next actions: i) assess vital signs every 10 minutes; ii) check fluids and perform oxygen therapy after the patient suffered breathing obstruction; iii) administer Fentanyl within 10 minutes after the patient complained of abdominal pain; iv) administer a second bolus of Fentanyl after the patient complained of severe pain; and v) administer Ondansetron within 10 minutes after the patient suffered nausea.

Apparatus

All sessions were conducted in classrooms equipped with 5-6 patient manikins. Students' positioning was captured through wearable (Pozyx.io) tags at 2-3Hz. Students wore physiological wristbands (Empatica e4). These record electrodermal activity (EDA) at 4Hz and wrist acceleration at 32 Hz. Some student actions were automatically logged by the high-fidelity manikins (Laerdal 3G), including placing the oxygen mask, attaching blood pressure monitor, reading blood pressure, administering medicine, attaching the ECG device, and performing CPR. Other actions were logged by an observer (a researcher, but it could also be a student) using a console to associate actions, pre-configured by the teacher, with students who performed them (e.g. starting intravenous fluid, checking vital signs, calling the doctor). All sessions were video recorded and data streams synchronised and down sampled to 1 Hz.

A LAYERED APPROACH FOR MULTIMODAL DATA

Illustrative scenario

The approach can be illustrated through a scenario. When a patient is unconscious, care of the airway is critical [12]. In an in-hospital CPR situation (like in Sim 1), this means a nurse should be ideally positioned *behind the bed* to perform ventilation techniques. If a student does not do this after the patient stops breathing, the teacher normally provides corrective feedback in the debrief. This behaviour can be identified based on positioning data and logs from the manikin, but a question remains: *how to map from sensor data to meaningful feedback as a teacher would do?*

The approach

The aim of the approach is two-fold i) categorising the underlying multimodal data into meaningful layers of information, and ii) applying data storytelling to drive visual attention to key events of the learning activity. As a starting point we use Echeverria's et al. [32] model for adding meaning to multimodal data. This is shown in grey in Figure 2. Boxes represent artefacts such as data structures and interfaces. Circles represent *input* parameters expressed as rules or templates to generate data structures or configure visualisations. Multimodal low-level learning activity data (D) are encoded into a meaningful information structure (M) based on domain knowledge (K). This structure can be directly rendered visible as *exploratory* visualisations (V). By operationalising this model, low-level sensor data can be imbued with contextual meaning to bring key stakeholders into the sensemaking loop. This makes this model well-suited to be expanded to map from low-level data to educational *insights*.

The added components (shown in blue in Figure 2) include parameters from the assessment criteria (A), used to generate a higher-order information structure, the Learner Model (LM). This is a structured representation of students' performance, misconceptions or difficulties [11]. The LM can be visualised as data stories (S) shown in a layered *explanatory* interface, by operationalising DS principles. In sum, the proposed approach expands Echeverria et al.'s by 1) modelling assessment criteria to enrich the data stream and produce a learner model, and 2) adding components related to DS. These contributions are now detailed.

Multimodal modelling

Inspired by an educational data mining technique to encode sequences of logged interactions using *alphabets* [70], Echeverria's et al. [32] proposed a modelling representation termed the (m-by-n) multimodal matrix (M). Domain knowledge (K) can be used to create rules to encode each modality of data (D) into one or more of the n columns of a matrix. Segments (m rows) are the smallest units of meaning considered for analysis and contain instances of group behaviours. For qualitative data (e.g. discourse), each row could correspond to an utterance. For time series data (e.g. changes in physiological states or gestures) each row might instead represent a time window (e.g. one or more seconds) or critical incidents in the activity. For example, in the illustrative scenario, the stream of (x, y) positioning data is meaningless without a frame of reference (e.g. an indoor map or critical spaces in the classroom). Coordinates could be encoded into classroom zones as columns, such as: *behind the bed*, *bed side*, *bed footer* and *medicine room*. These are meaningful because nurses can perform specific actions in each. Rows can then contain information of the physical space each nurse was at every moment (e.g. every one second) during a simulation.

This enables automated coding of quantitative, low-level data into qualitative, behavioural markers that can be grounded in generic features of teamwork, and the specifics of the activity. Echeverria et al. [32] showed how this structure could be used to create social proxies of activity that invite users to explore the data (V). However, a social proxy is a simple graph aimed at improving awareness instead of communicating insights [33]. As we explain below, the introduction of learner modelling and data storytelling enable us to overlay additional information to guide interpretation and provide more insight.

Learner modelling based on the assessment criteria

Although the multimodal matrix can be useful for mapping from multimodal data to more meaningful information for teachers and students, it is still not possible to generate *explanatory* visualisations. What is missing is a model of the learner to close the interpretative gap. LMs are foundational to provide personalised and adapted support through LA and intelligent tutoring systems (ITS) [11, 18]. A LM is a computational model of what the learner has and has not understood. Generating a LM involves inferring students' knowledge and skills from logged behaviour. This has been conventionally achieved in ITSs by diagnosing behavioural markers (from clickstreams and keystrokes) based on an expert model to identify students' level of mastery of skills, and their misconceptions, in order to adapt the tasks they are given [24].

In our approach, multimodal data streams from the health sim replace clicks and keystrokes. The assessment criteria or pedagogical intentions (A) can be expressed as a set of rules or templates that can be used to identify data points or trends in the data streams that should be communicated to students or teachers. We emphasise the role of assessment since it has been recognised by the LA community as key for aligning pedagogy and the design of LA tools [49].

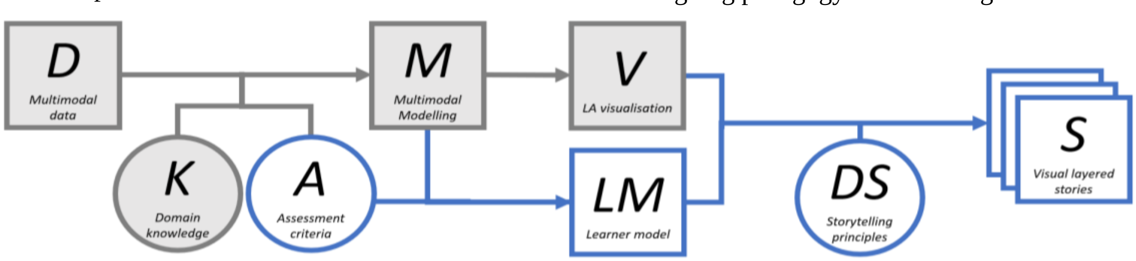

Figure 2: The multimodal layered storytelling approach.

In the illustrative scenario, an assessment criterion could be expressed as follows: "*the airway nurse should be behind the bed clearing the airway of the patient during CPR*". This statement could be converted into a computational series of rules to diagnose the behaviour of the nurses modelled in the multimodal matrix to assess whether someone was *behind the bed* while the manikin detected chest compressions. In this paper, the LM is modelled as a set of numeric parameters that match teacher´s assessment.

Data Storytelling principles.

Ryan [76] and Knaflic [48] identified the following set of DS *principles* that underpin our approach:

DS1. DS is goal oriented. The design of a data story should be aligned with a specific goal. This enables the identification of the data that should be highlighted.

DS2. The data story should rely on a fitting chart type. Some charts work better for certain purposes. For example, line charts effectively show changes over time [76].

DS3. The data story should be stripped down first. Aligned to Tufte's data-ink ratio [94], clutter adds complexity to visual interpretation. Decluttering can be achieved by removing unnecessary headers, borders, grids and data points that do not add informative value.

DS4. The data story should guide attention. Visual and narrative elements can be used to emphasise key data points to create meaning. This can be achieved by 1) *adding enhancements* such as arrows, lines, symbols or enclosures; 2) *changing* colour, contrast or thickness; and 3) *annotating* salient data features or *adding prescriptive titles* that summarise the message of the story.

These DS principles will be illustrated through the layered storytelling prototypes below.

Layered storytelling approach

Doyle proposed that a complex system can be simplified by breaking down it into individual layers [27]. Each *layer* can encapsulate a concept, idea or a small part of the complex system. According to Munzer [65] an approach to handle visual complexity is by combining multiple *layers* within a shared frame. Next, we list a set of principles for crafting layered visualisations of multimodal data.

L1. Each layer should represent a particular view of the learning activity, showing only one data story at a time [42]. This can facilitate the communication of insights while minimising distraction [31].

L2. A static layer should be used as a shared frame. Layers must share visual landmarks to facilitate user orientation as layers are revealed/hidden [65]. Each data story in a layer should be contextualised to the characteristics of the learning activity, such as duration, participants and critical incidents. This can be achieved by keeping the static layer (faded) in the background and overlaying others on top.

L3. The user should be able to select and combine layers. Each layer has its own purpose to fulfil as a part of the complex system [27, 65]. A layer presents a data story that communicates a class of insights about team activity. Teachers and students should be able to flexibly select and combine different layers to facilitate sense making.

L4. Data points highlighted in different layers should be clearly distinguished. As each layer represents different information, data points can be superimposed within the same coordinate system [27]. Therefore, different shapes/colours should be used in each layer.

L5. The content of each layer should be defined based on assessment criteria or pedagogical intentions. Each data story should be aligned with the learning design [38] to effectively provoke reflection on aspects that are relevant and that the teacher would normally provide feedback on.

Next section shows how these principles and components of our approach were materialised in two authentic studies.

STUDIES

Two data storytelling prototypes (1 and 2) were created based on the principles presented above. Each was tested with educators and students in the context of Sims 1 and 2 respectively. This section describes both studies in terms of participants, materials, methods and analysis.

Participants

Eight educators (E1-8, 1 male, 7 females) with extensive experience in simulation training (7-20 years) participated in *Study 1*. Six were staff at the Faculty of Health (2 professors, 2 lecturers – including the one who taught the program, 1 director of simulation, and 1 postdoctoral researcher) and two graduate students leading simulation programs (E4, E8). The study was run as an individual post-hoc reflection with each educator. *Study 2* was run a week after Sim 2 with 39 out of the 44 students involved in it. Students in each of the eight teams were invited to a post-hoc group debrief on their own activity.

Materials and prototypes.

Prototypes 1 and 2 showed insights on the activity of the three teams who performed Sim 1 to educators (e.g. Figure 4, left), and on students' own activity in Sim 2 (e.g. Figure 4, right), respectively. A static layer, a *timeline of actions* (e.g. Figure 3) performed by each nurse during the sim (captured by the manikin and the observer) was used as the background of both prototypes (in line with principle *L2)*. For *Study 1*, *four* layers, each representing a particular view of the multimodal data (*L1*), were generated based on the tasks of Sim 1 (*L5* and *DS1*). These focused on: *i*) *time responsiveness* of nurses performing critical tasks; *ii*) *mistakes made*, or actions performed poorly or in the wrong order; *iii*) *arousal (skin conductance) peaks* detected by an increase of 0.03 μs [7] using EDA Explorer [91]; and *iv*) *positioning* of nurses in key spaces. The prototype was coded using Javascript, with the functionality for users to select one or more layers in combination (*L3*).

For *Study 2*, *six* layers were generated based on the teacher's assessment criteria. Five layers focused on the actions that nurses were intended to perform in Sim 2: *i*) assess vital signs every 10 minutes; and reacting after the patient suffered *ii*) breathing obstruction; *iii*) abdominal pain; *iv*) severe pain; and *v*) nausea. A sixth layer (*vi*) presented insights in terms of nurses' arousal.

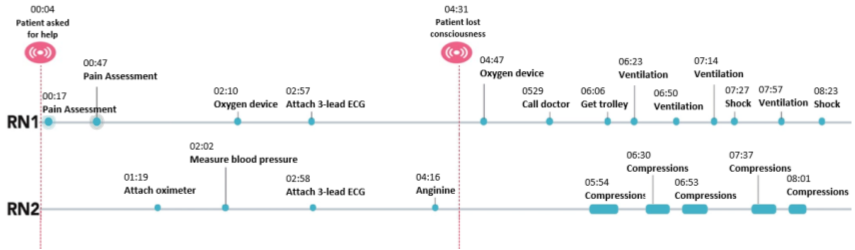

Figure 3: *Timeline of actions* that served as the common reference (background) of Prototypes 1 and 2. Example of a team of two nurses.

Each layer appears on top of the *timeline of actions* (Figure 3) to give context to each data story (*DS2*). The timeline is decluttered, and its opacity is reduced for selected data points to appear emphasised *(L4* and *DS3*). Examples in Figure 4 illustrates how a set of rules (in Table 1) were used to add visual enhancements according to the pedagogical intentions. Principle *DS4* was endorsed by highlighting data points using symbols (A), arrows (B), changes of colour/opacity (C) and enclosures (D). Salient data points were annotated directly (E), and statements were added as annotations or titles (F). A layer to expose the algorithm used to add the enhancements to educators and students was added to each prototype (see Figure 5).

Figure 4 (left) shows how a teacher selected two layers of prototype 1 (G- *mistakes* and *arousal peaks*). Key insights obtained from the LM were communicated, such as RN2 not presenting any arousal peak (F), CPR compressions being too shallow (E) and none of the nurses positioned behind the bed during CPR. Figure 4 (right) communicates results of the assessment of the frequency of vital signs checks contained in the LM. The layer *assess vital signs* (H) provides feedback to students about them either missing or correctly performing the vital signs check (see enclosure and annotations coloured in orange and blue respectively).

Table 1: Example rules used to highlight visual elements.

Data source	Learning intention	Rule pseudocode
Rules used to generate layers in Prototype 1 shown in Figure 4 (left)		
Manikin's logs	The compression depth should be between 5 and 6 cm	if compression_depth < 5: add_annotation("Compressions were shallow. CPR ..") highlight_keyword("shallow", "orange")
Positioning data **Human logs**	A nurse should be located at the top of the bed during CPR	if location_RN != *behind the bed* between(CPR_start_stop): add_annotation(airway) highlight_keyword("behind the bed", "orange")
Physiological data	Nurses should show any arousal state during the activity.	if arousal_peaks(RN) = 0: add_title("RN# did not show any arousal peak") highlight_keyword("arousal peaks", "purple")
Rules used to generate layers in Prototype 2 shown in Figure 4 (right)		
Manikin's logs **Human logs**	Patient's vital signs should be checked at least every 10 minutes	if time_between_two_checks > 10: add_annotation("More than X minutes passed between .....") add_arrows(annotation, data_point1, data_point2) highlight_datapoints(data_point1, data_point2, "orange") add_shaded area(data_point1, next_point, "orange") add_title("Remember to take vital signs every 10 minutes") else: add_annotation("Here you did it right! Well done!") highlight_datapoints(data_point_list, "blue")

Method

Both studies were conducted using LATEP, an elicitation protocol for understanding how non-data experts envisage the use of LA systems [58]. Based on this, our studies aimed at investigating: 1) the *added value* of making multimodal data visible through data stories; 2) the anticipated in-classroom *orchestration* of the tools; and 3) the potential impact on students' *accountability*. Inspired by the growing interest on explainable AI [96], we also sought to understand 4) the implications of *exposing the algorithms* used for crafting the stories.

Study 1 was conducted as 45-minutes individual interviews with educators (E1-8) and Study 2 as 30-minutes focus groups with students (S1-5) of each team (T1-8). This paper focuses on the next four tasks both educators and students were asked to do. Educators and students were first asked to think-aloud while exploring the *timeline of actions* (without layers) of three student teams and their own team respectively. Secondly, all participants explored the layers in no specific order. Then, they were asked about the added value of the layers, individually or in combination (*aim 1*). Third, they were asked about orchestration and accountability opportunities and concerns (*aims 2, 3*), as follows: *i*) how can the system be used in the classroom? and *ii*) who should be able see the interface, for which purpose and in what form? Finally, participants were asked to review the layer that exposes the algorithm (see Figure 5). Three questions were asked regarding: who should *i*) define, *ii*) see, and *iii*) change the rules (*aim 4*).

Analysis

Two researchers were present in each session, which were video-recorded and transcribed. We examined participants' statements and their actions exploring the prototypes.

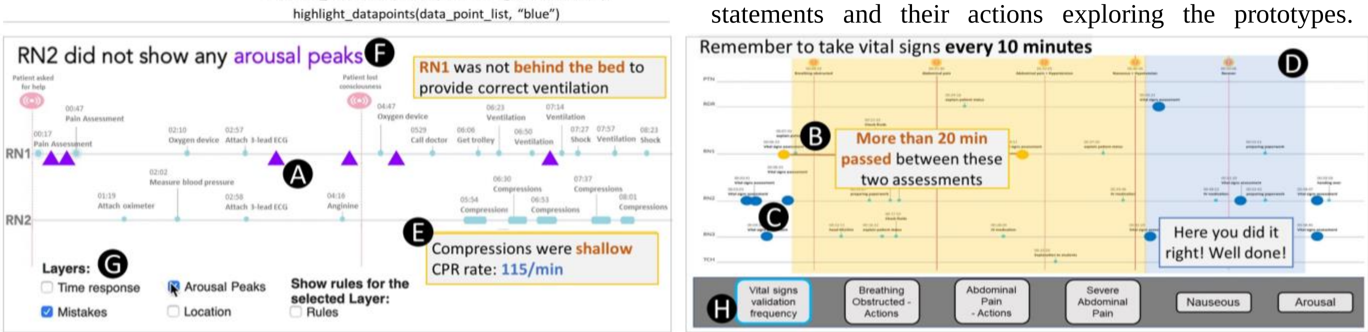

Figure 4: Prototypes of the layered storytelling interfaces (top). Left: Prototype 1 - layers *mistakes* and *arousal peaks* for a team of two nurses. Right: Prototype 2- layer *vital signs validation frequency* for a team of five nurses.

Following best practices of qualitative research in HCI ([61], p. 13), and given the direct alignment between the study protocol and the analysis themes, statements of interest were jointly coded [8] by two researchers according to the pre-set themes of the study protocol: a) *added value of the layers*; b) *anticipated classroom use*; c) *accountability*; and d) *algorithmic transparency*. Then, resulting coded statements were examined by authors who had several discussions to select instances that effectively illustrate opportunities and concerns of the approach to create MMLA interfaces that communicate insights.

RESULTS

This section examines evidence from Studies 1 and 2, organised by themes, that reflect the aims of the studies.

Added value of the layers

All educators agreed that the storytelling layers added significant value to the *timeline of actions*. Two educators explained that *"[the layers are] absolutely more valuable, because each tells a story behind the sim"* [E3] and *"can help students improve their practice"* [E7]. E7 added that *"the layers work really well together to give a really good picture of team performance"*. Students in all teams also appreciated the value of the layers to make reflection "*clearer*" [T2, S2] and "*easier*" [T3, S1]. Student [T3, S2] explained that *"[the layers] divide the data into sections. [Without layers] all actions look the same, and it doesn't say what timeframes should have been [considered]"*.

While exploring the layers that highlight delays and errors (e.g. Figure 4, left), three educators stated that time is "*critical*" and "*crucial*" in high-stakes situations. E8 explained: "*Time to detection, compression and shock are all crucial during a cardiac arrest"*. E5 also mentioned that "*It is important to identify when the team didn't do things in the right order or correctly"*. Two educators explained this is the kind of feedback they often would like to provide to students, but it is hard for them to assess the situation during class. E7 explained: "*It's good for students to know that compressions needed to be deeper as shown here* (see Figure 4, E). *If I was there, I could've only guessed that they were not deep enough"*. While exploring layers *i-v* of Prototype 2 (e.g. Figure 4, right), some students also stated this is the kind of feedback they need but they rarely receive. Student [T3, S4] explained this as follows: *"we always do things, but we're not sure if what we did was right or not. Teachers can't be there the whole time"*.

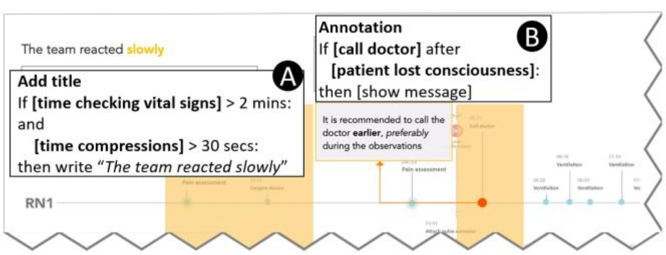

Figure 5: Partial view of the layer that exposes the rule-based algorithm used to add enhancements (e.g. A- a prescriptive title and B- an annotation) to layer *i-time responsiveness*, Prototype 1.

Educators reflected on the potential of the *arousal* layer to provoke reflection stress management, critical in nursing education [64]. E7 explained: "*this layer can give you insights into how the nurse was feeling"*. Educators illustrated how they would investigate students' increased arousal (e.g. "*was there something at this particular time that make you feel stressed?* [E2]); or if arousal was caused by a particular situation (e.g. "*you were doing pain assessment and then you were about to put the oxygen delivery device. What was the trigger there?"* [E1]).

Students reflected on their arousal in relation to their roles (e.g. *"I was pretty chill being the observer"* [T1, S4]; *"As I was the anaesthetist, once my job was done I was relaxed"* [T3, S3]) or actions they performed ("*I was stressed about the handover"* [T1, S3] or "*clearing the airway*" [T6, S1]). Some suggested ways in which this can trigger reflection. For example, student [T1, S1] said: "*I was so stressed. Maybe I should have delegated better? That was part of the feedback I got after the sim*". Another student explained that showing the arousal layer on top of the *actions* provides a more complete view of the simulation: "*it shows that stuff was still happening all through the sim because people are stressing throughout the scenario"* [T3, S3].

Some concerns were highlighted by educators. E3 noted that false negatives in the *mistakes* layer would be misguiding: *"if the layer did not highlight all the mistakes, it could make students think they performed well when in fact they missed critical things"*. Another concern was that some layers may be useful for certain sims but not for others. E5 stated this as follows: *"The time response [layer] can be very useful in a cardiac arrest simulation but not necessarily for other sims*" [E5]. Finally, concerns on the accuracy of the wristbands in detecting arousal were raised. Two educators said: *"The arousal doesn't match with what I would expect"* [E6]; and *"I'd expect to see a peak before they do anything"* [E8]. In addition, E3 and E6 pointed out that showing arousal peaks without certainly knowing their cause could lead to misinterpretations: *"You could think that [students] are calm, or that they're completely not interested. There's no gauging on what is going on"* [E6].

In sum, results showed how educators and students perceived the added value of the layers, as well as potential concerns to be considered in the design of LA interfaces that seek to communicate insights. The next section explores the potential in-classroom uses of the prototypes.

In-classroom orchestration opportunities

Educators suggested potential strategies to use the interface effectively. All educators agreed on the potential of providing feedback to students based on objective evidence. E8 expressed: "*I can tell students: this is what you did well, this is what you did poorly. This is what you can improve, you can do X, Y and Z"*. E5 explained she could "*make bolder statements that cannot [easily] be disputed*". This is part of the evidence-based culture in nursing education [43] in which "*capturing objective data on clinical scenarios is* *so important"* [E8]. Students also appreciated the evidence provided by the data stories. A student stated: "*It's not just that we were late, but we were late by that much time. We can precisely see how much we have to improve, and what we have done well. It makes a big difference!"* [T1, S2].

Some educators suggested the interface could be accessed individually by students to do a "*homework reflection activity*" [E1, E2]. However, all educators favoured the use of the tool to support teacher-led in-class debriefing. Most offered examples of questions that they would ask to spark conversations on: individual performance (e.g. *"Am I doing a good job? Am I getting things done?"*[E7]); arousal (*"Can you take me through what you were feeling at this point?"*[E2]); mistakes made (*"So, tell me a bit about how deep the compressions need to be? have you achieved that?"*[E2]); response time (*"when should you provide your first shock?"*[E8]); and positioning (*"What were you doing over here? Because you didn't go and get the resus trolley"* [E8]). Students also supported the idea of using the layered interface as a debriefing tool, but it "*needs to be facilitated by an instructor*" [T1, E4] as a "*group reflection*" [T3, E1].

Another potential role for the interface is in the re-design of simulations. E1, E2 and E5 explained that they could identify common mistakes across various teams of students (e.g. *"The interface can help us identify if a lot of groups of students are making the same errors"* [E5]). Teachers could then reflect on their teaching practice to address the most common errors, for example, by "*re-structuring a scenario, having a discussion around the concerns highlighted, and the things that are needed to work with teams"* [E1].

Finally, some educators mentioned both opportunities and risks of using the interface for summative assessment. For example, E8 noted that the interface could be used to *"provide evidence to determine competency"*. E3 also explained that *"one of the issues with high-stakes assessment, is making sure it is valid and consistent; and objective rather than subjective"*. However, E7 argued that using the system for summative assessment is against the intention of simulation. She explained that: "*Simulation is not a (summative) assessment. It is to give formative feedback to help students improve their practice"*.

In short, both educators and students strongly highlighted the potential benefit of using the layered interface to augment the teachers' capabilities in leading evidence-based group in-class debriefings, or with suitable scaffolding, a personal reflection task assigned to students.

Accountability

A consequence of making evidence about teamwork visible is that people can be made responsible for their actions [33]. This was recognised by educators in our study. E8 explicitly stated that: *"these layers keep people more accountable"*. To minimise misuse, educators explained that the data captured in one sim should only be accessed by the students and teachers involved in it. E5 explained: "*I would not go and talk about [this information] with other tutors or students. I would only talk about this with the students of my class*.". Yet, educators also recognised that, in practice, the tool would need to be used to lead reflection with multiple teams in the classroom and that this can indeed be beneficial for students to reflect on others' data. E5 explained that this creates a "*good opportunity to discuss teams' performance with the whole classroom*".

Educators raised some concerns about other students or external people looking at the layers. Students could feel "*a bit uncomfortable*" [E7] or could focus on "*comparing their performance with other teams"* [E3, E5], especially "*low performers"* [E3]. Two strategies were suggested to address this. The information could be kept strictly "*confidential"* [E3] (teams could only access their own information), but this would mean other students would not be able to learn from others' experiences. Alternatively, the interface could be de-identified and "*shared with other students within the same classroom, to avoid criticism*" [E2]. A teacher suggested that she could conduct a classroom reflection without pointing at students directly by: "*picking a random team, having a look at the interface and seeing what went well and explain what can be improved for next time"* [E2].

In contrast, we were interested to find that students were not concerned about others looking at their data. All but one student was keen for other students to reflect on their data stories if it "*helps in their learning*" [T3, S1], especially if "*the teacher sits down with them and talks it through*" [T4, S2]. The only concerned student explained that "*others would find it very boring because it doesn't relate to them*" [T1, S2]. Surprisingly, none of the students raised concerns in sharing their arousal traces "*as long as it doesn't include [their] names*" [T5, S1]. However, some students mentioned that if they knew before the sim that their data would be shown to others, it would affect their performance. One student explained: "*I would feel like being in an assessment. It would be so stressed*" [T6, S3].

In sum, prescriptive data stories can raise accountability and privacy concerns. However, educators and students suggested ways to address some of these through pedagogical strategies and interface design features.

Algorithmic transparency and manipulability

All educators agreed that teachers and students should be able to see the *if-then* rules, as an additional layer that could be *"turned on or off"* [E2] as in our prototypes. Educators considered that showing the rules during the debrief could help them guide the discussion by contrasting team's performance with the learning expectations. E1 explained this as follows: "*It might be useful to say in the debrief: 'here is the time it took you to complete this action and these are the rules that have been set to that. Do you think that's a reasonable performance?'"*.

Educators also indicated that showing the rules to students could help them recognise the gap between their personal expectation and the standard to which they should perform. Although *"students commonly have a timeframe in their mind"* [E7] and they may *"have a clue on how to fulfil a task"* [E2], students often do not have the perception of time and how critical the situation they are handling is [E7]. As one educator suggested: "*These rules are a reminder [for students] to compare their steps and the processes they did and what was required in the simulation"* [E8].

In contrast, not all students understood the rules initially, some needing additional explanation (e.g. "*I was a bit confused. But after you explained it, I got it. It's similar to coding. If we achieve 'this', you're going to show 'that'"* [T8, S5]). After understanding the rules, all students appreciated the importance of knowing how data stories were generated. One student explained that understanding the algorithm can be useful for reflection, as follows: "*Looking at these rules is useful because it's like if you ask a question and someone says no, you're like, well why is that wrong? Like this is constructive feedback or whatever it's called"* [T7, S1]. However, although some students mentioned that they would not mind seeing the rules, most agreed that they prefer to only reflect on the data stories (e.g. "*Can't you just incorporate the rules into each [layer] instead of doing it in an extra layer? Like explaining: 'this is [highlighted] in orange because…'*" [T8, S3]).

In terms of the flexibility to *change the rules*, all educators agreed that the rules should be defined by the teacher who coordinates the subject. For instance, E3 commented that *"a tutor may want five key things that matter the most in a simulation, so there could be rules related to each of those five things"*, which reflects how Prototype 2 was designed matching the teacher's assessment criteria items. E1 also explained that *"[there should be] flexibility so the coordinator could adapt the rules slightly"*. Two educators suggested that rules' parameters could be manipulated according to the level of experience of students. E6 explained: "*I would expect final-year nurses get their things done perfectly, more so than a first-year student"*. In contrast, two educators agreed that rules should remain the same. E7 explained: "*rules shouldn't change to accommodate the student. Instead, these rules should be seen as a teaching practice to help students [understand] what needs to be achieved*". Another idea was that rules should be defined based on international guidelines. E8 explained: "*It's the potential between saving a life and not saving a life. This is what's needed to be done and this is how quickly you need to do it".* Students mentioned they would not try to change the rules ("*I would like to keep [the parameters] like that*" [T5, S1]). Two students stated that teachers would be in a better position to "*understand the parameters*" [T5, S2]) and that "*basic [concepts of] coding*" [T8, S1] would be needed to change them.

In sum, both educators and students appreciated being able to see the rules. Educators appreciated the idea of being able to both interrogate and modify rules, and envisaged potential ways to provoke reflection. In contrast, most students preferred to only interact with the consequences of the rules (the data stories) without needing to see the rules.

DISCUSSION

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

The potential of the layered storytelling approach

In Dewey's [22] view, reflection involves the observation of the experience retrospectively to discern explanations for what happened. We heard consistently positive responses from educators and students, that the way in which data was communicated provoked **deeper reflection** on simulation experiences. In contrast to the current educational situation — in which debriefings are dependent on expert educator observation (but often stretched over multiple teams), and the partial (sometimes stressed, and always biased) memories of students — they recognised the value of capturing **objective evidence** of collocated activity and rendering it visible as data stories. Educators also envisaged the use of this same evidence to **review their instructional designs**. This is consistent with Lockyear et. al. [51] proposal that analytics-enabled feedback should support curriculum planning and redesign.

Although the utility of some layers can be tied to particular simulations (e.g. the *time responsiveness* layer is more relevant in a CPR sim than in non-emergency situations), from an educational point of view, this is expected. Learning is an epistemically, socially and physically situated activity strongly shaped by the instructional design [36]. **Context specificity is precisely the rationale for aligning the data stories with the educators' desired learning outcomes**. This can explain, for example, how a short delay in performing an action may be a critical error in some cases (e.g. a resuscitation scenario) and not having much importance in others (e.g. in non-life threatening situations). Each layer can thus be useful in a family of learning situations that share similar pedagogical intentions. We illustrated this with the *mistakes* and *arousal* layers which were instantiated differently in both studies.

We argue, therefore, that there is no reason in principle why the components of our approach could not enable interfaces beyond nursing simulation, to other domains with well-defined protocols (e.g. firefighting [73] and medical education [14]), or to contexts using multimodal learning analytics to model complex behaviours, such as collocated collaborative learning [20], designing in maker-spaces [98], or teacher activity in the classroom [71].

Risks from bias and errors

Like any symbolic representation, visualisations are never neutral [25], but this is particularly dangerous if they are endowed with an aura of objectivity that disguises biases [90]. As we have detailed, *explanatory* visualisations seek intentionally to **reduce complexity by focusing users'**

attention on specific target features. This is made very clear to users, but they may still contain biases and errors, whose risks must be assessed in context. To illustrate, while some educators were interested in the potential of the approach for **assessing teamwork summatively**, others were concerned about the **risk of false negatives** in the *mistakes* and *arousal* layers. Although false negatives were not identified in any of our prototypes, multimodal data stories can certainly introduce bias in the way annotations are written, how rules are created, and how accurately sensors capture data. We concur with this more cautious view, since the validity and reliability required for automated grading is significantly beyond the current maturity of this infrastructure. We lower the stakes by focusing on formative feedback, to provoke deeper reflection and dialogue, in which the human agents determine the ultimate meaning and consequences.

Human-centered design, and its tensions

The adoption of human-centered design is an obvious way to address concerns such as the above [63]. The potential damage of false negatives (and positives) in automated feedback can be reduced by engaging teachers and students in deciding what information is to be included in the learner model and storytelling interfaces, and iteratively prototyping to gauge the risks. That being said, it has also been argued that human-centered learning analytics faces the novel situation that (unlike most HCI contexts) the end-users are not experts in the task at hand: students by definition have not mastered the domain of study and invariably know very little about the processes of *learning* in a formal sense [10]. Thus, researchers and designers must discern when to accommodate students' views on what will help them learn and when to favour educators' views.

An interesting example arose in the context of the *mistakes* layer. Currently, there is much debate around **algorithmic transparency** in both academic and mainstream discourse. While many of the algorithms required to generate the storytelling layers are hidden from users, we experimented with the selective exposure of rules driving the *mistakes* layer. This was met with enthusiasm from some educators who saw potential in manipulating them for pedagogical purposes, while most students saw no particular value in seeing the rules. Future work should investigate whether good tool+activity design can help resolve these tensions, i.e., through tools and activities that *embrace imperfection*, which can be effective despite known imperfections of data, algorithms, and models of learning [47].

Data privacy

Finally, data privacy concerns were raised by educators and students. Educators' proposed uses of the tool beyond the classroom raised questions around **data ownership, data sharing and de-identification**, and how to notify students about when, why and by whom their data is being used. While privacy guidelines for systems that expose student data exist [28, 69, 72, 85], most focus on online systems in which the exploration of the data is often detached from the physical spaces in which it was collected. But capturing multimodal team data raises some acute concerns. For example, sensor data have a personal dimension to it not found in the more abstracted data from clickstreams, such as physiology, posture, gaze and movement; and interfaces that communicate insights on teamwork would directly reveal such information at least to other team members.

Limitations and future work

A scope limitation concerns the degree of formal task structure required for this interface to work. In nursing simulations, task procedures are well established and are often measurable using quantitative parameters. Since the approach has been applied in different simulations, there is confidence that it is applicable in simulation-based training situations. However, in cases such as open-ended collaborative learning, or meetings, the lack of formal protocols makes it much harder to formalise the learner model and rules to customise the visualisations. Future work could investigate to what extent our approach can still add value to more open-ended educational scenarios.

A technical limitation concerns the workflow architecture. Although the data capture and visualisation are fully automated, the data integration from sensors still requires human intervention (e.g. EDA peaks are automatically detected and then a researcher runs a script to plot them on the timeline). The MMLA community continues to develop pipelines to automate the transformation of multimodal data ([17, 23, 79] reviewed in [83]) that should benefit our work. Moreover, the challenge of programmatically annotating and enhancing charts includes not only content selection and generation [39, 62, 75], but also automated layout of visual elements [9, 45, 87]. These advances should help to move our infrastructure towards better data integration, automated annotation, and new visualisation techniques.

Finally, while the storytelling design approach is generic, the designs reported in this paper were specifically tuned to the requirements and forms of feedback our stakeholders co-designed with us. Exploring other ways of combining layered information, using alternative visualisation techniques and including quantification of metrics in the layers are all promising areas for future work.

CONCLUSION

As humans struggle with the data tsunami, we are now awash with dashboard prototypes and products. Yet, there is growing evidence that these are far from intuitive. This paper documents how we have wrestled with the challenge of designing activity-based feedback visualisations which draw the attention of non-technical users to key insights in the data. We argue, supported by user studies, that this work advances the state of the art in making multimodal data streams intelligible to non-data experts. The approach should enable similar collocated activities to benefit from these novel collaboration analytics.

REFERENCES

[1] Sara Alspaugh, Anna Swigart, Ian MacFarland, R Katz and Marti Hearst. 2015. *Rethinking the first look at data by framing it*. Technical Report UCB/EECS-2015-208, University of California at Berkeley.

[2] David Beer. 2018. Envisioning the power of data analytics. *Information, Communication & Society*. 21, 3, 465-479. https://doi.org/10.1080/1369118X.2017.1289232.

[3] Paulo Blikstein and Marcelo Worsley. 2016. Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*. 3, 2, 220-238. https://doi.org/10.18608/jla.2016.32.11.

[4] Paulo Blikstein and Marcelo Worsley. 2018. Multimodal learning analytics and assessment of open-ended artifacts. In *Learning Analytics in Education*, David Niemi, Roy D. Pea, Bror Saxberg and Richard E. Clark (Eds.). IAP, Charlotte, USA, 89.

[5] Robert Bodily and Katrien Verbert. 2017. Trends and issues in student-facing learning analytics reporting systems research. In *Proceedings of the International Conference on Learning Analytics and Knowledge, LAK'17*, 309-318. https://doi.org/10.1145/3027385.3027403.

[6] Robert Bodily, Judy Kay, Vincent Aleven, Ioana Jivet, Dan Davis, Franceska Xhakaj and Katrien Verbert. 2018. Open learner models and learning analytics dashboards: a systematic review. In *Proceedings of the International Conference on Learning Analytics and Knowledge, LAK'18*, 41-50. https://doi.org/10.1145/3170358.3170409.

[7] Jason J. Braithwaite, Derrick G. Watson, Robert Jones and Mickey Rowe. 2013. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*. 49, 1, 1017-1034.

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology*. 3, 2, 77-101. https://doi.org/10.1191/1478088706qp063oa.

[9] Chris Bryan, Kwan-Liu Ma and Jonathan Woodring. 2017. Temporal summary images: An approach to narrative visualization via interactive annotation generation and placement. *IEEE Transactions on Visualization and Computer Graphics*. 23, 1, 511-520. https://doi.org/10.1109/TVCG.2016.2598876.

[10] Simon Buckingham Shum, Rebecca Ferguson and Roberto Martinez-Maldonado. 2019. Human-Centred Learning Analytics. *Journal of Learning Analytics*. 6, 2, 1-9. https://doi.org/10.18608/jla.2019.62.1.

[11] Susan Bull and Judy Kay. 2016. SMILI☺: a Framework for Interfaces to Learning Data in Open Learner Models, Learning Analytics and Related Fields. *International Journal of Artificial Intelligence in Education*. 26, 1, 1-39. https://doi.org/10.1007/s40593-015-0090-8.

[12] Kenneth H Butler and Brian Clyne. 2003. Management of the difficult airway: alternative airway techniques and adjuncts. *Emergency Medicine Clinics*. 21, 2, 259-289. https://doi.org/10.1016/S0733-8627(03)00007-5.

[13] Davide Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Hans-Jörg Schulz, Marc Streit and Christian Tominski. 2016. Characterizing guidance in visual analytics. *IEEE Transactions on Visualization Computer Graphics*. 23, 1, 111-120. https://doi.org/10.1109/TVCG.2016.2598468.

[14] Teresa Chan, Stefanie Sebok-Syer, Brent Thoma, Alyssa Wise, Jonathan Sherbino and Martin Pusic. 2018. Learning analytics in medical education assessment: The past, the present, and the future. *AEM education training* 2, 2, 178-187. https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Sherbino%2C+Jonathan.

[15] Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang and Kwan-Liu Ma. 2019. TalkTraces: Real-Time Capture and Visualization of Verbal Content in Meetings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'19*, 577:571--577:514. https://doi.org/10.1145/3290605.3300807.

[16] Qing Chen, Zhen Li, Ting-Chuen Pong and Huamin Qu. 2019. Designing Narrative Slideshows for Learning Analytics. In *Proceedings of the IEEE Pacific Visualization Symposium, PacificVis'19*, 237-246. https://doi.org/10.1109/PacificVis.2019.00036.

[17] Yi Han Victoria Chua, Preman Rajalingam, Seng Chee Tan and Justin Dauwel. 2019. EduBrowser: A Multimodal Automated Monitoring System for Co-located Collaborative Learning. In *Learning Technology for Education Challenges*, Lorna Uden, Dario Liberona, Galo Sanchez and Sara Rodríguez-González (Eds.). Springer, Cham, 125-138.

[18] Albert T Corbett, Kenneth R Koedinger and John R Anderson. 1997. Intelligent tutoring systems. In *Handbook of human-computer interaction*, Marting G. Helander, Thomas K. Landauer and Prasad V. Prabhu (Eds.). Elsevier, 849-874. https://doi.org/10.1016/B978-044481862-1.50103-5.

[19] Rodney N. Cuff. 1980. On casual users. *International Journal of Man-Machine Studies*. 12, 2, 163-187. https://doi.org/10.1016/S0020-7373(80)80016-2.

[20] Mutlu Cukurova, Rose Luckin, Eva Millán and Manolis Mavrikis. 2018. The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education*. 116, Jan 2018, 93-109. https://doi.org/10.1016/j.compedu.2017.08.007.

[21] Dan Davis, Ioana Jivet, René F Kizilcec, Guanliang Chen, Claudia Hauff and Geert-Jan Houben. 2017. Follow the successful crowd: raising MOOC completion rates through social comparison at scale. In *Proceedings of the International Conference on Learning Analytics & Knowledge, LAK'17*, 454-463. http://dx.doi.org/10.1145/3027385.3027411.

[22] John Dewey. 1997. *How we think*. Courier Corporation, Chelmsford, USA.

[23] Daniele Di Mitri, Jan Schneider, Roland Klemke, Marcus Specht and Hendrik Drachsler. 2019. Read Between the Lines: An Annotation Tool for Multimodal Data for Learning. In *Proceedings of the International Conference*

on Learning Analytics and Knowledge, LAK'19, 51-60. https://doi.org/10.1145/3303772.3303776.

[24] Pierre Dillenbourg and John Self. 1992. A Framework for Learner Modelling. *Interactive Learning Environments*. 2, 2, 111-137. https://doi.org/10.1080/1049482920020202.

[25] Marian Dörk, Patrick Feng, Christopher Collins and Sheelagh Carpendale. 2013. Critical InfoVis: exploring the politics of visualization. In *Proceedings of the CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 2189-2198. https://doi.org/10.1145/2468356.2468739.

[26] Rodrigo Guimarães dos Santos Almeida, Beatriz Maria Jorge, Valtuir Duarte Souza-Junior, Alessandra Mazzo, José Carlos Amado Martins, Elaine Cristina Negri and Isabel Amélia Costa Mendes. 2018. Trends in research on simulation in the teaching of nursing: an integrative review. *Nursing education perspectives*. 39, 3, E7-E10. http://dx.doi.org/10.1097/01.NEP.0000000000000292.

[27] John C Doyle and Marie %J Proceedings of the National Academy of Sciences Csete. 2011. Architecture, constraints, and behavior. In *National Academy of Sciences 108, Supplement 3*, Donald W. Pfaff (Ed.), Rockefeller University, New York, NY, 15624-15630. https://doi.org/10.1073/pnas.1103557108.

[28] Hendrik Drachsler and Wolfgang Greller. 2016. Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. In *Proceedings of the International Conference on Learning Analytics & Knowledge, LAK'16*, 89-98. https://doi.org/10.1145/2883851.2883893.

[29] Brent Dykes. 2015. Data storytelling: What it is and how it can be used to effectively communicate analysis results. *Applied Marketing Analytics*. 1, 4, 299-313.

[30] V. Echeverria, R. Martinez-Maldonado, R. Granda, K. Chiluiza, C. Conati and S. Buckingham Shum. 2018. Driving Data Storytelling from Learning Design. In *Proceedings of the International Conference on Learning Analytics and Knowledge*, *LAK'18*, 131-140. https://doi.org/10.1145/3170358.3170380.

[31] Vanessa Echeverria, Roberto Martinez-Maldonado, Simon Buckingham Shum, Katherine Chiluiza, Roger Granda and Cristina Conati. 2018. Exploratory versus Explanatory Visual Learning Analytics: Driving Teachers' Attention through Educational Data Storytelling. *Journal of Learning Analytics*. 5, 3, 72-97. https://doi.org/10.18608/jla.2018.53.6.

[32] Vanessa Echeverria, Roberto Martinez-Maldonado and Simon Buckingham Shum. 2019. Towards Collaboration Translucence: Giving Meaning to Multimodal Group Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'19*, 39, 31-16. https://doi.org/10.1145/3290605.3300269.

[33] Thomas Erickson and Wendy A. Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction*. 7, 1, 59-83. https://doi.org/10.1145/344949.345004.

[34] Stephen Few. 2004. *Show me the numbers: Designing Tables and Graphs to Enlighten*. Analytics Press, Burlingame, CA, USA.

[35] Kristin L Fraser, Michael J Meguerdichian, Jolene T Haws, Vincent J Grant, Komal Bajaj and Adam Cheng. 2018. Cognitive Load Theory for debriefing simulations: implications for faculty development. *Advances in Simulation*. 3, 1, 1-28. https://doi.org/10.1186/s41077-018-0086-1.

[36] Peter Goodyear and Yannis Dimitriadis. 2013. In medias res: reframing design for learning. *Research in Learning Technology*. 21, 19909 1-13. http://dx.doi.org/10.3402/rlt.v21i0.19909.

[37] Richard H. R. Harper. 2019. The Role of HCI in the Age of AI. *International Journal of Human–Computer Interaction*. 35, 15, 1331-1344. https://doi.org/10.1080/10447318.2019.1631527.

[38] Davinia Hernandez-Leo, Roberto Martinez-Maldonado, Abelardo Pardo, Juan Muñoz-Cristobal and Maria Rodriguez-Triana. 2016. Analytics for Learning Design: A Layered Framework and Tools *British Journal of Educational Technology*. 50, 1, 139–152. https://doi.org/10.1111/bjet.12645.

[39] Jessica Hullman, Nicholas Diakopoulos and Eytan Adar. 2013. Contextifier: automatic generation of annotated stock visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *CHI'13*, 2707-2716. https://doi.org/10.1145/2470654.2481374.

[40] Noah Iliinsky and Julie Steele. 2011. *Designing data visualizations: representing informational relationships*. O'Reilly Media, Inc., Sebastopol, CA, USA.

[41] Kori Inkpen, Stevie Chancellor, Munmun De Choudhury, Michael Veale and Eric PS Baumer. 2019. Where is the Human?: Bridging the Gap Between AI and HCI. In *Proceedings of the CHI'19 Extended Abstracts on Human Factors in Computing Systems*, W09. https://doi.org/10.1145/3290607.3299002.

[42] Daryl Inniss and Rubenstein Roy. 2017. Layers and the Evolution of Communications Networks. In *Silicon Photonics*, Daryl Inniss and Rubenstein Roy (Eds.). Morgan Kaufmann, Oxford, 21-40. https://doi.org/10.1016/B978-0-12-802975-6.00002-8.

[43] Mary Ireland. 2008. Assisting students to use evidence as a part of reflection on practice. *Nursing Education Perspectives*. 29, 2, 90-93. https://doi.org/10.1097/00024776-200803000-00009.

[44] Ioana Jivet, Maren Scheffel, Marcus Specht and Hendrik Drachsler. 2018. License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the International Conference on Learning Analytics and Knowledge, LAK'18*, 31-40. https://doi.org/10.1145/3170358.3170421.

[45] Eser Kandogan. 2012. Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology, VAST'12*, 73-82. https://doi.org/10.1109/VAST.2012.6400487.

[46] Jong W. Kim, Robert A. Sottilare, Keith Brawner and Timothy Flowers. 2018. Integrating Sensors and Exploiting Sensor Data with GIFT for Improved Learning Analytics. In *Proceedings of the Annual GIFT Users Symposium, GIFTSym6*, 299-312.

[47] Kirsty Kitto, Simon Buckingham Shum and Andrew Gibson. 2018. Embracing imperfection in learning analytics. In *Proceedings of the International Conference on Learning Analytics and Knowledge, LAK'18*, 451-460. https://doi.org/10.1145/3170358.3170413.

[48] Cole Nussbaumer Knaflic. 2015. *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, Hoboken, NJ, USA.

[49] Simon Knight, Simon Buckingham Shum and Karen Littleton. 2014. Epistemology, assessment, pedagogy: where learning meets analytics in the middle space. *Journal of Learning Analytics*. 1, 2, 23-47. https://doi.org/10.18608/jla.2014.12.3.

[50] B. Lee, N.H. Riche, P.; Isenberg and S. Carpendale. 2015. More than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications*. 35, 84-90. https://doi.org/10.1109/MCG.2015.99

[51] Lori Lockyer, Elizabeth Heathcote and Shane Dawson. 2013. Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*. 57, 10, 1439-1459. https://doi.org/10.1177/0002764213479367.

[52] Steven Lonn, Stephen J Aguilar and Stephanie D Teasley. 2015. Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*. 47, 90-97. https://doi.org/10.1016/j.chb.2014.07.013.

[53] Stephen MacNeil, Kyle Kiefer, Brian Thompson, Dev Takle and Celine Latulipe. 2019. IneqDetect: A Visual Analytics System to Detect Conversational Inequality and Support Reflection during Active Learning. In *Proceedings of the ACM Conference on Global Computing Education*, 85-91. https://doi.org/10.1145/3300115.3309528.

[54] Karin Maier, Philipp Leitner and Martin Ebner. 2019. Learning Analytics Cockpit for MOOC Platforms. In *Emerging Trends in Learning Analytics*, Myint Swe Khine (Ed.), Brill Sense, Leiden, Netherlands, 271-285. https://doi.org/10.1163/9789004399273_014.

[55] Katerina Mangaroska and Michail N Giannakos. 2018. Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, (in press), 1-19. https://doi.org/10.1109/TLT.2018.2868673.

[56] Bette Mariani and Jessica Doolen. 2016. Nursing simulation research: What are the perceived gaps? *Clinical simulation in nursing*. 12, 1, 30-36. https://doi.org/10.1016/j.ecns.2015.11.004.

[57] Roberto Martinez-Maldonado, Kalina Yacef and Judy Kay. 2015. TSCL: A conceptual model to inform understanding of collaborative learning processes at interactive tabletops. *International Journal of Human-Computer Studies*. 83, november 2015, 62-82. http://dx.doi.org/10.1016/j.ijhcs.2015.05.001.

[58] Roberto Martinez-Maldonado, Vanessa Echeverria, Doug Elliott, Carmen Axisa, Tamara Power and Simon Buckingham Shum. 2019. Designing translucent learning analytics with teachers: an elicitation process. *Interactive Learning Environments*, (in press), 1-30.

[59] Roberto Martinez-Maldonado, Judy Kay, Simon Buckingham Shum and Kalina Yacef. 2019. Collocated Collaboration Analytics: Principles and Dilemmas for Mining Multimodal Interaction Data. *Human–Computer Interaction*. 34, 1, 1-50. https://doi.org/10.1080/07370024.2017.1338956.

[60] Wannisa Matcha, Dragan Gasevic and Abelardo Pardo. 2019. A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies*, (in press), 1-20. https://doi.org/10.1109/TLT.2019.2916802.

[61] Nora McDonald, Sarita Schoenebeck and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM: Human-Computer Interaction (PACM: HCI). CSCW*, (in press).

[62] Ronald Metoyer, Qiyu Zhi, Bart Janczuk and Walter Scheirer. 2018. Coupling Story to Visualization: Using Textual Analysis as a Bridge Between Data and Interpretation. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI'18*, 503-507. https://doi.org/10.1145/3172944.3173007.

[63] Miriah Meyer and Jason Dykes. 2019. Criteria for Rigor in Visualization Design Study. *IEEE Transactions on Visualization and Computer Graphics*, (in press), 1-13. http://dx.doi.org/10.1109/TVCG.2019.2934539.

[64] Lars Müller, Verónica Rivera-Pelayo, Christine Kunzmann and Andreas Schmidt. 2011. From Stress Awareness to Coping Strategies of Medical Staff: Supporting Reflection on Physiological Data. In *Proceedings of the International Workshop on Human Behavior Understanding*, 93-103. https://doi.org/10.1007/978-3-642-25446-8_11.

[65] Tamara Munzner. 2014. *Visualization Analysis and Design*. CRC Press / Taylor & Francis Group, Boca Raton, FL.

[66] Omid Noroozi, Iman Alikhani, Sanna Järvelä, Paul A Kirschner, Ilkka Juuso and Tapio Seppänen. 2019. Multimodal data to design visual learning analytics for understanding regulation of learning. *Computers in Human Behavior*. 100, Nov 2019, 298-304. https://doi.org/10.1016/j.chb.2018.12.019.

[67] Xavier Ochoa, Katherine Chiluiza, Roger Granda, Gabriel Falcones, James Castells and Bruno Guamán. 2018. Multimodal Transcript of Face-to-Face Group-Work Activity Around Interactive Tabletops. In *Proceedings of the CROSS-MMLA workshop on Multimodal Learning Analytics across Spaces*, 1-6.

[68] Evelyn Palominos, Tracy Levett-Jones, Tamara Power and Roberto %J Nurse Education Today Martinez-Maldonado. 2019. Healthcare students' perceptions and experiences of making errors in simulation: An integrative review. *Nurse Education Today*. 77, Jun 2019. https://doi.org/10.1016/j.nedt.2019.02.013.

[69] Abelardo Pardo and George Siemens. 2014. Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*. 45, 3, 438-450. https://doi.org/10.1111/bjet.12152.

[70] Dilhan Perera, Judy Kay, Irena Koprinska, Kalina Yacef and Osmar R. Zaïane. 2008. Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*. 21, 6, 759-772. https://doi.org/10.1109/TKDE.2008.138.

[71] Luis Pablo Prieto, Kshitij Sharma, Łukasz Kidzinski, María Jesús Rodríguez-Triana and Pierre Dillenbourg. 2018. Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning*. 34, 2, 193-203. https://doi.org/10.1111/jcal.12232.

[72] Paul Prinsloo and Sharon Slade. 2015. Student privacy self-management: implications for learning analytics. In *Proceedings of the International Conference on Learning Analytics and Knowledge*, 83-92. https://doi.org/10.1145/2723576.2723585.

[73] Hemant Purohit, Samantha Dubrow and Brenda Bannan. 2019. Designing a Multimodal Analytics System to Improve Emergency Response Training. In *Learning and Collaboration Technologies. Designing Learning Experiences*, Zaphiris P. and Ioannou A. (Eds.). Springer, Cham, 89-100. https://doi.org/10.1007/978-3-030-21814-0_8.

[74] Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos and Sheelagh Carpendale. 2018. *Data-driven storytelling*. CRC Press, Boca Raton, FL, USA.

[75] Puripant Ruchikachorn. 2016. Integrating Visualizations into the Authoring of Narratives. Ph.D Dissertation Dissertation, State University of New York at Stony Brook, NY, USA.

[76] Lindy Ryan. 2016. *The Visual Imperative: Creating a Visual Culture of Data Discovery*. Morgan Kaufmann, Burlington, USA. https://doi.org/10.1016/C2015-0-00786-9.

[77] Bertrand Schneider and Roy Pea. 2014. Toward collaboration sensing. *International Journal of Computer-Supported Collaborative Learning*. 9, 4, 371-395. https://doi.org/10.1007/s11412-014-9202-y.

[78] Bertrand Schneider and Paulo Blikstein. 2015. Unraveling students' interaction around a tangible interface using multimodal learning analytics. *Journal of Educational Data Mining*. 7, 3, 89-116. Retrieved from https://jedm.educationaldatamining.org/index.php/JEDM/article/view/JEDM102.

[79] Jan Schneider, Daniele Di Mitri, Bibeg Limbu and Hendrik Drachsler. 2018. Multimodal learning hub: A tool for capturing customizable multimodal learning experiences. In *Proceedings of the European Conference on Technology Enhanced Learning, ECTEL'18*, 45-58. https://doi.org/10.1007/978-3-319-98572-5_4.

[80] Hans-Jörg Schulz, Marc Streit, Thorsten May and Christian Tominski. Year. Towards a characterization of guidance in visualization. In *Proceedings of the IEEE Conference on Information Visualization (InfoVis'13)*, 1-2.

[81] Beat A. Schwendimann, Maria J. Rodríguez-Triana, Andrii. Vozniuk, Luis P. Prieto, Mina S. Boroujeni, Adrian Holzer, Denis Gillet and Pierre Dillenbourg. 2017. Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies*. 10, 1, 30-41. http://doi.org/10.1109/TLT.2016.2599522

[82] E. Segel and J. Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*. 16, 6, 1139-1148. http://doi.org/10.1109/TVCG.2010.179.

[83] Shashi Kant Shankar, Luis P Prieto, María Jesús Rodríguez-Triana and Adolfo Ruiz-Calleja. 2018. A Review of Multimodal Learning Analytics Architectures. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies, ICALT'18*, 212-214. https://doi.org/10.1109/TLT.2016.2599522.

[84] Milo Shields. 2005. Information literacy, statistical literacy, data literacy. *International Association for Social Science Information Service and Technology quarterly*. 28, 2, 7-14. https://doi.org/10.29173/iq790.

[85] Sharon Slade and Paul Prinsloo. 2013. Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*. 57, 10, 1510-1529. https://doi.org/10.1177/0002764213479366.

[86] Robert A Sottilare, C Shawn Burke, Eduardo Salas, Anne M Sinatra, Joan H Johnston and Stephen B Gilbert. 2018. Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence in Education*. 28, 2, 225-264. https://doi.org/10.1007/s40593-017-0146-z.

[87] Arjun Srinivasan, Steven M Drucker, Alex Endert and John Stasko. 2018. Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *IEEE Transactions on Visualization and Computer Graphics*. 25, 1, 672-681. https://doi.org/10.1109/TVCG.2018.2865145.

[88] Manuel Stein, Thorsten Breitkreutz, Johannes Haussler, Daniel Seebacher, Christoph Niederberger, Tobias Schreck, Michael Grossniklaus, Daniel Keim and Halldor Janetzko. Year. Revealing the Invisible: Visual Analytics and Explanatory Storytelling for Advanced Team Sport Analysis. In *Proceedings of the International Symposium on Big Data Visual and Immersive Analytics (BDVA)*, 148-156. https://doi.org/10.1109/BDVA.2018.8534022.

[89] Marc Streit, Hans-Jorg Schulz, Alexander Lex, Dieter Schmalstieg and Heidrun Schumann. 2011. Model-driven design for the visual analysis of heterogeneous data. *IEEE Transactions on Visualization and Computer Graphics*. 18, 6, 998-1010. https://doi.org/10.1109/TVCG.2011.108.

[90] Danielle Albers Szafir. 2018. The good, the bad, and the biased: five ways visualizations can mislead (and how to fix them). *Interactions*. 25, 4, 26-33. https://doi.org/10.1145/3231772.

[91] Sara Taylor, Natasha Jaques, Weixuan Chen, Szymon Fedor, Akane Sano and Rosalind Picard. 2015. Automatic identification of artifacts in electrodermal activity data. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'15*, 1934-1937. https://doi.org/10.1109/EMBC.2015.7318762.

[92] Alice Thudt, Jagoda Walny, Theresia Gschwandtner, Jason Dykes and John Stasko. 2018. Exploration and Explanation in Data-Driven Storytelling. In *Data-Driven Storytelling*, Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos and Sheelagh Carpendale (Eds.). AK Peters/CRC Press, Natick, MA, USA, 59-83.

[93] Chao Tong, Richard Roberts, Rita Borgo, Sean Walton, Robert Laramee, Kodzo Wegba, Aidong Lu, Yun Wang, Huamin Qu and Qiong Luo. 2018. Storytelling and visualization: An extended survey. *Information*. 9, 3, 65. https://doi.org/10.3390/info9030065.

[94] Edward Rolf Tufte. 1983. *The visual display of quantitative information*. Graphics Press, Cheshire, CT, USA.

[95] Konstantinos Vassakis, Emmanuel Petrakis and Ioannis Kopanakis. 2018. Big data analytics: applications, prospects and challenges. In *Mobile Big Data*, Georgios Skourletopoulos, George Mastorakis, Constandinos X. Mavromoustakis, Ciprian Dobre and Evangelos Pallis (Eds.). Springer, Cham, 3-20. https://doi.org/10.1007/978-3-319-67925-9.

[96] Danding Wang, Qian Yang, Ashraf Abdul and Brian Y Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'19*, 601:601-615. https://doi.org/10.1145/3290605.3300831.

[97] Colin Ware. 2012. *Information Visualization: Perception for Design*. Elsevier, Waltham, MA, USA.

[98] Marcelo Worsley and Paulo Blikstein. 2018. A multimodal analysis of making. *International Journal of Artificial Intelligence in Education*. 28, 3, 385-419. https://doi.org/10.1007/s40593-017-0160-1.

[99] Joanna Young and Jan Wessnitzer. 2016. Descriptive statistics, graphs, and visualisation. In *Modern statistical methods for HCI*, Kaptein M Robertson J. (Ed.), Springer, Cham, 37-56. https://doi.org/10.1007/978-3-319-26633-6_3.

[100] Yangjinbo Zhang. 2018. Converging Data Storytelling and Visualisation. In *Proceedings of the International Conference on Entertainment Computing, ICEC 2018*, 310-316. https://doi.org/10.1007/978-3-319-99426-0_36.