The effects of visualisation literacy and data storytelling dashboards on teachers' cognitive load

Yuchen Liu
Faculty of Information Technology, Monash University, Australia

Stanislav Pozdniakov
Faculty of Information Technology, Monash University, Australia; School of Electrical Engineering and Computer Science, University of Queensland, Australia

Roberto Martinez-Maldonado
Faculty of Information Technology, Monash University, Australia

Learning analytics (LA) dashboards are becoming increasingly available in various learning settings. However, teachers may face challenges in understanding and interpreting the data visualisations presented on those dashboards. In response to this, some LA researchers are incorporating visual cueing techniques, like data storytelling (DS), into LA dashboard design to reduce the data visualisation skills – often referred to as visualisation literacy (VL) – and cognitive effort required by teachers to effectively use dashboards. However, despite the potential of DS principles in simplifying data visualisations, there is limited evidence supporting their effectiveness in actually reducing teachers' cognitive load. The study presented in this paper addresses this gap by investigating the potential impact of LA dashboards, with and without DS elements, on teachers with varying VL levels. Through a quasi-experimental study involving 23 teachers, we analysed changes in pupil dilation – a proxy for cognitive load – as they examined LA dashboards featuring student data captured while participating in synchronous, online collaborative learning tasks. Our findings suggest DS can reduce cognitive load, particularly for teachers with lower VL. These results provide insight into the effects of DS and VL on teachers' cognitive load, thereby informing the design of LA dashboards.

Implications for practice or policy:
- Developers of LA dashboards need to pay more attention to incorporating visual and narrative elements that are easily comprehensible and target-oriented, based on users' visualisation literacy levels.
- Educational providers and LA designers can recommend dashboards with DS elements to teachers with low VL to enhance their work efficiency.

Keywords: learning analytics (LA), visualisation literacy (VL), data storytelling (DS), cognitive load, eye tracking

Introduction

Learning analytics (LA) dashboards are becoming increasingly available in a wide variety of learning settings (Verbert et al., 2013). However, teachers may face challenges when attempting to comprehend and interpret the data representations, or data visualisations, presented on LA dashboards (Jivet et al., 2017). Moreover, the increasingly data-driven nature of instructional decision-making requires teachers to possess a certain level of proficiency in data analysis techniques. Yet, this proficiency might exceed the professional skill set of some educators. Furthermore, aside from the complexities of specific visualisations, the overall design of the dashboard interface, in terms of its complexity and intuitiveness, can also pose challenges to teachers, particularly those without a strong foundation in data analysis (Fernandez Nieto et al., 2022). Finally, it may be challenging for teachers to identify the types of educational questions they may want to answer through LA dashboard use (Li et al., 2021).

Indeed, to fully benefit from LA dashboards, teachers may need a certain level of visualisation literacy (VL) (Pozdniakov et al., 2023). However, even with a high level of VL, if the data visualisation elements in LA dashboards are excessively complex, they might challenge teachers' information processing abilities and increase their cognitive load (Sweller, 2011). In response, some LA researchers have started incorporating visual cueing techniques into LA dashboard design, aiming to decrease the data visualisation skills and cognitive efforts teachers need to interpret student data visualisations effectively. Data storytelling (DS) is one such technique aimed at simplifying data visualisation, making it more understandable to audiences who may not have a strong background in data analysis.

Although research has begun to explore teachers' VL and the methods for its assessment (Pozdniakov et al., 2023), the actual effects of LA dashboards and DS on teachers of varying VL levels remain unclear. Teachers might encounter cognitive overload both when processing complex data visualisations and when interpreting DS elements. Moreover, there is limited evidence as to whether DS can aid teachers in responding more effectively to certain types of educational questions. This study seeks to fill the aforementioned gaps by examining how cognitive load, indicated by pupil dilation, differs based on the type of dashboard (with or without DS elements) used by teachers and their respective VL when responding to specific educational questions using a set of LA dashboards. For in-depth exploration, we conducted a quasi-experimental study involving 23 teachers of diverse VL levels. These teachers were tasked with responding to various types of questions while examining LA dashboards, containing student data from synchronous, online collaborative learning discussions. We employed eye tracking to document changes in pupil dilation throughout this process. Our research design permitted us to directly compare and understand the differences in cognitive load among teachers with varying VL levels when utilising dashboards with or without DS elements. The findings of this study can provide insights into how DS and VL interact with teachers' cognitive load, offering practical implications for the design of more effective LA dashboards in educational contexts.

Background and related work

LA dashboards for teachers

LA dashboards have become important tools for teachers in various educational settings (Verbert et al., 2020). They can support teachers' tasks, such as monitoring student learning progress (Schwendimann et al., 2016), identifying potential learning barriers that students may encounter (Yan & Lin, 2021) and providing information that teachers can use to personalise feedback (Martinez-Maldonado et al., 2015). However, despite LA dashboards providing a wealth of data and information, parsing and effectively utilising these data for educational purposes remains a challenge (Jivet et al., 2017; Schwendimann et al., 2016). Teachers using LA dashboards are often tasked with comprehending and handling a large volume of data, while also considering how to incorporate these data into their teaching practice. Merely displaying data on a dashboard may not meet teachers' needs: it could potentially lead to a high cognitive load (Van Merriënboer & Sweller, 2010). Therefore, the design and optimisation of LA dashboards to better serve teachers is a significant issue currently faced in educational technology research (Verbert et al., 2020). The key challenge lies not only in providing more data but also in displaying data in a way that allows teachers to quickly and effortlessly understand and utilise the information for pedagogical purposes (Verbert et al., 2013).

Moreover, identifying the types of questions teachers may answer based on the data displayed in a dashboard is essential for the effective use of a given LA dashboard (Li et al., 2021; Verbert et al., 2013). For instance, Li et al. highlighted that various categories of questions from teachers can be addressed by different dashboard designs. They also proposed that some LA dashboards offer more suitable affordances for addressing specific educational questions. Pozdniakov et al. (2022) advocated for a dashboard design that explicitly includes the analytic questions teachers can answer based on certain dashboard visualisations, using these questions as headings for those visualisations. Van Leeuwen and Rummel (2020) also compared the different interaction patterns that teachers adopt while using dashboards designed for either simply mirroring information or for providing more explicit advice to

teachers. Despite these explorations, there is a noticeable gap in empirical research demonstrating how dashboard design can influence teachers' interpretation process in order for them to address different types of questions about students' progress. These factors are critically important for the design of effective LA tools.

DS in LA

To address some of the current challenges that teachers commonly face while interacting with LA dashboards (mentioned above), some LA researchers are incorporating visual cueing techniques, like DS, into LA dashboard design. DS integrates charts, text and other resources to emphasise important data points and evidence in a dashboard interface, aiming to assist the audience in focusing on responses to specific questions (Echeverria et al., 2018; Pozdniakov et al., 2023). For instance, Martinez-Maldonado et al. (2020) suggested a layered dashboard design and a progressive visual cueing approach to aid teachers in gradually understanding multimodal data of student teams. Echeverria et al. created a teaching-facing dashboard that displayed textual annotations to drive the attention of teachers towards relevant data points that were critical according to the learning design.

However, it is crucial to note that if a LA dashboard, embedded with DS elements, is overly complex, it might counterproductively increase teachers' cognitive load rather than aid in data comprehension. The potential impact of this design complexity suggests that teachers' data VL skills can significantly influence their interaction with LA dashboards containing DS elements (Pozdniakov et al., 2023). Therefore, teachers' level of VL emerges as a key factor worth considering in this context.

Teachers' VL

In recent years, the notion of VL has been increasingly recognised by researchers in the field of information visualisation (Donohoe & Costello, 2020). VL refers to the ability and skill to read and interpret visualised data and extract information from data visualisations (Lee et al., 2016). The significance of this ability is also relevant to LA. For example, many teachers still struggle with understanding and interpreting dashboards partly because they lack relevant data skills (Ndukwe & Daniel, 2020). An insufficient VL among teachers can lead to a limited understanding of various chart types and visualisation elements, thereby impacting their ability to accurately assess student learning (Donohoe & Costello, 2020).

Therefore, understanding the potential role that teachers' VL may play in the LA dashboard interpretation process is crucial. Recent studies have suggested tools, like the Visualisation Literacy Assessment Tool (VLAT) test, that can be effectively used to measure teachers' VL levels (Lee et al., 2016). This test can aid in determining teachers' capabilities in understanding and interpreting data visualisations (Pozdniakov et al., 2023). Moreover, Bloom's taxonomy (Bloom et al., 1984), which has been extensively used in education to categorise learning objectives into varying levels of complexity, has also been proposed as a way to assess the kind of tasks that a data visualisation can enable (Arneson & Offerdahl, 2018) and also to assess and support the development of visualisation skills in educational contexts.

Although some research has delved into teachers' VL and methods to assess it (Pozdniakov et al., 2023), and the potential influence of teachers' characteristics on the ways they interact with dashboards (van Leeuwen et al., 2021), there remains a dearth of studies examining the impact of LA dashboard design on teachers' cognitive load in relation to teachers' VL skills and the kind of educational questions they can ask of the analytics.

Measurement of teachers' cognitive load

Cognitive load theory is an instructional theory based on our knowledge of human cognition (Sweller, 2022). The basic idea of cognitive load theory is that cognitive capacity in working memory is limited, so that if a learning task requires too much capacity, learning will be hampered (De Jong, 2010). Working memory is the small amount of information that can be held in mind and used in the execution of cognitive tasks (Cowan, 2014). To address the challenges arising from the limitations of cognitive capacity, designing

instructional systems that optimise the use of working memory capacity and avoid cognitive overload is recommended (De Jong, 2010). Moreover, accurate measurement of cognitive load becomes crucial to ensure that strategies to communicate information are effective and to prevent the excessive consumption of working memory capacity.

Sweller et al. (2019) evaluated the progress made since the introduction of cognitive load theory, highlighting three major developments in the measurement of cognitive load. Two of these developments are particularly concerned with the evolution and refinement of subjective measurement methods. Indeed, even to this day, a considerable number of studies continue to rely on subjective report assessments for evaluating the cognitive load of teachers.

However, recently in the field of LA, there has been an increasing focus on the third development mentioned by Sweller et al. (2019): objective measurement, particularly using eye tracking technology. For instance, the study by Wang et al. (2021) explored the measurement of cognitive load in collaborative and intelligent learning environments based on eye tracking data. Nonetheless, these studies primarily concentrate on the cognitive aspects of learners, rather than specifically investigating the cognitive load of teachers. The extent to which the cognitive load of teachers is influenced by the design aspects of LA dashboards remains underexplored.

Aim and research questions

In response to the aforementioned research gap, we conducted a study aimed at generating a deeper understanding of the impact of DS in LA dashboard design on teachers' cognitive load, considering teachers' VL skills (measured using the VLAT test) and the kind of specific educational questions, identified by their Bloom's taxonomy level (Arneson & Offerdahl, 2018). To achieve this aim, we addressed the following research questions (RQs):

- RQ1: To what extent does the LA dashboard design (with and without DS elements) impact the cognitive load experienced by teachers, considering their level of VL?
- RQ2: To what extent does responding to different types of educational questions when inspecting different LA dashboard designs affect the cognitive load of teachers?
- RQ3: To what extent does responding to different types of educational questions affect teachers' cognitive load in relation to teachers' level of VL?

Methods

Participants

A total of 23 teachers were recruited for this study (13 of them identified themselves as females and 10 as males). The age range of the participants was between 22 and 60 years (M = 31, SD = 8). Except for one, all the teachers had attained a postgraduate degree: seven held a master's degree, 12 were enrolled in a doctoral programme and two had already earned a doctoral degree. Each participant had a degree in a field related to science, technology, engineering or mathematics and was either engaged in a teaching role at a university or had an average of 3 years of teaching experience (SD = 3.4).

Design

Context
The study was set in the context of an authentic online learning setting. A coordinator of a unit of study in an information technology programme utilised a publicly available analytics tool known as ZoomSense (Bartindale et al., 2020). ZoomSense enhances the cloud-based video conferencing service Zoom by adding virtual agents within each Zoom breakout room to autonomously record student participation continuously. The virtual agents were configured based on the teachers' predefined settings, enabling them to automatically generate corresponding Google Docs files for each breakout room and share the

links with the students. The Google Docs are pre-configured by ZoomSense to provide students with a specific text box where students are expected to write for each particular task set by their teacher. This system set-up enables the automatic logging of whether students have started to work on or completed a specific task. The data captured by ZoomSense during a class facilitated via breakout rooms includes the number of times and amount of time each student within a breakout room speaks, the progress in the completion of tasks by each group of students within a breakout room in the Google Docs and the number of times and duration a teacher visits a breakout room.

ZoomSense comes with a default dashboard that does not include any DS elements (as shown in Figure 1). Two additional dashboard prototypes, which included DS elements, were created as a part of a co-design process described by Pozdniakov et al. (2022). Based on information visualisation principles (Bach et al., 2022), in all three dashboards, a timeline chart is included at the top as a frame of reference indicating the activity in which the class is currently located (see Timeline in Figures 1, 2 and 3, which contains three activities along with the start and final times of the class. In these figures, the current activity selected by the teacher is highlighted in navy blue over grey, as shown with Activity 3). At the beginning of each class, the teacher can set the number of activities that will be displayed on the timeline. The rest of this subsection provides details about the three dashboard designs.

Dashboard without DS elements
This dashboard contains four different types of visualisations:

- Pie charts illustrate the proportion of time each student has contributed to the group discussion (see V1 in Figure 1).
- Sociograms charts, represented as one chart per breakout room, visually depict the communication dynamics between participants. Each node in the chart represents a student or a teacher, while the thickness of the links connecting two nodes represents the duration of communication between those participants consecutively (see V2 in Figure 1).
- Timeline sequence diagrams visualise the interaction patterns among students within a group by depicting the constellation of utterances over time, enabling teachers to observe the dynamic nature of turn-taking behaviour (see V3 in Figure 1).
- Google Docs progress is shown using a horizontal line chart (see V4 in Figure 1).

Teachers can navigate from one visualisation to another visualisation to answer their questions by clicking on the navigation buttons at the top of the interface (see Navigation in Figure 1).
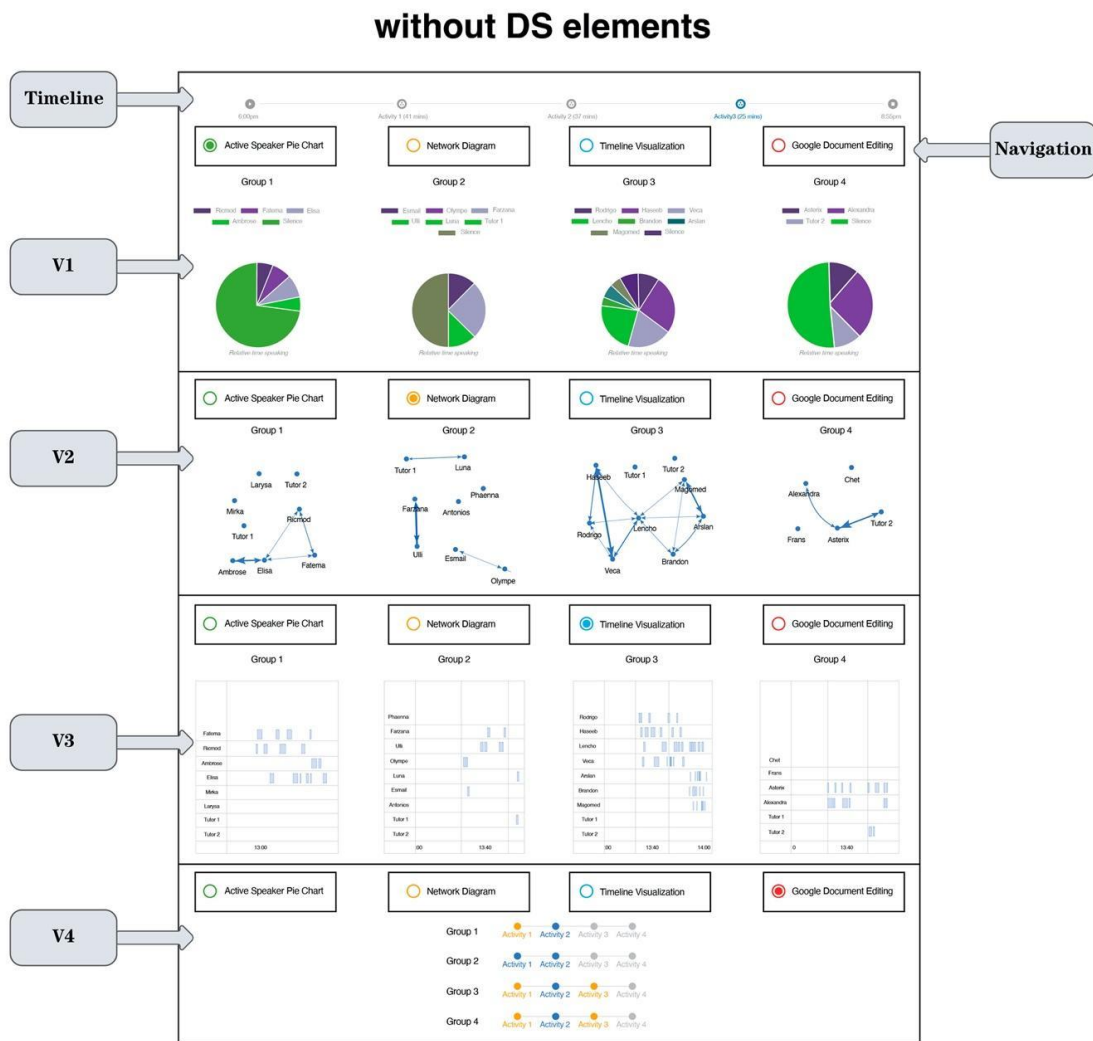
Figure 1. Dashboard without DS elements used in the study

Dashboards with DS elements

These two dashboards with DS elements were developed as part of the iterative design process described by Pozdniakov et al. (2022), with the second version representing a more refined iteration of the first. Both dashboards share the following DS elements:

- The question that the teacher needs to answer serves as the title of each corresponding visualisation chart (see DS1 in Figure 2).
- Semantic colour coding is implemented in the dashboard design to ensure colour emphasis is accessible for individuals with colour blindness. In this approach, three contrasting colours are utilised in a colour-blind friendly manner. The majority of visual elements are displayed in grey and navy, while orange and red are reserved for highlighting situations that require closer attention (see DS2 in Figure 3): for example, to highlight the groups that have discussed the most in V6 or the rooms where the teacher has spent the most time in V7.
- V6 is present in both versions, depicted as a horizontal bar graph which visualises the detected speech volume in each group. When the speech volume of a group surpasses the overall average, its associated bar is automatically accentuated in orange.
- V7 is a component of both versions and employs a bar chart to show the cumulative duration (expressed in minutes) that the teacher spends in each breakout room. The bar representing the room in which the teacher has spent the most time is distinctly highlighted in orange.

The differences between both dashboards are the following:

- V5 present text narratives explaining the student's behaviour in version 1. This uses Zoom speech data to categorise students as "active" (when the student is engaged in speaking), "inactive" (when not engaged in speaking), and "dominant" or "quiet" (if one individual is predominantly speaking or, conversely, is notably silent compared to others).
- In the second version, V5 disappears and, as requested by teachers (Pozdniakov et al., 2022), the narratives are embedded into V2 (see DS3 in Figure 3).
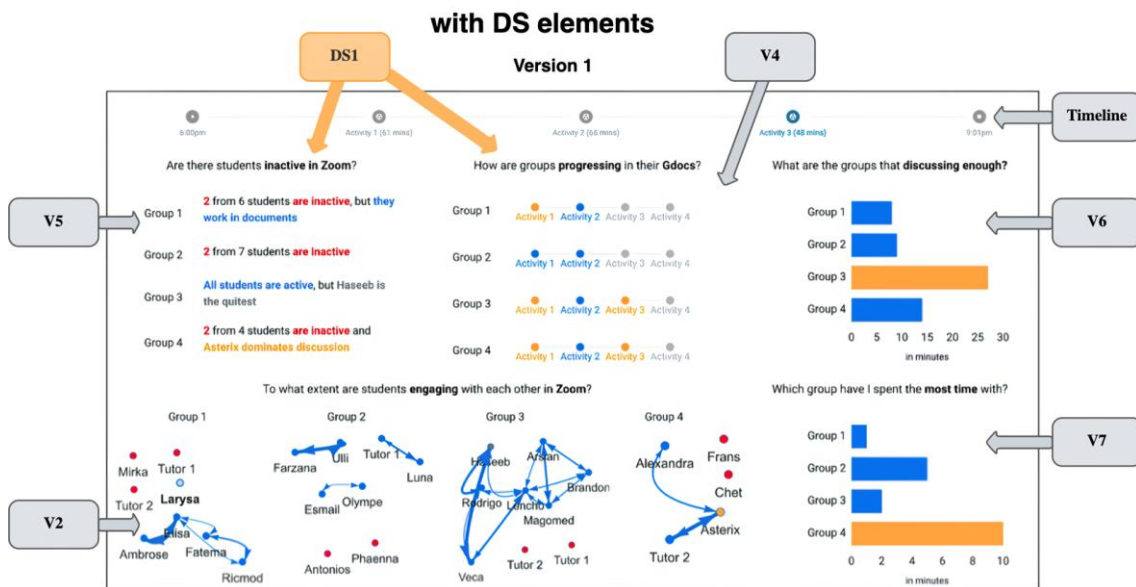

Figure 2. Dashboard with DS elements version 1 used in the study. DS elements are brought into realisation through DS1 questions presented as chart titles.
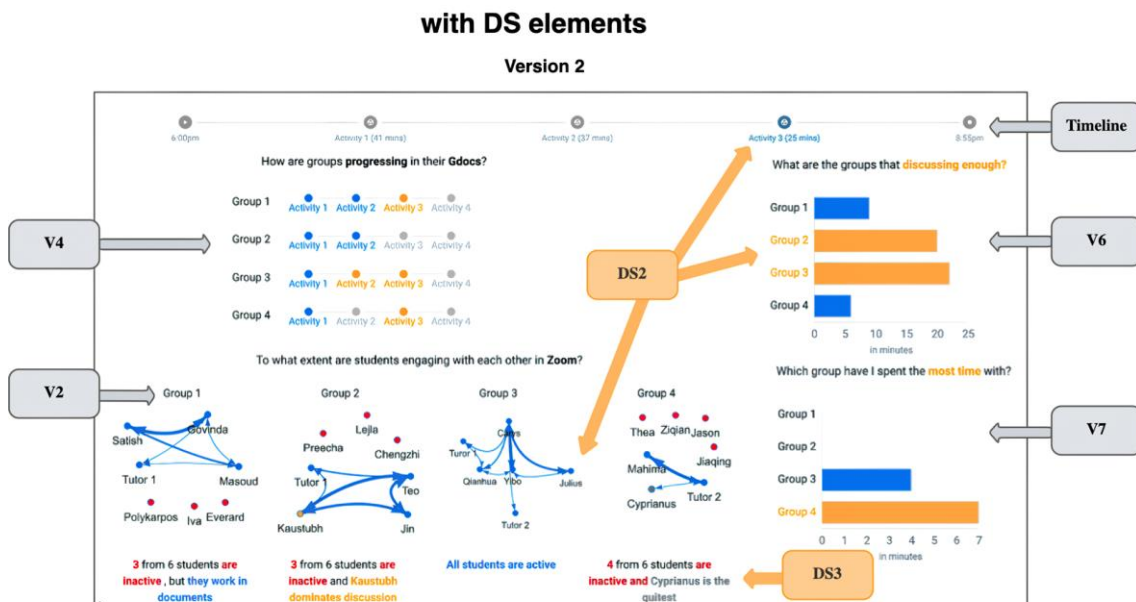

Figure 3. Dashboard with DS elements version 2 used in the study. DS elements are brought into realisation through DS2 emphasised elements, like highlighting bars in V6 and V7 charts, and nodes in Timeline, V2 and V4 charts; and DS3 textual summaries.

Procedure

The study consisted of two parts: an eye tracking study and VL testing.

Eye tracking study
This part of the study involved a session lasting approximately 40 minutes for each of the 23 participating teachers. We placed a Tobii Pro Nano eye tracker at the bottom of a 24-inch desktop monitor, which had a resolution of 1920 x 1200 and a sampling rate of 60 Hz. To ensure a consistent environment for all participants, we maintained the same laboratory conditions throughout the study: the screen luminance remained constant, the laboratory had moderate and consistent illumination, and we arranged an external video camera to record the audio data. Additionally, we provided a stationary chair for participants to ensure their position remained stable.

At the beginning of the study, all participants received a brief onboarding session to help them understand the meaning of each element, colour, and text displayed on each of the dashboards. Subsequently, participants completed a short (5 minutes) training activity to simulate the tasks they would be performing as a part of the study. As part of the task execution, we showed participants the three dashboard versions and asked them to assess students' engagement. To support this process and guide teachers' interpretation of the visualisations, we provided six educational questions (see details below). The study followed a within-subject design, which meant that the same set of questions was asked of each participant for each version of the dashboard. To minimise any potential order effect, the order in which each dashboard version was presented was randomised. Specifically, we used a three-by-three Latin square matrix design to pseudo-randomly assign the order of the versions presented to the participants.

The same six educational questions were posed to teachers while inspecting each dashboard. In this experiment, we particularly focused on the first two Bloom's levels (Arneson & Offerdahl, 2018), namely 1 – knowledge and 2 – comprehension, and constructed educational questions in the experiment based on these two levels to guide teachers in using LA dashboards for specific teaching purposes. In the context of data visualisation (Arneson & Offerdahl, 2018), questions at the knowledge level focus mainly on retrieving specific facts and data points. The three knowledge questions that tasked teachers to identify explicit facts and data on the dashboard were the following: "Which group or groups have the most inactive students?", "In which group or groups are students participating in the discussion equally?" and "With which group have you (imagine that you are 'Tutor 2') spent the most time?" By contrast, questions at the comprehension level require participants to interpret and understand information on the dashboard, involving higher-order cognitive processes such as those that require making comparisons, summarising the information being presented or making predictions (Arneson & Offerdahl, 2018). This level consisted of the following three questions: "Which group is the most engaged in writing activity?", "Which group is the most engaged in discussion?" and "In your opinion, which group or groups might have experienced issues with progressing in the learning activity?"

VL testing
After the eye tracking, each teacher was asked to fully complete the VLAT test (Lee et al., 2016). The validated VLAT test, consisting of 53 multiple-choice questions, required participants to interpret various data charts, including bar and line charts, to assess their general VL. To standardise the test, we followed the recommendations by Lee et al. (2016). Each question was timed with a 40-second limit, and participants had the option to skip any question if they were unsure of the correct response. Given that exposure to multiple charts via the VLAT test can potentially activate or improve VL skills, and following the suggestion by Lee et al., the test was deliberately scheduled 1–3 days post-task completion to minimise potential bias to the greatest extent possible.

Analysis

Pupil dilation measures

In our study, we used pupil dilation as a proxy for participants' cognitive load when using various types of dashboards. A subtractive baseline correction for pupil diameter was applied as suggested by Krejtz et al. (2018). The baseline was set during the reading-learning activity before the dashboard was presented to the participants, a period that could be considered a preparation period before the participants had started the main task. We calculated the intra-trial change in pupil dilation, which represents the change in pupil diameter during the participant's use of the dashboard, relative to the baseline phase. To reflect the cognitive load experienced by participants when using the dashboard, we set such a baseline phase before each use of the dashboard. This measure was calculated in a similar way to that reported in Krejtz et al. We first calculated the running mean of the corresponding baseline phase and then used the difference between pupil diameter and the mean of the baseline phase as this measure.

VL measures

After the completion of the VL test by the participants, individual scores were calculated for each participant. Questions that were skipped or not answered within the time limit were marked incorrect. The median score of VL from all participants (Mdn = 39) was used to categorise the participants into low or high VL levels. The low VL level comprised 12 participants (Mdn = 35, min = 17, max = 39) and the high VL level had 11 participants (Mdn = 45, min = 40, max = 48).

Given the observed range of scores in the low VL group, particularly the presence of an extreme value (score of 17), further analysis was undertaken to assess the impact on the group categorisation. This involved reassessing the grouping after the exclusion of the extreme value and also exploring the feasibility of a three-level categorisation using quartile-based thresholds. However, the limited sample size led to impractically small subgroups in a three-level categorisation. Consequently, we determined that maintaining the original two-level categorisation was the appropriate approach, even when including the outlier in the analysis. In this decision-making process, we re-analysed all research questions, particularly considering the impact of excluding the extreme value on the results. We found that even in the absence of the extreme value, the main findings of the study remained consistent. Therefore, the decision to retain the outlier score of 17 was not only to acknowledge its significance within the data set but also to preserve the authentic representation of the characteristic profile of the low VL group.

Modelling

To address the research questions, we selected the intra-trial change in pupil dilation as the dependent variable, while the dashboard type, types of educational questions classified according to Bloom's taxonomy and VL levels were used as independent variables. As the data for this research was continuously collected over time, we noted a potential significant autocorrelation within the data, suggesting an inherent correlation among observations in the time series. To this end, we utilised generalised additive mixed models (GAMMs) for our analysis. GAMMs not only allow for more flexible modelling of non-linear relationships (van Rij et al., 2019) but also effectively handle autocorrelation in the data and consider both fixed and random effects. We transformed the educational question type and VL level into dummy variables and included them along with the dashboard type as fixed effects in the model. Additionally, we introduced random intercepts for each participant to account for individual differences.

The model we used is described by the following equation:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 (x_{1ij} \cdot x_{2ij}) + u_i + \rho e_{ij-1} + e_{ij}$$

where y_{ij} represents the dependent variable for the j-th observation within the i-th participant, x_{1ij} and x_{2ij} are the independent variables, u_i is the random intercept for each participant, and $\rho e_{ij-1} + e_{ij}$ denotes the first-order autocorrelation of the residuals within each participant.

We conducted a Durbin-Watson test to see whether there is a need to account for an autocorrelation. After that, we fitted the GAMM model using the mgcv package in R. In order to decide on the final model's

structure, we used the Akaike information criterion (AIC) and likelihood ratio test. Furthermore, in this project, post hoc analyses involved further exploring the interactions among dashboard type, question type and VL level. We used the emmeans R package to compute marginal means of these interaction effects. To account for family-wise errors and increase the reliability of our findings, the Benjamini-Hochberg false discovery rate procedure was applied to adjust statistical significance for pairwise marginal comparisons. Therefore, the *p* values are reported after such adjustments were computed. We use Cohen's *d* when reporting effect sizes.

Results

The effect of DS and VL on teachers' cognitive load (RQ1)

As explained above, to investigate the effects of DS and VL on teachers' cognitive load, we used GAMM. The Durbin-Watson test indicated a substantial autocorrelation (DW = 0.490, $p < 0.001$). This motivated the inclusion of an autocorrelation structure to be a part of the model. Secondly, we checked whether the inclusion of an interaction effect between DS and VL leads to a better model fit. Based on the AIC results, the model incorporating the interaction term showed a better fit (AIC = −1011.046) compared with the model without the interaction term (AIC = −1007.029). This was further confirmed by a likelihood ratio test ($\chi^2(1) = 8.017$, $p < 0.05$).

The GAMM findings (see Table 1), with the interaction term included, revealed significant impacts of the dashboard type on cognitive load. Specifically, compared to the dashboard without DS elements, both dashboards with DS elements, versions 1 and 2, showed a significant impact. Additionally, there was a significant interaction effect between the dashboard type and VL on cognitive load. In contrast, the VL's main effect on cognitive load was not significant.

Table 1
Summary of the results of the GAMM for the effect of dashboard and VL

Coefficients	Estimate	SE	*t* value	*p* value	95% CI
(Intercept)	0.119	0.029	4.131	< 0.001***	[0.063, 0.176]
With DS elements version 1	-0.078	0.017	-4.498	< 0.001***	[-0.112, -0.044]
With DS elements version 2:	-0.045	0.021	-2.162	0.031*	[-0.085, -0.004]
VL	0.002	0.042	0.051	0.960	[-0.080, 0.084]
With DS elements version 1: VL	0.070	0.025	2.789	0.006**	[0.021, 0.119]
With DS elements version 2: VL	0.060	0.030	2.023	0.044*	[0.002, 0.119]

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Difference in cognitive load between without DS and with DS conditions for low VL and high VL teachers

In the post-hoc analyses, we assessed the significance of differences across groups with paired comparisons. Figure 4 provides a visual representation of the estimated marginal means of cognitive load and their 95% confidence intervals (CI) for different groups.

For teachers with low VL, the dashboard with DS elements (version 1) significantly reduced cognitive load when compared to a dashboard without DS elements. This significant reduction is supported by a large effect size (*d* = 0.715). The estimated value and adjusted *p* value (*estimate* = 0.041, adjusted *p* < 0.01) further confirm this statistical significance, indicating that DS elements play an important role in reducing cognitive load for teachers with lower VL.

However, we cannot determine whether the dashboard with DS elements (version 2) reduces cognitive load. The smaller effect size (*d* = 0.349) suggests a lower level of impact, and although the estimated value

was 0.075, the adjusted *p* value (adjusted *p* > 0.05) suggests that the impact of version 2 on cognitive load is not statistically significant.

Among teachers with high VL, the impact of DS elements on cognitive load was not prominent. Specifically, neither version 1 (*d* = 0.067, *estimate* = 0.113) nor version 2 (*d* = −0.135, *estimate* = 0.137) demonstrated a significant effect of DS elements in reducing cognitive load, as evidenced by the adjusted *p* values being greater than 0.05. Additionally, the sizes of the *d* values and estimates indicate that, even when effects are present, they remain relatively small. Therefore, these results suggest that, in this group, the effectiveness of DS elements in reducing cognitive load was not substantial.
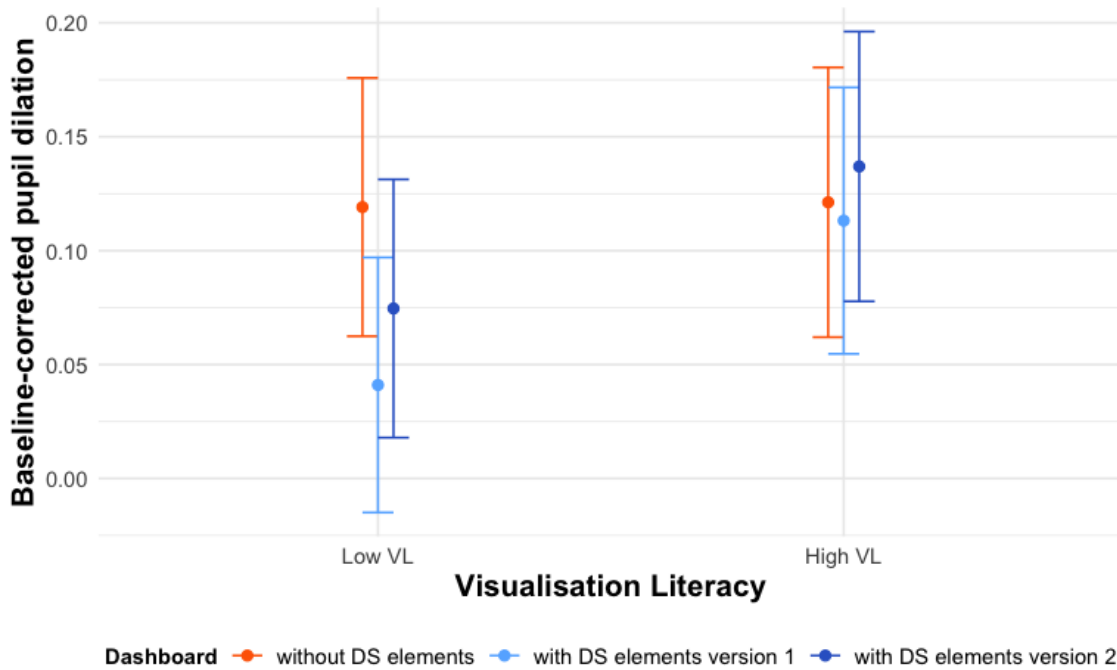

Figure 4. Estimated marginal means of cognitive load and their 95% CI for different groups. Post-hoc analysis revealed one significant pairwise contrast: For teachers with low VL, the dashboard with DS elements (version 1) significantly reduced cognitive load (*d* = 0.715, *estimate* = 0.041, adjusted *p* < 0.01). Error bars show 95% CI.

The effect of DS on teachers' cognitive load when answering different types of educational questions (RQ2)

Based on the above analysis, we considered whether teachers' cognitive load was affected when answering different types of educational questions using the different types of dashboards. Therefore, following the same procedure as before, the Durbin-Watson test results (DW = 0.485, *p* < 0.001) indicated that there was autocorrelation in the data. GAMM was employed, and models with and without interaction between the dashboard type and educational questions type were fitted. The AIC results showed that the model without the interaction term demonstrated a better fit (AIC = −1005.918) compared to the model with the interaction term (AIC = −1003.015). This was further confirmed by a likelihood ratio test, which failed to provide enough evidence to favor the model with the interaction term ($\chi^2(1)$ = 1.097, *p* > 0.05).

Addressing RQ2, according to the results of the model without interaction (see Table 2), comparing to the dashboard without DS elements, the dashboard with DS elements version 1 had a significant impact on cognitive load. The impact of the dashboard with DS elements version 2 on cognitive load was not significant. Additionally, within the context of different dashboard types, the impact of the type of educational questions that teachers were tasked to address on cognitive load was not significant.

Table 2
Summary of the results of GAMM for dashboard and educational questions types

Coefficients	Estimate	SE	t value	p value	95% CI
(Intercept)	0.122	0.022	5.617	< 0.001***	[0.079, 0.164]
With DS elements version 1	-0.045	0.013	-3.557	< 0.001***	[-0.070, -0.020]
With DS elements version 2	-0.017	0.015	-1.115	0.266	[-0.047, 0.013]
Educational questions	-0.002	0.005	-0.380	0.704	[-0.012, 0.008]

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

The effect of different levels of VL on teachers' cognitive load when answering different types of educational questions (RQ3)

Similar to previous analyses, here we mainly investigated the effect of different levels of VL on teachers' cognitive load when answering different types of educational questions. The DW test result (DW = 0.485, $p < 0.001$) still indicated the presence of autocorrelation in the data. Therefore, we continued to use GAMM for analysis. Similarly, we fitted two models: one with an interaction term and one without. Based on the AIC results, it was apparent that the model without the interaction term delivered a more optimal fit (AIC = −995.031) relative to the model that included the interaction term (AIC = −993.032). A subsequent likelihood ratio test corroborated this observation, indicating insufficient evidence to prefer the model with the interaction term ($\chi^2(1) = 0.007$, $p > 0.05$).

In addressing RQ3, according to the GAMM results (see Table 3), VL did not significantly impact cognitive load. Additionally, in the context of VL, educational question type also had no significant impact on cognitive load.

Table 3
Summary of the results of GAMM for VL and educational questions types

Coefficients	Estimate	SE	t value	p value	95% CI
(Intercept)	0.082	0.027	3.057	0.002**	[0.029, 0.134]
VL	0.043	0.038	1.132	0.258	[-0.032, 0.119]
Educational questions	-0.001	0.005	-0.163	0.871	[-0.011, 0.009]

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Discussion

Summary of results

Regarding RQ1, we found that there were some differences in cognitive load between teachers with different VL levels when using dashboards with or without DS elements. Specifically, teachers with low VL had significantly lower cognitive load when using dashboards with DS elements. In particular, the impact of the dashboard with DS elements version 1 in reducing cognitive load was significant, supporting the notion that the ease of comprehension of DS (Echeverria et al., 2018) can effectively reduce teachers' cognitive load when working with complex data. However, the impact of dashboard with DS elements version 2 was not statistically significant, which suggests that further research may be required to clarify its impact. The differences in the impacts of dashboard designs, versions 1 and 2, could be attributed to the subtle variations in the design elements of the visual interface. The key difference is that in version 2 of the dashboard, the network visualisations are complemented by explanatory text, making the resulting visualisation denser. In contrast, version 1 segregated the same information into two separate interface components: the explanatory text and the network visualisations, each with a unique header. This design divergence served to simplify the visual interface by using two distinct visual components (in version 1) instead of one (in version 2) to present the same information, thereby aiding teachers with low VL to comprehend the data. This aligns with dashboard design patterns that advocate not posing too many

questions for a single graph (Bach et al., 2022). On the other hand, for teachers with high VL, DS had a lesser impact on their cognitive load when using different dashboards. This could be because they were already proficient in processing complex information. This also validated Pozdniakov et al.'s (2023) study, which suggested that teachers with high VL can fully leverage LA dashboards.

Regarding RQ2 and RQ3, the LA literature has highlighted the importance of the types of questions that teachers may answer based on the data displayed in a dashboard and how this ultimately shapes the kinds of interactions that may occur (Li et al., 2021; Verbert et al., 2013). Yet, in our study, we did not find any significant impact of the different types of educational questions teachers were tasked to address on their cognitive load in relation to their VL level. This may suggest that teachers can flexibly adjust their cognitive strategies based on different educational questions, and this adaptability does not heavily rely on their level of VL. However, as indicated in our RQ2 results, the dashboard with DS elements version 1 significantly affects teachers' cognitive load when answering any type of educational questions. This highlights the importance of selecting an appropriate dashboard to manage cognitive load during teaching tasks. For example, the dashboard design needs to be simplified if teachers need to do other tasks concurrently but the complexity of the dashboard could be higher if the teachers' task is focused on the exploration of student data or for reporting purposes.

Implications for research and practice

Our findings contribute with new insights to the body of LA literature focused on understanding how to design teacher-facing dashboards and data-intensive user interfaces that teachers can actually use (Jivet et al., 2017; Verbert et al., 2013). A key focus of our work is to provide evidence on the impact of the dashboard design on teachers' cognitive load, an aspect that has just started to be explored in recent literature, as shown by Pozdniakov et al. (2023). Additionally, we emphasise the potential role of DS as a support tool for teachers who may not have strong data visualisation skills (Echeverria et al., 2018; Fernandez Nieto et al., 2022).

Firstly, our findings indicate that dashboards with DS elements can be beneficial for teachers with low VL by potentially reducing their cognitive load. This can offer a new direction for dashboard design, where researchers might need to pay more attention to incorporating elements that are easily comprehensible and target-oriented, based on users' VL levels. Simultaneously, educational providers and LA designers can leverage this finding to recommend dashboards with DS elements to teachers with low VL to enhance their work efficiency. However, we observed differences in the impact on cognitive load among different dashboards with DS elements, possibly attributed to variations in the DS elements incorporated within the dashboards. Therefore, further investigation into how these distinct DS elements specifically affect teachers' cognitive load becomes particularly crucial. Nonetheless, the application of DS principles into dashboard and LA interface design also opens up a new research direction that can focus on understanding the potential unforeseen ethical concerns of incorporating visual cueing and visual guidance to direct teachers' attention to certain data points (Echeverria et al., 2018). For instance, an overemphasis on certain data points could potentially skew teachers' perceptions and interpretations of student performance, thereby influencing their educational strategies and decision-making processes.

Limitations and future work

This study has some limitations. Firstly, the dashboard designs employed in this study were pre-determined by the ongoing learning analytics project our research builds on, which stopped us from testing alternative dashboard designs and might have restricted our understanding of how diversity in dashboard designs could influence teachers' cognitive load. To further explore how various dashboard designs impact teachers' cognitive load, future studies could consider incorporating a wider variety of dashboard designs. Secondly, although we divided the teachers into two levels based on their VL in this study, VL might indeed be more complex, with subtle differences between various levels. Therefore, future studies could delve into a more nuanced classification of VL levels to examine its effect on teachers' cognitive load more accurately. After reassessing the categorisation approach, including considering a three-level system, we encountered challenges with small sample sizes. This led us to maintain the two-

level categorisation to preserve statistical robustness. Future research could explore a more detailed classification of VL with larger sample sizes for a deeper understanding of its impact. Thirdly, the types of educational questions that the teachers had to answer in our study were predetermined. However, in actual teaching scenarios, the types of educational questions that teachers might have to deal with could be more varied and complex. As such, future research could attempt to incorporate a broader and more realistic range of educational question types to better simulate actual teaching scenarios, providing more relevant insights for real-world teaching. Finally, as already mentioned when reporting our results, the sample size was relatively small. A large sample of teachers could have provided more certainty regarding some of the results that were not statistically significant.

Conclusion

In this study, we explored the effects of VL and DS on teachers' cognitive load. We found that for teachers with low VL, dashboards with DS elements could effectively reduce their cognitive load. However, for teachers with high VL, the type of dashboard and types of analytic questions have a smaller impact on their cognitive load. Based on these findings, we suggest that the design of teaching dashboards should fully consider the teacher's level of VL. For those teachers with low VL who deal with a large amount of data tasks, DS may better help them reduce cognitive load, thus more effectively processing data information.

Author contributions

Author 1: Conceptualisation, Investigation, Formal analysis, Writing – original draft, Writing – review and editing; **Author 2:** Conceptualisation, Data curation, Investigation, Writing – review and editing; **Author 3:** Conceptualisation, Supervision, Writing – review and editing

Acknowledegements

References

Arneson, J. B., & Offerdahl, E. G. (2018). Visual literacy in Bloom: Using Bloom's taxonomy to support visual learning skills*. CBE—Life Sciences Education*, *17*(1). https://doi.org/10.1187/cbe.17-08-0178

Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y., & Chen, M. (2022). Dashboard design patterns*. IEEE Transactions on Visualization and Computer Graphics*, *29*(1), 342–352. https://doi.org/10.1109/TVCG.2022.3209448

Bartindale, T., Olivier, P., Chen, P., Richardson, D., Mackie, T., & Pozdniakov, S. (2020). *ZoomSense* (Version 1.0.3) [Computer software]. Monash University.

Bloom, B. S., Krathwohl, D. R., & Masia, B. B. (1984). *Taxonomy of educational objectives: The classification of educational goals*. Longman.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology FReview*, *26*, 197–223. https://doi.org/10.1007/s10648-013-9246-y

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, *38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

Donohoe, D., & Costello, E. (2020). Data visualisation literacy in higher education: An exploratory study of understanding of a learning dashboard tool. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(17), 115–126. https://www.learntechlib.org/p/218005/

Echeverria, V., Martinez-Maldonado, R., Buckingham Shum, S., Chiluiza, K., Granda, R., & Conati, C. (2018). Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling. *Journal of Learning Analytics*, *5*(3), 73–97. https://doi.org/10.18608/jla.2018.53.6

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., & Buckingham Shum, S. (2018). Driving data storytelling from learning design. In *Towards user-centred learning analytics—Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 131–140). Association for Computing Machinery. https://doi.org/10.1145/3170358.3170380

Fernandez Nieto, G. M., Kitto, K., Buckingham Shum, S., & Martínez-Maldonado, R. (2022). Beyond the learning analytics dashboard: Alternative ways to communicate student data insights combining visualisation, narrative and storytelling. In *Learning analytics for transition, disruption and social change—Proceedings of the 12th International Learning Analytics and Knowledge Conference* (pp. 219–229). Association for Computing Machinery. https://doi.org/10.1145/3506860.3506895

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Lecture notes in computer science: Vol. 10474. Data driven approaches in digital education* (pp. 82–96). Springer. https://doi.org/10.1007/978-3-319-66610-5_7

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE, 13*(9), Article e0203629. https://doi.org/10.1371/journal.pone.0203629

Lee, S., Kim, S.-H., & Kwon, B. C. (2016). VLAT: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics, 23*(1), 551–560. https://doi.org/10.1109/TVCG.2016.2598920

Li, Q., Jung, Y., & Friend Wise, A. (2021). Beyond first encounters with analytics: Questions, techniques and challenges in instructors' sensemaking. In *The impact we make: The contributions of learning analytics to learning—Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (pp. 344–353). Association for Computing Machinery. https://doi.org/10.1145/3448139.3448172

Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., & Buckingham Shum, S. (2020). From data to insights: A layered storytelling approach for multimodal learning analytics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376148

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. (2015). The LATUX workflow: designing and deploying awareness tools in technology-enabled learning settings. In *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/2723576.2723583

Ndukwe, I. G., & Daniel, B. K. (2020). Teaching analytics, value and tools for teacher data literacy: A systematic and tripartite approach. *International Journal of Educational Technology in Higher Education, 17*(1), 1–31. https://doi.org/10.1186/s41239-020-00201-6

Pozdniakov, S., Martinez-Maldonado, R., Tsai, Y. S., Cukurova, M., Bartindale, T., Chen, P., & Gasevic, D. (2022). The question-driven dashboard: How can we design analytics interfaces aligned to teachers' inquiry? In *Learning analytics for transition, disruption and social change—Proceedings of the 12th International Learning Analytics and Knowledge Conference* (pp. 175–185). Association for Computing Machinery. https://doi.org/10.1145/3506860.3506885

Pozdniakov, S., Martinez-Maldonado, R., Tsai, Y. S., Echeverria, V., Srivastava, N., & Gasevic, D. (2023). How do teachers use dashboards enhanced with data storytelling elements according to their data visualisation literacy skills? In *Towards trustworthy learning analytics—Proceedings of the 13th International Learning Analytics and Knowledge Conference* (pp. 89–99). Association for Computing Machinery. https://doi.org/10.1145/3576050.3576063

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies, 10*(1), 30–41. https://doi.org/10.1109/TLT.2016.2599522

Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation, 55,* 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Sweller, J. (2022). The role of evolutionary psychology in our understanding of human cognition: Consequences for cognitive load theory and instructional procedures. *Educational Psychology Review, 34*(4), 2229–2241. https://doi.org/10.1007/s10648-021-09647-0

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Van Leeuwen, A., & Rummel, N. (2020). Comparing teachers' use of mirroring and advising dashboards. In *Celebrating 10 years of LAK: Shaping the future of the field—Proceedings of the 10th International Conference on Learning Analytics & Knowledge* (pp. 26–34). Association for Computing Machinery. https://doi.org/10.1145/3375462.3375471

van Leeuwen, A., Knoop-van Campen, C. A., Molenaar, I., & Rummel, N. (2021). How teacher characteristics relate to how teachers use dashboards: Results from two case studies in K-12. *Journal of Learning Analytics, 8*(2), 6–21. https://doi.org/10.18608/jla.2021.7325

Van Merriënboer, J. J. G,, & Sweller, J. (2010). Cognitive load theory in health professional education: Design principles and strategies. *Medical Education, 44*(1), 85–93. https://doi.org/10.1111/j.1365-2923.2009.03498.x

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing, 23*, Article 2331216519832483. https://doi.org/10.1177/2331216519832483

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500–1509. https://doi.org/10.1177/0002764213479363

Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., & De Laet, T. (2020). Learning analytics dashboards: The past, the present and the future. In *Celebrating 10 years of LAK: Shaping the future of the field—Proceedings of the 10th International Conference on Learning Analytics & Knowledge* (pp. 35–40). Association for Computing Machinery. https://doi.org/10.1145/3375462.3375504

Wang, Y., Lu, S., & Harter, D. (2021). Towards collaborative and intelligent learning environments based on eye tracking data and learning analytics: A survey. *IEEE Access*, *9*, 137991–138002. https://doi.org/10.1109/ACCESS.2021.3117780.

Yan, H., & Lin, F. (2021). Including learning analytics in the loop of self-paced online course learning design. *International Journal of Artificial Intelligence in Education, 31*, 878–895. https://doi.org/10.1007/s40593-020-00225-z

---

Corresponding author: Yuchen Liu, yuchen.liu@monash.edu