JOURNAL OF LEARNING ANALYTICS

Volume 12(1), 367–390. https://doi.org/10.18608/jla.2024.8471

Investigating the Effect of Visualization Literacy and Guidance on Teachers' Dashboard Interpretation

Stanislav Pozdniakov[1]*, Roberto Martinez-Maldonado[2]*, Yi-Shan Tsai[3], Vanessa Echeverria[4], Zachari Swiecki[5] and Dragan Gašević[6]

Abstract

Recent research on learning analytics dashboards has focused on designing user interfaces that offer various forms of *visualization guidance* (often referring to notions such as *data storytelling* or *narrative visualization*) to teachers (e.g., emphasizing data points or trends with colour and adding annotations), aiding them in interpreting visual elements to gain a comprehensive understanding of students' learning processes. Yet, while some studies have explored how teachers interpret students' data through these dashboards, many have overlooked the diverse technical capabilities of teachers, which can significantly impact their use of LA dashboards. In particular, *visualization literacy* (VL) skills can greatly influence how effectively teachers interpret dashboards. To the best of our knowledge, no comprehensive account exists that details how teachers with varying VL skills interpret visual representations of students' data. In this paper, we address this gap by investigating how teachers interpret LA dashboards, both with and without visualization guidance, taking into account their VL. We illustrate this by analyzing teachers' think-aloud sessions as they engage with dashboards in the context of monitoring synchronous online learning tasks undertaken by student groups using Zoom and Google Docs. Using epistemic network analysis, we examine the differences in interpretations between teachers with varying VL levels. Our findings revealed that teachers with low VL exhibited shallower dashboard interpretations than those with high VL. However, the association of VL with successful task completion rate was not significant. Also, visualization guidance did not enable teachers to deepen their interpretations. While some visualization guidance helped teachers to complete tasks correctly, excessive visualization guidance can also be detrimental.

Notes for Practice

- Teacher-facing learning analytics (LA) dashboards are ultimately aimed at supporting teachers in the interpretation of student data in order to gain a deeper understanding of students' learning processes.
- However, the effectiveness of LA dashboards depends on both the guidance offered to teachers in interpreting visual elements and teachers' varying visualization literacy (VL).
- We found that teachers with low VL in general exhibited shallower dashboard interpretations than teachers with high VL.
- Teachers with low VL commonly used single visualizations during their interpretation, while teachers with high VL often performed deeper interpretations using multiple visualizations.
- Visualization guidance did not affect teachers' interpretation, regardless of their VL.
- Teachers' VL was not associated with the correct completion of tasks. Yet, less visualization guidance was actually associated with an improved rate of correct task completion; thus, explicit visualization guidance, e.g., juxtaposing the text summary with the corresponding charts, might become a distraction for teachers.

Keywords

Learning analytics, dashboards, human-centred design, data literacy, visualization guidance, data storytelling.

Submitted: 12/05/2024 — **Accepted:** 11/11/2024 — **Published:** 18/12/2024

[1]* *Corresponding author Email: pozdniakov@uq.edu.au Address: Centre for Learning Analytics Monash (CoLAM), Faculty of Information Technology (FIT), Monash University, Australia, and School of Electrical Engineering and Computer Science, The University of Queensland, Australia. ORCID iD: https://orcid.org/0000-0003-4451-9181*

[2]* *Corresponding author Email: roberto.martinezmaldonado@monash.edu Address: CoLAM, FIT at Monash University, Australia. ORCID iD: https://orcid.org/0000-0002-8375-1816*

[3]Email: : yi-shan.tsai@monash.edu Address: CoLAM, FIT at Monash University, Australia. ORCID iD: https://orcid.org/0000-0001-8967-5327
[4]Email: : vanessa.echeverria@monash.edu Address: CoLAM, FIT at Monash University, Australia, and Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador. ORCID iD: https://orcid.org/0000-0002-2022-9588
[5]Email: : zach.swiecki@monash.edu Address: CoLAM, FIT at Monash University, Australia. ORCID iD: https://orcid.org/0000-0002-7414-5507
[6]Email: : dragan.gasevic@monash.edu Address: CoLAM, FIT at Monash University, Australia. ORCID iD: https://orcid.org/0000-0001-9265-1908

1. Introduction

There is growing interest in using learning analytics (LA) to provide teachers with technology-enhanced support (Salas-Pilco et al., 2022; Sousa et al., 2021). Dashboards have become a preferred end-user interface design in education, offering a direct view to teachers of students' data (Fernandez Nieto et al., 2022). They are increasingly used in various learning environments, including online courses (Dourado et al., 2021), MOOCs (Chavan & Mitra, 2022), collaborative learning settings (Bao et al., 2021), and K–12 classrooms (Dickler et al., 2021). For most of the educational community, dashboards represent the most tangible aspect of LA. They are often the first and most direct interaction that members of the educational community have with LA (Verbert et al., 2020). These dashboards depict a variety of information, such as student engagement and progress toward learning objectives.

Yet, despite efforts to create effective teacher-facing dashboards, there is mixed evidence regarding how well they support teachers (Campos et al., 2021; Molenaar & Knoop-van Campen, 2019; Wise & Jung, 2019; Herodotou et al., 2019). This can be attributed to at least two main factors: (i) inappropriate dashboard designs not meeting teachers' needs or the intentions in their learning designs (Hernández-Leo et al., 2018; Kaliisa et al., 2022; Sampson, 2017) and (ii) teachers' limited proficiency in using these tools (Herodotou et al., 2019). In particular, when using dashboards to monitor student progress, teachers may struggle to interpret the data correctly due to insufficient technical capabilities (Campos et al., 2021).

Indeed, recent research flagged the need to investigate teachers' technical capabilities to effectively use dashboards (van Leeuwen et al., 2021). Ndukwe and Daniel (2020) suggested that teachers with limited technical proficiency are at risk of misinterpreting student data, which could lead to poor decision-making. These technical capabilities, specifically the ability to proficiently use dashboards that contain data visualizations, are defined as *visualization literacy* (VL). VL involves the skill of engaging with and understanding visualizations (Lee et al., 2017), including the ability to interpret visual elements representing data and draw accurate conclusions from them. Although teacher-facing dashboards often feature complex visualizations (Kaliisa et al., 2023), the concept of VL is not sufficiently addressed in the existing LA literature.

A different stream of research in LA has proposed to support the interpretation of dashboards by introducing *visualization guidance* approaches into the design of LA dashboards. Visualization guidance is a dynamic computer-assisted procedure, purposed to proactively address any knowledge gaps that users might encounter throughout their engagement with visualizations (Ceneda et al., 2017). This procedure typically comes in the form of visual cues and textual prompts that guide users to focus on relevant data charts and visual elements (Barral et al., 2021). As a result, visualization guidance is aimed at reducing the burden on users of finding contextually relevant information, simplifying the interpretation process, and more explicitly communicating insights (Ceneda et al., 2017). In LA, research on guidance in teacher-facing dashboards commonly draws from two main areas, namely, teachers' professional vision (van Leeuwen et al., 2019) and sensemaking (Wise & Jung, 2019; Poquet, 2024), focusing primarily on how visualization guidance as a whole affects teachers' use of dashboards (van Leeuwen et al., 2019, 2021; van Leeuwen & Rummel, 2022), and works drawing from literature on the information visualization literature and human-centred LA (Echeverria, Martinez-Maldonado, Granda, et al., 2018; Martinez-Maldonado et al., 2020; Fernandez Nieto et al., 2022; Pozdniakov, Martinez-Maldonado, Tsai, Echeverria, et al., 2023), focusing on the novel visualization approaches to enable or disable such guidance.

Extensive research conducted by van Leeuwen and colleagues (van Leeuwen et al., 2017; van Leeuwen, 2019; van Leeuwen & Rummel, 2022) suggests that to enable teachers, guidance on the dashboard should be aiming at *detection of pedagogically important events* and aiding in *teachers' interpretation of the important events*, where teachers could explore how these events are situated within the broader context. To enable such support for teachers, these authors explored how the increasing level of guidance in teacher dashboards could support the detection and interpretation of important events related to students' learning (van Leeuwen et al., 2019). While these works emphasized the benefits of incorporating visualization guidance in teacher-facing dashboards, Nazaretsky, Cukurova, and Alexandron (2022) and Nazaretsky, Ariely, and colleagues (2022) suggested that educational technology, which incorporates explicit direct textual recommendations or prescriptions, should necessarily allow teachers to disable such prescriptions to ensure that they can make their own instructional judgments.

A deeper focus on the visualization design to guide teachers' attention is jointly presented in works by Echeverria and colleagues (Echeverria, Martinez-Maldonado, Buckingham Shum, et al., 2018; Martinez-Maldonado et al., 2020; Fernandez Nieto et al., 2022). For example, Echeverria, Martinez-Maldonado, Granda, and colleagues (2018) recommended enhancing dashboard visualizations with elements that highlight key data points and trends, along with explanatory text, to aid teachers' interpretation. Similarly, some authors have proposed using metaphoric visualizations (Molenaar et al., 2020), drill-down narrative slideshows (Chen et al., 2019), and textual explanations and recommendations (Echeverria, Martinez-Maldonado, Buckingham Shum, et al., 2018; Fernandez Nieto et al., 2022). Additionally, integrating visualizations into layers based on the pedagogical context (Martinez-Maldonado et al., 2020) and employing novel visualization techniques for insights from multimodal analytics in classrooms, have been explored (G. Fernandez-Nieto et al., 2022). These works collectively examine how visualization guidance in dashboards can help teachers focus their attention and simplify the interpretation of student data. However, while some studies have investigated how teachers interpret student data through these dashboards, the aspect of teachers' VL has largely been overlooked in LA research.

In the LA community, van Leeuwen and colleagues (2019) suggest a visualization guidance framework that differentiates teacher-facing dashboards into three categories: dashboards containing no visualization guidance (i.e., "mirroring" dashboards), dashboards containing varying extents of visualization guidance (i.e., "alerting dashboards"), and "advising" or "guiding" dashboards. Building on this work, the study conducted by Kasepalu and colleagues (2023) introduced changes to this operationalization, combining dashboards implementing alerting and guiding visualization guidance into one category. The authors investigated how teachers' situational awareness and workload vary depending on the visualization guidance embedded in the dashboard used. Overall, visualization guidance can reinforce their understanding and interpretation of the data, allowing them to draw more accurate and nuanced conclusions. The reasons for these improvements stem from the fact that visualization guidance provides a structured approach to data interpretation, highlighting key data points and trends and providing explanatory text to aid understanding. Visualization guidance can significantly improve dashboard interpretations for teachers with varying levels of VL in several ways. Visualization guidance could scaffold low VL such that it would help to improve plot comprehension by emphasizing relevant elements on the charts (Barral et al., 2020), simplifying plot legends and thus helping teachers focus their attention on the most relevant data, and simplifying the interpretation process (Lalle et al., 2021). In the case of high VL, it helps make plot interpretation faster by decluttering chart comprehension (Pozdniakov, Martinez-Maldonado, Tsai, Echeverria, et al., 2023), which would result in faster interpretation of students' data depicted on the dashboards (Barral et al., 2020).

The potential interplay between VL, visualization guidance, and teachers' interpretation of LA dashboards has been emphasized by Pozdniakov, Martinez-Maldonado, Tsai, Echeverria, and colleagues (2023) and Liu and colleagues (2024), who found differences in teachers' dashboard use depending on teachers' VL. However, these authors focused solely on a quantitative analysis of teachers' gaze behaviours and pupil diameter when inspecting the dashboards, without conducting an in-depth, qualitative exploration of how teachers interpret dashboards with and without visualization guidance. Consequently, the degree to which teachers' interpretation is influenced by their VL and the guidance provided by dashboards remains unexplored. The knowledge regarding the differences in teachers' interpretation is central to advancing and generalizing research results on teacher-facing sensemaking. However, Poquet (2024) suggested that applications of sensemaking theories in LA lack shared theoretical foundations. In the context of teacher-facing dashboards, this issue complicates the generalization of results regarding teachers' use of the dashboards. To address this problem and enable further generalization of results across studies, Poquet (2024) proposed a structured framework to operationalize and describe sensemaking used in LA studies.

To address this gap, we investigate how teachers with varying VL skills interpret students' data depicted via alerting and advising (with visualization guidance) and mirroring (no guidance) dashboards. As such, this study relies on the visualization guidance taxonomy used in Kasepalu and colleagues (2023), which proposed to combine dashboards implementing alerting and advising visualization guidance into a single type. In response to the call of Poquet (2024), we use the proposed framework to operationalize the sensemaking constructs used in our study. We asked 23 consenting teachers, with different VL levels, to participate in think-aloud sessions, where they interpreted students' data depicted via two types of LA dashboards: with and without visualization guidance. To analyze these data, we employed epistemic network analysis (ENA) to contrast how teachers with varying VL interpret students' data depicted via the dashboards, depending on the dashboard type used. The dashboards used in the study were designed to support teachers monitoring synchronous small-group online learning sessions. This paper contributes to the literature on LA teacher-facing dashboards by empirically investigating (a) the role of teachers' VL in the interpretation of students' data, (b) the role of visualization guidance in teachers' interpretation, and (c) the association of teachers' VL and visualization guidance in scaffolding the correct completion of monitoring tasks.

2. Background

2.1 Teacher-Facing Dashboards

Teacher-facing dashboards are used in different settings and for various instructional purposes. Dashboards can be used in real-time *monitoring* scenarios where teachers need to make sense of students' data and act immediately (Martinez-Maldonado et al., 2020; van Leeuwen et al., 2021; Dickler et al., 2021; Pozdniakov et al., 2022). Teachers might also use dashboards asynchronously, at their convenience. Dashboards are used to monitor students' progress in assignments and adjust the assignment design accordingly (Chavan & Mitra, 2022; Wise & Jung, 2019). Teachers can use dashboards when they provide feedback to students (Ahn et al., 2019; Khosravi et al., 2021; Rets et al., 2023; Echeverria et al., 2024) and when engaging in high-level educational management (Deakin Crick et al., 2017).

However, as the complexity of data gathered from various educational platforms and settings increases (Martinez-Maldonado et al., 2023), a more thoughtful approach to how diverse dashboards for teachers can aid in understanding and determining appropriate pedagogical responses is needed (Jivet et al., 2018). Designing teacher dashboards that are easy to understand and can aid in interpreting students' data is a constant challenge (Jørnø & Gynther, 2018; Wise & Jung, 2019; Echeverria et al., 2024). Despite the growing number of teacher-facing dashboards designed to assist teachers, teachers often struggle to interpret student activity and make informed decisions using most of the current dashboards designed for them (Sloan-Lynch & Morse, 2024).

Although the literature has repeatedly emphasized the importance of careful consideration and investigation of teachers' interpretation and sensemaking of data when using teacher-facing dashboards (Pozdniakov, Martinez-Maldonado, Singh, et al., 2023) and explainable LA interfaces (Khosravi et al., 2022), reviews indicate that only a minority of studies explicitly consider how dashboards scaffold teachers' interpretation during the evaluation (Schwendimann et al., 2017; Sahin & Ifenthaler, 2021). This oversight is significant, as teachers' interpretation process is crucial for gaining insight into students' learning (Wise & Jung, 2019; Verbert et al., 2013). Such understanding could enable practitioners and researchers to refine dashboard designs (Ahn et al., 2019) and potentially develop innovative methods to support teachers in their interpretation tasks.

2.2 Teachers' Data Interpretation in LA

Data interpretation is a key process for teachers using dashboards. This involves orientation, where they overview the data visualized in the dashboard (Wise & Jung, 2019) and then identify and describe specific elements, integrating their broader knowledge (van Leeuwen et al., 2019). "Teachers' noticing" is the ability to select specific indicators of student behaviour, such as classroom events or signs of misunderstanding (Van Es & Sherin, 2002). It also includes prioritizing pedagogically significant elements and evaluating student progress in context. Van Leeuwen and colleagues (2019) proposed a coding scheme for teachers' dashboard interactions, based on teachers' noticing literature (Van Es & Sherin, 2002). The scheme detailed teachers' interpretation, including (i) dashboard elements used, (ii) specific performance indicators, and (iii) interpretation depth. The scheme structured teachers' responses during dashboard interactions.

The sensemaking process typically happens non-deliberately when users face analytics. Interpretation is considered to be one of the main elements of teachers' sensemaking when they face analytics (Wise & Jung, 2019; Pozdniakov, Martinez-Maldonado, Singh, et al., 2023). Verbert and colleagues (2013) presented a sensemaking model that is applicable for both students and teachers. Building on this work, Wise and Jung (2019) presented a sensemaking model targeted specifically at teachers using LA. This model suggests two types of sensemaking: question-driven, where teachers have a clear question in mind when they face LA, and data-driven, where teachers do not have a clear question but rather explore students' data to find patterns. As such, this model highlights the fundamental role of the questions teachers have in mind to support their sensemaking. Campos and colleagues (2021) introduced their version of the teacher-facing dashboard sensemaking model, including notions of emotional response to analytics, overfitting, and vantage points. Recently, Poquet (2024) critically reviewed existing LA studies that use sensemaking theories and found that these studies combine different paradigms without coherently describing sensemaking. To address this problem, the author revised key sensemaking theories from broader fields, such as human-computer interaction and cognitive theory, and proposed novel unified lenses on sensemaking in LA. In particular, Poquet (2024) suggested that LA studies using sensemaking lenses should operationalize *activities* to be undertaken in the typical *situation* of dashboard use, as well as what and how stakeholders *notice* (e.g., specific dashboard indicators) and *perceive* (e.g., using single or multiple visualizations) when interpreting students' data.

Previous studies have examined two main aspects of teachers' interpretations of data depicted via dashboards: (a) the indicators teachers focus on, their data interpretation strategies, the decisions they make, and the impact of their actions (Wise & Jung, 2019; van Leeuwen et al., 2019; Molenaar & Knoop-van Campen, 2019; Bao et al., 2021), and (b) the effect of basic teacher characteristics on interpretation, dashboard use, and feedback type (Campos et al., 2021; van Leeuwen et al., 2021). Wise and Jung (2019) identified two main strategies teachers use when interpreting student data from LA dashboards: *reading data* and *explaining patterns*. Van Leeuwen and colleagues (2019) found that dashboard type had an effect on the interpretation of student progress but not the detection of problematic learning situations. Similar findings were reported by Bao and colleagues (2021) regarding teachers' diagnostic capabilities. Molenaar and Knoop-van Campen (2019) found that frequent dashboard consultation led to more diverse feedback and a greater focus on specific interpretation strategies. Campos and colleagues (2021) found that novice and expert teachers use different strategies to interpret dashboard data. Van Leeuwen and colleagues (2021) investigated whether teacher characteristics could account for variance in dashboard use. They concluded that teacher characteristics like experience, age, gender, and self-efficacy did not explain differences in dashboard use, contrasting with Campos and colleagues (2021)'s findings.

In summary, research has examined teachers' interpretation of student data on dashboards, focusing on (i) *indicators of students' progress* such as group trends or individual characteristics (van Leeuwen et al., 2019; Wise & Jung, 2019); (ii) *visualization elements* or *reference frames* like single or multiple visualizations (van Leeuwen et al., 2019; Wise & Jung, 2019; Li et al., 2021); and (iii) *interpretation depth*, referring to the extent of teachers' contextualization and elaboration (van Leeuwen et al., 2019; Campos et al., 2021). While these studies offer empirical insights into dashboard use, only van Leeuwen and colleagues (2019) compared interpretations across different dashboards, finding a correlation between the inclusion of attention-guiding visual elements and deeper interpretation.

2.3 Visualization Guidance in Teacher-Facing Dashboards

Visualization guidance is an interactive process that aids users in interacting with visualizations (Ceneda et al., 2017). It uses visual indicators and written hints to direct attention to charts and visual components (Barral et al., 2021). The goal is to streamline interpretation and convey insights more clearly (Ceneda et al., 2017). It is emerging in LA to highlight key insights and declutter visualization (Milesi & Martinez-Maldonado, 2024).

Various approaches have been proposed for visualization guidance in teacher-facing dashboards. Echeverria, Martinez-Maldonado, Granda, and colleagues (2018) formulated a conceptual model aligning dashboard design with learning design. A follow-up study found improved interpretation capabilities with visualization guidance (Echeverria, Martinez-Maldonado, Buckingham Shum, et al., 2018). Novel approaches continue to evolve, including a narrative drill-down approach to simplify complex MOOC data (Chen et al., 2019), a design approach for mapping key events of students' progress (Martinez-Maldonado et al., 2020), and a dashboard design approach that maps visualizations to teachers' questions (Pozdniakov et al., 2022). However, the appropriateness of traditional visualization techniques for teacher dashboards has been questioned (Fernandez Nieto et al., 2022).

Two taxonomies are proposed to characterize visualization guidance in LA. Echeverria, Martinez-Maldonado, Buckingham-Shum, and colleagues (2018) delineated visualization guidance in LA into exploratory and explanatory, primarily focusing on the latter type. Furthermore, van Leeuwen and colleagues (2019) suggested differentiating teacher-facing dashboards based on the extent of visualization guidance they contain, classifying dashboards into three types: mirroring (no guidance), alerting (some guidance), and advising (high guidance). Mirroring dashboards typically reflect existing data without changes to help scaffold the interpretation and detection of instructionally important events. Alerting dashboards contain certain data transformations and employ heuristics, algorithms, or machine learning models to identify relevant details, which are then reflected or emphasized on the dashboard. Advising or guiding dashboards, as operationalized in work conducted by Kasepalu and colleagues (2023), not only depict students' data but also employ various information visualization techniques to aid teachers in interpreting relevant details.

The exploratory visualization guidance is more appropriate for the cases of sensemaking which happen when teachers do not yet have clear questions in mind and do not have clear expectations from analytics; hence, guidance support for divergent thinking is required, as illustrated by the work of Chen and colleagues (2019). The explanatory visualization guidance is more suited for situations when teachers have well-formulated questions when facing the analytics; thus, elements such as colour emphasis and textual explanations could be added to support this process. This visualization guidance support is illustrated by works of Pozdniakov and colleagues (2022) and Fernandez Nieto and colleagues (2022). The work of Martinez-Maldonado and colleagues (2020) lies in between, effectively incorporating visualization guidance to support both exploratory and explanatory purposes. Importantly, for both types of visualization guidance, providing support for the interpretation of students' data is crucial.

Evidence is limited on whether visualization guidance affects teachers' interpretation capabilities. Kasepalu and colleagues (2023) explored how teachers' situational awareness and workload vary depending on using no dashboard at all, using dashboards without visualization guidance, and using dashboards with visualization guidance. They found that even a dashboard without guidance (mirroring) increased teachers' situational awareness compared to not using any dashboard at all, enabling them to better perceive and understand classroom events. In contrast, dashboards with guidance (alerting and advising) significantly reduced teachers' perceived workload. While this study provided evidence regarding the positive effects on perceived cognitive load when teachers use dashboards with visualization guidance, it did not explore the differences in teachers' interpretation. Only two studies have investigated this (van Leeuwen et al., 2019; Pozdniakov, Martinez-Maldonado, Tsai, Echeverria, et al., 2023). Van Leeuwen and colleagues (2019) found differences in interpretation depending on the type of dashboard used but no association between visualization guidance and task completion. This suggests that while visualization guidance may influence data interpretation, it does not necessarily improve task performance. Van Leeuwen and colleagues (2019) implemented guidance by emphasizing the active group with a specific icon and including textual explanations. Pozdniakov, Martinez-Maldonado, Tsai, Echeverria, and colleagues (2023) used a combination of colour emphasis, textual explanations, and questions as titles to guide teachers' attention.

Pozdniakov, Martinez-Maldonado, Tsai, Echeverria, and colleagues (2023) explored how teachers with different VL visually inspect dashboards with visualization guidance. They found that teachers with low VL benefited more from explicit visualization guidance. However, the study only included dashboards with visualization guidance. Subsequent studies found differences in how teachers used visualization guidance dashboards depending on their VL but did not include dashboards without visualization guidance (Pozdniakov, Martinez-Maldonado, Tsai, Srivastava, et al., 2023; Liu et al., 2024). This leaves a gap in our understanding of how visualization guidance is associated with teachers' abilities to interpret data.

2.4 Teachers' VL

While understanding how teacher characteristics correlate with dashboard use is important, van Leeuwen and colleagues (2021) found little evidence that basic characteristics like age, gender, and teaching experience account for variation in use. Thus, exploring more specific constructs related to dashboard use, like VL, was suggested (Pozdniakov, Martinez-Maldonado, Tsai, Srivastava, et al., 2023; Liu et al., 2024). VL, the ability to read, interpret, and draw conclusions from visual data representations (Lee et al., 2017), can be an appropriate construct for exploring teachers' interpretation of learning data on dashboards. Several methods to estimate VL exist. Burns and colleagues (2020) proposed using Bloom's taxonomy to analyze user interactions with visualizations and task completion. Lee and colleagues (2017) created the Visualization Literacy Assessment Test (VLAT) to quantitatively evaluate VL. VLAT and its variations (Pandey & Ottley, 2023) have been used to study the connection between users' VL and users' gazing behaviour when introduced to novel visualization techniques (Lalle et al., 2021) and to study the use of teacher-facing dashboards. Donohoe and Costello (2020) used VLAT to study teachers' VL against self-reported confidence in data interpretation, finding that teachers often overestimate their interpretation capabilities. Pozdniakov, Martinez-Maldonado, Tsai, Echeverria, and colleagues (2023) used VLAT to compare how teachers with low and high VL interact with two dashboards with visualization guidance. They found that teachers with low VL began to exhibit strategies similar to those of high-VL teachers when examining dashboards with more explicit visualization guidance (Pozdniakov, Martinez-Maldonado, Tsai, Echeverria, et al., 2023). However, there is currently no evidence regarding how teachers' interpretation of students' data depicted via dashboards would be associated with the interplay between dashboard design (with and without visualization guidance) and teachers' VL skills. This gap necessitates further research.

2.5 Contribution and Research Questions

As such, this study focuses on investigating teachers' interpretation in sensemaking situations when teachers have a clearly defined question when they face mirroring dashboards (no visualization guidance) and alerting and advising dashboards (with visualization guidance). The literature has flagged the role of VL in interpreting students' data depicted via dashboards with and without visualization guidance but has not provided any conclusive evidence. Following the dashboard classification outlined in van Leeuwen and colleagues (2019), dashboards without visualization guidance correspond to mirroring dashboards, while dashboards with visualization guidance correspond to alerting and advising dashboards. The current study focuses on the interpretation of students' data using dashboards during sensemaking when teachers have a clearly defined question. Answering the recent call for generalizability of studies involving sensemaking theories, this work also uses a shared lens framework suggested to describe the results of sensemaking studies (Poquet, 2024). To bridge this gap in knowledge, we formulated the following overarching research question (**RQ**): How do teachers with varied VL interpret students' data presented via LA dashboards with and without visualization guidance? To answer this question, the following sub-RQs were formulated:

RQ1: What are the differences in the ways teachers with varying VL skills interpret students' data presented via LA dashboards?

RQ2: What are the differences in the ways teachers with varying VL skills interpret students' data presented via LA mirroring (without guidance) and alerting and advising (with guidance) dashboards given that teachers complete tasks of varying difficulty?

RQ3: To what extent are teachers' VL and LA dashboards with varying visualization guidance associated with correct task completion?

3. Methods

This section first describes the dependent and independent variables used in the study and the learning context where the study was conducted; then operationalizes the sensemaking constructs used in the study; and then describes various teacher-facing dashboard designs presented to teachers, the study procedure, information about the study participants, and the analysis.

Two independent variables are used in the study: *teachers' VL*, consisting of two levels, low VL and high VL, and *visualization guidance*, consisting of three levels, no visualization guidance (mirroring dashboard) and dashboards with less and more explicit visualization guidance (two alerting and advising dashboards). Four dependent variables are used in this study: teachers' *depth of interpretation*, i.e., participants' tendency to engage in simpler or deeper interpretations, operationalized as dimension MR1, resulting from the ENA model; teachers' *use of single or multiple visualizations* during the interpretation, operationalized as dimension SVD2, resulting from the ENA model; *correct task completion*, operationalized as the number of correct responses participants provided when they used a dashboard (ranges from 0 to 5); and a *task category* consisting of two levels, "comprehension" and "retrieval," operationalized using Bloom's taxonomy.

3.1 The Learning Context

The study was conducted in the context of an authentic graduate course at Monash University to aid teachers in overseeing synchronous online tasks in the IT Research Methods course. Weekly 3-hour online classes were conducted by two teachers, with students completing various tasks in groups of four to six using tools like Zoom and Google Docs. Two examples of learning activities happening in a single online session, which students are required to work on in groups, were as follows:

- **Activity 1—Audio Reflection** (15 min): Post-presentation, discuss challenges. Consider (1) helpful aids for description, (2) better descriptive language strategies, and (3) useful visual aids.

- **Activity 2—Visual Metaphors** (30 min): Reflect on your audience and plan: (1) Identify two challenging visual elements from your plan. (2) Discuss these with your group and document them in the Activity 2 table. (3) Brainstorm visual representation ideas, and provide online content links if useful. (4) Justify each idea's effectiveness for audience understanding.

3.2 Operationalization of Sensemaking Constructs Used in the Study

Following the shared lens framework suggested by Poquet (2024), we operationalize the sensemaking-related constructs used in the current study. The *focus* is a think-aloud study where teachers inspect LA dashboards with and without visualization guidance. The *activity system* used in the study is characterized by data collected from synchronous collaborative online students' activities when they work using audiovisual and text-based tools, such as Zoom and Google Docs.

The *definition of a situation*, from a teacher's perspective, involves detecting which group might be stuck or experiencing issues with progressing in the planning discussion or executing the activity and determining which group the teachers should visit next to help students. The following exemplar situations depict types of situations teachers could face: (i) Students are silent in the breakout room, with no indication of either discussion or execution visible on the dashboards. (ii) Several groups have high discussion time, but there are no traces of executing the activity in Google Docs. (iii) There are signs of activity in Google Docs, but none of the students have talked to each other.

Noticing is operationalized as teachers' ability to identify information related to students' behaviour (see Section 3.6, Table 2, *(1) Specificity of students' performance*). *Perceiving* is operationalized as teachers referring to single or multiple affordances, such as referring to single or multiple visualizations during their think-aloud (see Section 3.6, Table 2, *(2) Specificity of using visualization devices*). Additionally, teachers' reflections related to general techniques or strategies they use when interpreting data, or reflections regarding their ability to identify relevant elements of visualizations that scaffold interpretation (see Section 3.6, Table 2, *(4) Metacognitive reflections*), might also be attributed to perceiving.

Interpreting is defined similarly to how it was done by van Leeuwen and colleagues (2019) and is characterized by the depth and quality of participants' evidence triangulation and argumentation regarding student or group behaviour. In this case, it partially encompasses both noticing and perceiving.

3.3 Dashboard Design

The dashboards, developed to assist teachers in Zoom monitoring, use data from an open-source analytics system, ZoomSense (Bartindale et al., 2021). This system uses virtual agents in Zoom breakout rooms to monitor student interactions in Google Docs. The section elaborates on three dashboards. The pre-existing dashboard, part of ZoomSense, with no guidance (**NG**), is depicted in Figure 1. Dashboards with visualization guidance elements (**VG1** and **VG2**), created with teachers (Pozdniakov et al., 2022), are shown in Figure 2. These dashboards are designed for real-time teacher monitoring during student collaboration in breakout rooms. Authentic deployment results suggest that dashboards aid teachers in interpreting the data. All dashboards feature a timeline chart (*T*) at the top, marking class activities and times. The current activity is highlighted in blue. The original dashboard, displayed in Figure 1, includes four visualizations: pie charts, sociograms, timeline sequence diagrams, and Google Docs horizontal bars. Teachers switch views using buttons (see *N* in Figure 1). Pie charts (*V6* in Figure 1) show each student's speaking time. Sociograms (*V4* in Figure 1) depict participant communication. Timeline sequence diagrams (*V7** in Figure 1) reveal turn-taking dynamics. A horizontal bar (*V2* in Figure 1) tracks Google Docs progress, with blue, orange, and grey indicating completed, in-progress, and not-started activities, respectively.

NG dashboard—no teachers involved in the design The dashboard with no visualization guidance was originally developed as an open-source project and teachers were not involved in the design of that dashboard version. The importance of accounting for the needs of educational stakeholders when designing LA to enable its benefits has been emphasized before and constitutes the heart of human-centred LA (Buckingham Shum et al., 2019). However, less attention has been paid to design approaches and visual techniques that can facilitate the interpretation of dashboards to support sensemaking.

VG1 dashboard designed using a participatory approach To overcome the dual limitation of absence of teacher involvement and no explicit focus on visual techniques facilitating the interpretation of students' progress in dashboards, we followed

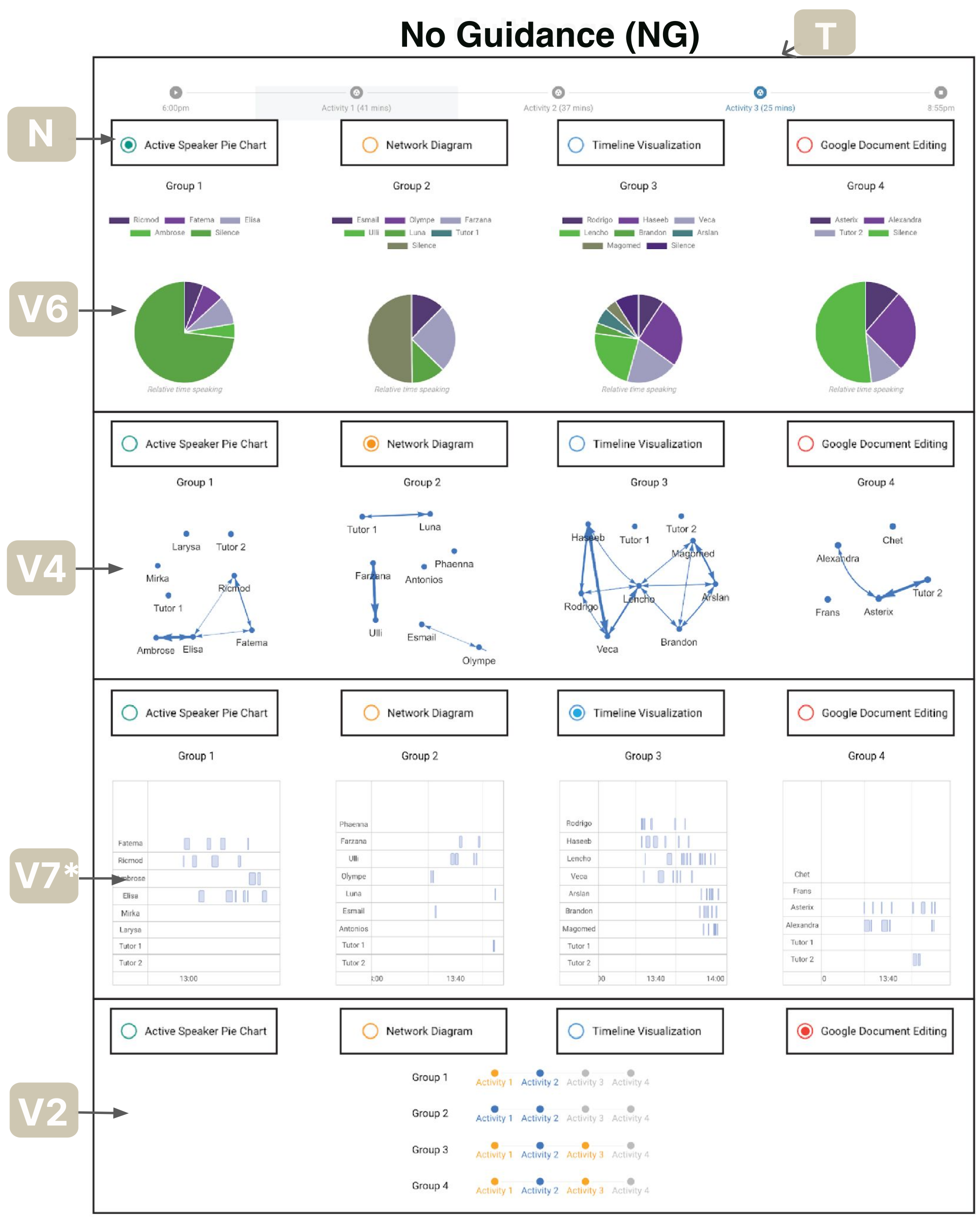

Figure 1. The initial dashboard (NG) comprises *T*, a timeline chart; *N*, navigation buttons for view transitions; *V6*, pie charts showing the proportion of time each student spoke in a group; *V4*, sociogram charts depicting the presence or absence of communication between students; *V7**, timeline sequence diagrams depicting the order in which students spoke in a group; *V2*, a horizontal bar showing Google Docs progress. *V6* and *V7** are charts unique to the NG dashboard.

participatory methods guidelines (Sarmiento & Wise, 2022; Alfredo et al., 2024), conducting semi-structured interviews with 15 consenting teachers who had experience teaching online courses via Zoom. We aimed to elicit the requirements for the dashboards suited for monitoring synchronous group activities conducted online. By conducting an inductive thematic analysis, we identified the most critical challenges teachers experienced during their instruction. This was followed by the identification of the salient questions that teachers commonly have. These questions were formulated using at least some of the words that participants used to express their actual concerns and the sources of evidence they would require. The elicited questions served as a main driver for the re-design of the dashboard with no visualization guidance. Particularly, the questions that addressed the concerns of the majority of the teachers were explicitly added to the dashboard. This was followed by deciding which visualization from the original dashboards, without guidance, could be primarily used to answer a particular question. If no existing dashboard could be used for this purpose, a new appropriate visualization that could provide evidence supporting

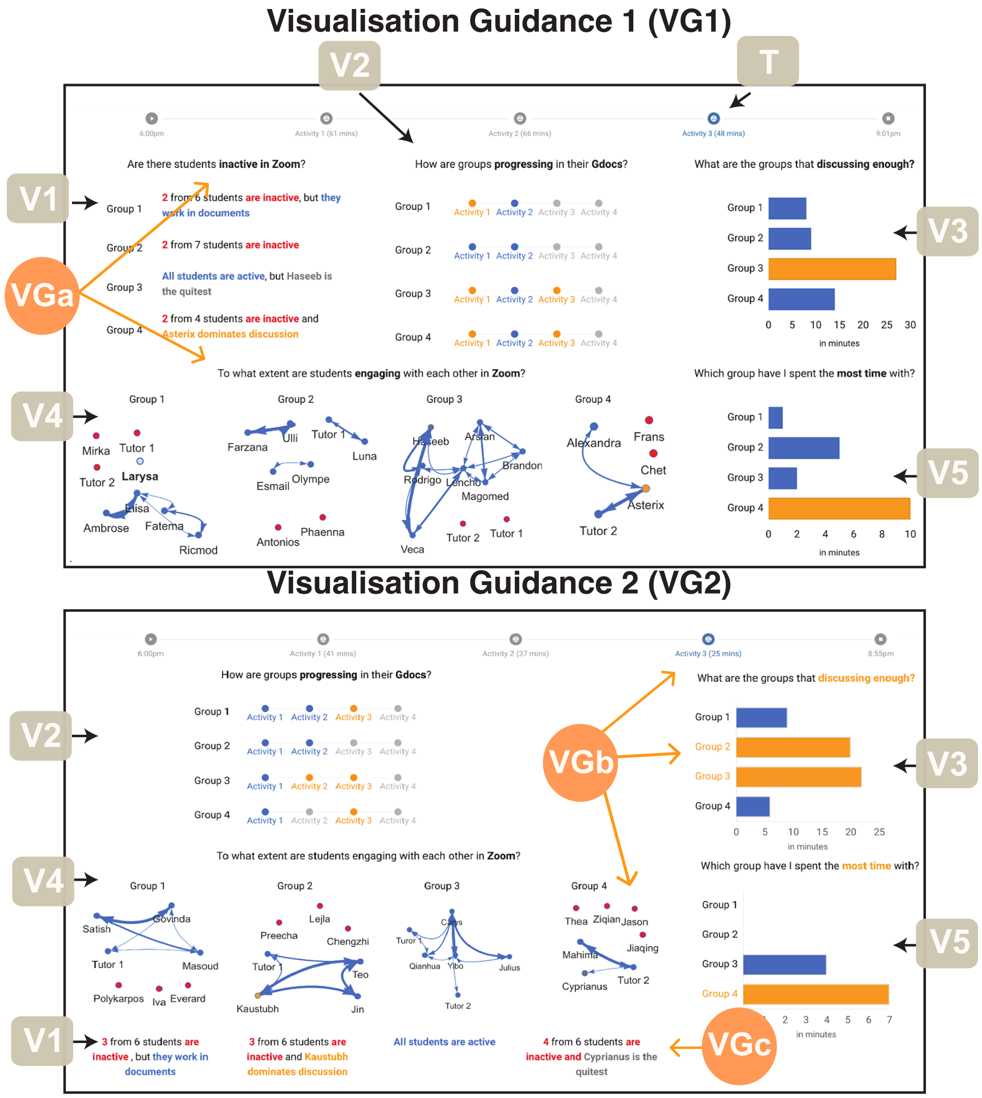

Figure 2. Two iterations of teacher dashboards with visualization guidance were designed, resulting in VG1 and VG2. These dashboards feature (VGa) teachers' questions, (VGb) colour emphasis, and (VGc) textual narratives. *T* is a timeline chart. *V1* categorizes students based on Zoom speech data. *V2* indicates progress in Google Docs. *V3* and *V5* show students' speech amounts in each group. *V4* is a sociogram chart representing communication between students or teachers.

answering the question was proposed. Lastly, we added the visualization guidance elements supporting the interpretation of the data depicted via a particular chart.

Both original dashboards and dashboards with guidance share two visualizations—sociograms and Google Docs horizontal bars. Based on Echeverria, Martinez-Maldonado, Granda, and colleagues (2018), we added the following visualization guidance elements to the dashboard: (VGa) teachers' questions, (VGb) colour emphasis, and (VGc) textual narratives. *VGa-Questions*, V1–V5 in Figure 2 (top panel), were defined by teachers in a prior study (Pozdniakov et al., 2022)). *VGb-Colour emphasis* uses three contrasting colours consistently, with grey and navy blue for most elements and orange and red for emphasis. *VGc-Textual narrative*, V1 in Figure 2 (bottom panel), briefly describes students' behaviours. The teacher-designed dashboards include two additional visualizations. A narrative-based chart classifies students based on Zoom speech data. Two bar charts show students' speech amount and teacher's time in a breakout room. Visualization guidance elements are applied to these visualizations. For sociograms, nodes are coloured if a student did not communicate or communicated more than peers in the last 5 minutes.

VG2 dashboard resulting from the evaluation with teachers We conducted a two-step evaluation with four consenting teachers. We first asked them to use the dashboard in a situation resembling an authentic classroom. This was followed by deployment where the first author of this work participated in the online sessions when teachers used the dashboard in the authentic online sessions. The results of the formal evaluation (Pozdniakov et al., 2022) and observations made when observing teachers using the dashboard VG1 were as follows. During the deployment of dashboard VG1, teachers explained that they needed more time to interpret sociograms (Figure 2, V4). Moreover, some of the teachers treated V1 and V4 as visualizations that represent two separate channels of data, when in fact they represented the same data in different ways. Based on this evidence, and drawing on prior research (Choudhry et al., 2021; Lalle et al., 2021), which found adding text to highlight key features of visualizations can improve understanding, we combined visualizations V1 and V4 (Figure 2). We envisaged that this would further declutter the dashboard and would improve comprehension of data depicted via sociograms by easing interpretation since it would be easier to use sociograms either by themselves or accompanied with narrative-based charts under the sociograms to interpret the data. We also added additional emphasis to the bar visualization labels (Figure 2, V3 and V5), aiming to drive teachers to areas that they should pay attention to in order to scaffold their interpretation, namely, the relevant part of the bar visualization group labels. We added colour emphasis to parts of the "closed questions," where a one-word answer would suffice, i.e., for two bar plots located under the questions "What are the groups that discuss enough?" and "Which group have I spent the most time with?" (see bar charts in Figure 2, V3, V5). As such, we envisaged that this change would help teachers better distribute attention in dashboard VG2 compared to VG1. These changes resulted in dashboard VG2. Both VG1 and VG2 dashboards provided the same information using six and five visualizations, respectively.

3.4 Participants

The study was advertised among faculty instructors, and consenting instructors were invited to the study lab on campus. All participants had prior experience with teaching classes via Zoom. A total of 23 university teachers (13 females and 10 males), ranging in age from 22 to 60 years old ($M = 31$, $SD = 8$), agreed to take part in the study. Except for one teacher, all had attained a postgraduate degree, with eight holding master's degrees, 12 being PhD students, and two possessing a PhD qualification. Additionally, all teachers held STEM degrees and either currently worked as university teachers (i.e., teaching associates and lecturers) or had an average of three years of teaching experience ($SD = 3.4$).

3.5 Study Procedure

I. Dashboard inspection Participants were invited to a specialized room on campus to use three dashboard versions for student progress evaluation. They were introduced to the dashboard's design, elements, rules, and colour semantics. Participants were asked to inspect the dashboard and interpret student progress in a simulated online class monitoring situation, guided by six tasks (Table 1). Tasks were classified using Bloom's taxonomy (Burns et al., 2020). Tasks 1, 2, and 5 targeted *recalling factual information*, while tasks 3, 4, and 6 targeted *understanding information*. Task durations were $M = 73.7$ (no guidance), $M = 63.0$ (VG1 guidance), and $M = 55.0$ (VG2 guidance). The study used a within-subject design. All participants inspected all dashboard versions and answered the same questions. To avoid memorization and order effects, different classroom data were used for each dashboard, and dashboard order was pseudo-randomly assigned using a 3×3 Latin square matrix design. The concurrent think-aloud protocol was used, with participants encouraged to verbalize their thoughts during inspection (Bowles, 2018). Silent participants were prompted to reflect. The study procedure is depicted in Figure 3.

Table 1. Study tasks categorized according to Bloom's taxonomy.

Bloom Category	Task	Description
Retrieval	T1	What group or groups have no inactive students?
Retrieval	T2	In what group or groups are students participating in the discussion equally?
Comprehension	T3	What group is the most engaged in the writing activity?
Comprehension	T4	Which group is the most engaged in the discussion?
Retrieval	T5	What group has Tutor 2 spent the most time in?
Comprehension	T6	In your opinion, what group or groups might have experienced issues with progressing in the learning activity?

II. VL assessment After the main study, participants were emailed a link to the VLAT test to evaluate their VL. The test was to be completed within 1 to 3 days to minimize bias, as VLAT could influence initial VL skills due to its extensive use of multiple charts. The test included 53 multiple-choice questions covering common chart types. We followed Lee and colleagues (2017)'s recommendations for conducting the test, including a 40-second time limit per question, with the option to skip unanswered questions. Participants were compensated $50 for the whole study.

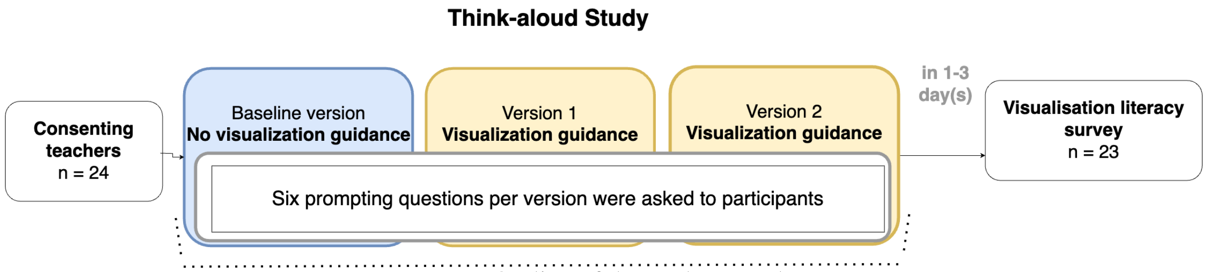

Figure 3. Outline of the study procedure.

3.6 Analysis

Instrument The audio recordings were transcribed and the transcripts were used to examine teachers' interpretation processes during tasks (McIntyre et al., 2022). The coding scheme comprised 14 codes across four dimensions:

1. *specificity of using visualization devices*, which identified the type and use of visualization affordances during interpretation;
2. *specificity of students' performance*, which noted the indicators that participants focused on during interpretation;
3. *depth of interpretation*, which assessed the quality of participants' evidence triangulation and argumentation regarding student or group behaviour[1]; and
4. *metacognitive reflections*, which captured reflections not directly related to the answer or interpretation, but to the general techniques or strategies used when interpreting data.

Dimensions 1 to 3 were derived from previous literature, such as the teacher noticing framework (Gamoran Sherin & van Es, 2009) (cited in van Leeuwen et al., 2019), while dimension 4, metacognitive reflections, was inspired by literature on VL (Burns et al., 2020) and strategies used when working with information visualizations (Germanakos et al., 2019). Codes were formulated inductively after the initial analysis of the transcribed data by one researcher. The coding scheme and frequencies of code instances in the transcript data are presented in Table 2.

Table 2. Coding scheme, code examples from the transcript data, and number of instances coded for each VL group (low and high VL). The "Code" column includes abbreviations in bold, which correspond to code names in the ENA visualizations.

Dimension	Code	Description	Protocol segment	Low VL	High VL
(1) Specificity of students' performance	Number of inactive or active students (**InactiveStudents**)	Mentioning a number of students that are active or inactive	"There are four students; they are not active at all."	43	49
	Teacher-student interaction (**TeacherStudent Int.**)	Mentioning situations in which students and teachers interacted or did not interact with each other	"Because even though there are two teachers, they have not talked to students . . . ."	58	66
	Student name (**StudentName**)	Mentioning a specific student name	"Holt might be dominating the conversation."	21	18
(2) Specificity of using visualization devices	Colour or emphasis (**Colour**)	Mentioning the colour of a visualization	"Because I can see the timeline of the activities is marked with blue colour."	22	14
	Visualization element (**VisElement**)	Mentioning a specific visual element of a visualization, e.g., a number from one of the visualizations or a specific element from a visualization, e.g., legend	"There is a strong thick line between the teacher and other students."	71	89
	Frame of Reference (FoR) comparing based on 1 vis (**SingleVis**)	Comparing students or groups over the same visualization. (1) There should be explicit mention of student groups (e.g., group 1 and group 2, student A and student B, etc.). (2) There should be no mention of more than a single visualization. If the participant does not mention any specific visualization, it is not included in the current code.	"Because compared with groups 3 and 4, they are way behind."	72	57
	Frame of reference (FoR) comparing based on multiple vis (**MultipleVis**)	Comparing the same student or group of students over multiple visualizations	"All groups have no inactive students [network]; all of them have inactive students in here but group 3 has very high engagement [bar chart]."	76	95

[1]Code D2 encompasses subcodes like D2.Intuition and denotes teachers reflecting on student progress based on assumptions and data, and D2.ActivityOutcome denotes teachers comparing students' progress to the expected learning activity outcome. Any instance coded as a subcode is also classified as D2.

(3) Depth	D1 (**Depth1**)	Describing students' behaviour only. These instances might stem from previous encounters with the same visualizations. No judgment should be present.	"I think it's group 3."	86	34
	D2 (**Depth2**)	Describing and judging students' behaviour, providing some potential explanation, such as comparing various groups and justifying their opinion (participants might refer to the learning progress of students by mentioning visualization elements, e.g., "because thickness of lines")	"They still have three people that are not very active, but at least they constantly discuss with each other."	116	133
	D2 intuition (**D2.Intuition**)	Arguing about the progress of some students, but rather based on assumptions and loosely situated in the data	"I think the answer to this question should be related to the time the teacher spent on the group. Because as teachers, we spend the most time in group 3. So I think we may be answering the questions from group 3. They might be facing some issues with the learning activity."	16	27
	D2 weighing against activity goal (**D2.ActivityOutcome**)	Arguing about the progress of some students stemming from weighing against the expected outcome of the learning activity. For example, it might be mentioned that students are not required to write in this activity.	"I think this activity does not require too much writing so I still think that is [group X]."	4	7
	D3 (**Depth3**)	Describing and judging students' behaviour, with a more developed argumentation and more than one potential explanation about the progress of a student or a group	"So engaged in discussion, in this case, is going to be group 3. They are spending most of the time doing discussions, at least around 22 minutes. They are all engaged with each other so they are either chatting or they are discussing with each other; we cannot tell unless I go to group 3 and discuss with them for 4 minutes so I assume they are okay."	14	31
(4) Metacognitive reflections	positive (**MetaPos**)	Reflecting on how a certain set of visualizations or visual elements helped them address a given task	"I think the answer to this question should be related to the time the teacher spent on the group."	36	63
	negative (**MetaNeg**)	Reflecting on how certain set of visualizations or visual elements did *not* help them address a given task	"But I couldn't count the number of students over here with the pie chart."	39	46

Transcript segmentation, i.e., dividing transcript structured parts based on the scope of the task performed by the participant, was done by one researcher. The typical unit of analysis is a short statement, i.e., "There are four students; they are not active at all," but many units include several sentences. Units of analysis were grouped within tasks since participants were required to interpret student data within the six given tasks.

A subsample of data (25% of tasks) was independently coded by two researchers. Each unit of analysis could be coded using more than one code. Cohen's k was computed for each code. We aimed for moderate agreement for each coding dimension, balancing a consistent coding scheme with the ability to capture diverse participant interpretations (N. McDonald et al., 2019). Disagreements were resolved through discussion and consensus (Campbell et al., 2013), leading to refined coding guide specifications. The consensus was achieved, yielding moderate reliability measures for each dimension: (1) specificity of using visualization devices ($k = 0.76$), (2) specificity of students' performance ($k = 0.95$), (3) depth ($k = 0.93$), and (4) metacognitive reflections ($k = 0.86$). Following this, one researcher coded the remaining data. Code examples for each dimension are in Table 2 (see also Section 4, below).

Coding think-alouds Think-alouds from 23 participants were recorded, transcribed, and coded using Nvivo. A total of 414 task responses were coded, resulting in 1403 coded references. Transcripts were segmented based on the scope of the task when using a dashboard.

VL test We used the median value of VL scores for all participants as a threshold to assign participants to low- or high-VL groups ($Mdn = 39$). This resulted in two groups of participants, where a total of 12 participants belonged to the low-VL group ($Q1 = 30$, $Mdn = 35$, $Q3 = 37$, $min = 17$, $max = 39$) and 11 participants were in the high-VL group ($Q1 = 43.5$, $Mdn = 45$, $Q3 = 46$, $min = 40$, $max = 48$).

To address RQs 1 and 2 (see details below), we employed a mixed-methods approach. Initial thematic coding was followed by ENA (Shaffer et al., 2009). ENA is a data analysis method that identifies and quantifies relationships in coded data, visualizing code relationships as networks. This method integrates qualitative and quantitative data analysis, enhancing understanding of think-aloud data and facilitating pattern identification using statistical techniques (Shaffer et al., 2009). ENA has been used in research exploring LA adoption (Tsai et al., 2021) and student resource use and project goal discussion (Bressler et al., 2022). ENA enables visual and statistical comparisons between conditions, making it useful for contrasting dashboard interpretations of teachers with different VL levels. It has been employed to compare visualization use by successful and unsuccessful students and performance differences in assignment completion (Elmoazen et al., 2022). Findings are illustrated with quotes in the results section, with each point represented by an appropriate code instance (Saldaña, 2015).

For both RQs 1 and 2, the total number of observations used is 414, where a single observation included a teacher's think-aloud transcript while they were using a dashboard to complete a task. In total, it equalled 23 teachers using three dashboards to complete six tasks. RQ3 focused on task completion accuracy; the number of observations used is 345. Task completion accuracy was computed as the number of correct responses to a study task. Since task 6 was an open question with no correct answer, the observations from this task were excluded.

RQ1 To address the research question "What are the differences in the ways teachers with varying VL skills interpret students' data presented via LA dashboards?", we used a modelling approach. We constructed ENA representations using the R package `enaR` (Marquart et al., 2017). The following ENA specifications were used:

$$\text{VL} > \text{participant} > \text{visualization guidance} > \text{task}. \tag{1}$$

In our study, a *unit of analysis* is a statement from a task using a dashboard, encapsulating the teacher's thought process. We calculated code co-occurrences within a task performed by a participant using a specific dashboard. Each task was assigned a unique ID for tracking and analysis. We used the entire task, not a moving window, for adjacency matrix computation. This was due to the brevity of our think-aloud responses, which rarely exceeded five sentences. This approach captured the full context of the teacher's thought process without information loss. Finally, data were aggregated using binary summation. This method assigns a value of one to codes that occur at least once during a task, regardless of frequency. Codes that did not occur are assigned a value of zero (Shaffer et al., 2009). This approach captures the presence of a code without overemphasizing frequent but non-insightful codes.

In our study, each *unit of analysis*—a single instance of a teacher's verbal response during a task using a dashboard—produced an ENA network. These networks were averaged across two groups, low-VL and high-VL participants, yielding two separate networks. We compared these networks using ENA's subtraction technique, which visualizes network differences. We employed a means rotation to maximize the variance between the two groups on the first latent dimension (MR1) (Shaffer et al., 2009), enabling a clear distinction between the low-VL and high-VL groups. The second dimension, SVD2, is derived from a singular value decomposition. These dimensions were used to map each unit of analysis and the corresponding codes and to draw connections based on code co-occurrences. The resulting model was visually represented as a graph, with nodes representing codes and edges representing connections between codes. Nodes are placed in X and Y coordinates mapped to MR1 and SVD2. We then "subtracted networks" to contrast the two network models, each representing a different VL group (Shaffer et al., 2009). This was achieved by subtracting the connection weights between nodes of one network from the other, resulting in a network graph that illustrates the differences. We then used linear mixed models (LMMs), modelling two dimensions, MR and SVD2, using the participants' VL as fixed effects, and included a random intercept for participants. We used the `lmerTest` R package to fit the model (Kuznetsova et al., 2017). We applied the mean rotation technique, which positions VL group means on the X axis, allowing a visual comparison between participants with different VL. Importantly, no differences are expected between participants' VL on Y axis coordinates (mapped to SVD2), meaning there will be no association between SVD2 and VL. The resulting equation to model MR1 and SVD2 was as follows:

$$
\begin{aligned}
&\mathrm{Y}_{ij} = \beta_0 + \beta_{\mathrm{VL}} \cdot \mathrm{VL}_{ij} + u_{0j} + \varepsilon_{ij},\\
&\mathrm{Y}_{ij} = \left[\mathrm{MR1}_{ij}, \mathrm{SVD2}_{ij}\right],\\
&u_{0i} \sim \mathcal{N}(0, \sigma^2_{\mathrm{prt}}).
\end{aligned}
$$

We also reported within-group (σ^2) and between-group ($t00$) variances; the intraclass correlation coefficient (ICC), which indicates the proportion of variances explained by the random effect structure; the marginal R^2, which considers only the variance of the fixed effects; and conditional R^2, which takes both the fixed and random effects into account. For reporting effect size, we used Cohen's d.

RQ2 To address this research question, "What are the differences in the ways teachers with varying VL skills interpret students' data presented via LA dashboards with varying visualization guidance levels given that teachers complete tasks of varying difficulty?", we used the same ENA representation as reported in RQ1 (see Equation 1). We answered this RQ in two steps. Firstly, we conducted an exploratory visual analysis comparing how teachers' depth of interpretation (represented by dimension MR1) varies throughout task completion depending on teachers' VL and the type of dashboard used. Secondly, we conducted a statistical analysis. For the exploratory analysis, we calculated the average values of the MR1 dimension for each task, based on the teachers' VL and the type of dashboard they used. This allowed us to visually contrast how teachers' depth of interpretation varies when they complete different tasks, accounting for visualization guidance used and teachers' VL (see Figure 5 in Section 4.2).

Following this, we conducted a statistical analysis. We added two additional variables to the modelling done as part of RQ1, namely, task category (see Table 1, Bloom Category) and the visualization guidance used as fixed effects. We followed a backward stepwise procedure for model selection based on the Akaike information criterion (AIC). The resulting equation to model MR1 and SVD2 is as follows:

$$\mathrm{Y}_{ij} = \beta_0 + \beta_{\mathrm{VL}} \cdot \mathrm{VL}_{ij} + \beta_{\mathrm{TaskLevel}} \cdot \mathrm{TaskLevel}_{ij} + \beta_{\mathrm{Vis.Guidance}} \cdot \mathrm{Vis.Guidance}_{ij} + u_{0j} + \varepsilon_{ij},$$
$$\mathrm{Y}_{ij} = \left[\mathrm{MR1}_{ij}, \mathrm{SVD2}_{ij}\right],$$
$$u_{0i} \sim \mathcal{N}(0, \sigma^2_{\mathrm{prt}}).$$

RQ3 The research question "To what extent are teachers' VL and visualization guidance associated with correct task completion?" was addressed using logistic regression. This model family was chosen to determine the relationship between VL, dashboard type (i.e., with or without VG), and the teachers' task completion accuracy. Task completion accuracy was computed as the number of correct responses to a study task. For this analysis, we excluded task 6, since it was an open question with no correct answer. A binomial generalized LMM with random effects was initially considered, akin to RQ1. However, due to a singular fit resulting from insufficient data variance, this approach was discarded. Instead, a generalized linear model was employed. We followed a stepwise backward elimination procedure for model selection. Model fit was evaluated using the coefficient of discrimination, Tjur's R^2. Reported effect sizes were calculated by converting the odds ratio to Cohen's d.

4. Results

4.1 RQ1: Differences between Teachers with Low and High VL

The resulting ENA model is depicted as a graph in Figure 4. Larger values on the X axis (9.6% variance explained) show a higher tendency for the participants to engage in (i) *deep* (code 3, "Depth2") and *profound* (code 3, "Depth3") interpretations and are detailed in terms of explicitly mentioning *visual elements from the visualizations* during the interpretation (code 2, "VisElement"). Lower values on the X axis indicate the participants' tendency to engage in *simpler interpretations* and to rely on *colour emphasis* (code 2, "Colour"). Larger values on the Y axis (5.9% of variance explained) indicate the participants' tendency to use *multiple visualizations* (code 2, "MultipleVis") during their interpretation, while low values indicate interpretations within a *single visualization* (code 2, "SingleVis").

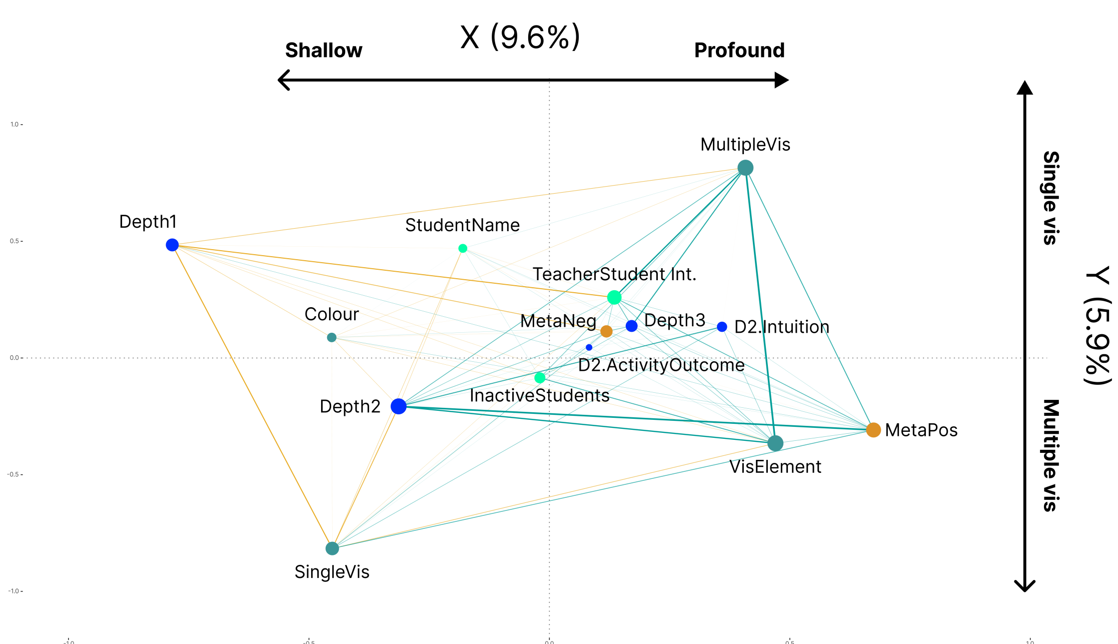

Figure 4. Subtracted ENA graph which depicts prominent connections between codes, specific for the participants with either low or high VL. The colour of the nodes represents the following: *specificity of visualization devices* (dimension 2) coloured as dark green, *specificity of students' performance* (dimension 1) as light green, *depth of the interpretation* (dimension 3) as blue, and *metacognitive reflections* (dimension 4) as orange. Edges coloured in orange and green indicate connections between codes, which are more prominent for the teachers with low VL and high VL, respectively. Code abbreviations are defined in Table 2. The thickness of the edges shows the strength of the connection between the codes.

The observed differences between participants with low and high VL are as follows:

Dimension 1: Specificity of students' performance Participants with varying VL levels commonly referenced specific student performance indicators (Table 2, dimension 1). They often cited teacher-student interaction, e.g., P1, "because we can see that it is teacher 2 who had conversations with other students," and the number of inactive or active students (code 1, "InactiveStudents") to support their interpretations, as shown by P1: "there are four students; they are not active at all." A notable difference in the specificity of low- and high-VL participants was the more frequent mention of specific student names by those with low VL (Figure 4 and Table 2, code 1, "StudentName").

Dimension 2: Specificity of using visualization devices Low-VL participants relied more on colour for interpretations than high-VL participants (Figure 4, code 2, "Colour"). They used colour to compare student groups and justify answers, as P19 illustrated: "Okay, that would be group 3, because again in the next visualization, group 3 has no *red* dots." Low-VL participants also tended to rely more on single visualizations for task answers, often mentioning specific student names during interpretations (code 1, "StudentName"), as shown by P11: "Kaden and teacher 2 have two-way communication, and then the teacher also communicates with Oscar and Ali. So, it means teacher 2 is communicating more compared to group 4. So, that's why I said that in group 3, teacher 2 is most engaged."

High-VL participants, however, were generally more specific when referring to visualizations (Figure 4, dimension 2, **specificity of vis devices**). They often relied on various visual elements for deep interpretation (code 3, "Depth2," and code 2, "VisElement"), as P17 demonstrated: "For example, 'the engaging with each other' visualization. Like the thickness of the edge. But it's normalized, so we can't really tell if they spend more time in discussion or not. That's a tricky one. Let me calculate the time. So, it's 4 minutes. So, let me ... 23 minus 4 is 19. And group 2 is at 20."

Dimensions 3 and 4: Interpretation depth and metacognitive reflections Participants engaged in shallow, deep, and profound interpretations (codes 3,"Depth1"; 3, "Depth2"; and 3, "Depth3"). *Low-VL participants* typically engaged in *shallow* interpretations, regardless of whether they used *single or multiple visualizations* (see connections in Figure 4, between codes 3, "Depth1"; 2, "SingleVis"; and 2, "MultipleVis"). During these interpretations, they indicated difficulty in definitively interpreting student data and answering questions, even when they identified the right set of visualizations, as illustrated by P11: "Then it should be group 3 or 4, but I can't say which one because both are yellow." Moreover, low-VL participants were more likely to be unsure or reach no conclusion about group performance, as illustrated by P3: "I think group 3. In terms of the network diagram, teacher 2 ... Well, in the network diagram, teacher 2 is not connecting with students with an arrow. Does that mean that teacher 2 didn't really ... It's hard to tell, but I can see group 2 completed two activities." However, when engaging in deep interpretations, they usually interpreted student data within a single visualization (codes 3, "Depth2," and 2, "SingleVis"). When low-VL participants could identify the right set of visualizations for tasks (code 4, "MetaPos"), they typically referred to visualization elements (code 2, "VisElement"), as illustrated by P19: "This visualization, Google Doc, is very clear because they outline the activity. So, you can see groups 1 and 2 start at activity 2." However, the vast majority of the co-occurrences between any metacognitive reflection were more prominent for participants with high VL, as can be seen from Figure 4, where most of the edges between codes 4, "MetaPos," and 4, "MetaNeg," are coloured in blue.

High-VL participants typically solved tasks by interpreting student data using multiple visualizations (code 2, "MultipleVis"), leading to a *deep interpretation* (code 3, "Depth2"). They demonstrated diversity in interpretation depth and enacted profound interpretations twice as often as low-VL participants (see Figure 4 and frequencies in Table 2, code 3, "Depth3"). No specific codes were exclusively linked to high-depth interpretations of high-VL participants. However, high-VL participants engaging in profound interpretation (code 3, "Depth3") frequently used multiple visualizations (code 2, "MultipleVis"), during which they suggested alternative explanations for the observed students' behaviour, as illustrated by P8: "Group 3 spent the most time discussing among each other. And, I would like to point out, that although they do spend time with the teacher, it could also mean that they have a lot of collaboration between each other and the teacher. So group 3 is the most engaged in discussion for this question. Contrasting them to group 4, they don't really discuss much because four of them are inactive. And they are also spending a lot more time engaging with the teacher." Moreover, high-VL participants were reflected in the fact of successfully identifying the right set of visualizations for the task (code 4, "MetaPos") when they used multiple visualizations.

In contrast, when using single visualizations (code 2, "SingleVis"), participants with high VL often relied on intuition (code 3, "D2.Intuition"). The use of single visualizations can limit interpretation, necessitating reliance on intuition, as P14 exemplified: "Group 4 is not participating in activity 1. Maybe it's just because they are not willing to participate in activity 1. So I don't think it's an issue. I think it's their attitude." They also interpreted student data against learning activity outcomes, as shown by connections between code 3, "D2.ActivityOutcome," and code 2, "SingleVis."

To summarize, *modelling results* confirm a significant difference in the *depth of interpretation* between participants with different VL (see Table 3, MR1 model, VLHigh). Participants with low VL generally had shallower interpretations of student data than those with high VL. There was no significant difference in the use of single and multiple visualizations based on participants' VL (see Table 3, SVD2 model, VLHigh; the full model table is available in Appendix A). This aligns with expectations as we used the means rotation technique, aligning group means on the Y axis (see Section 3.6).

Table 3. LMMs for MR1 and SVD2.

	MR1 ~ VL + (1|*prt*)				SVD2 ~ VL + (1|*prt*)			
Variables	Estimates	CI	p	eff.size	Estimates	CI	p	eff.size
(Intercept)	−0.06	[−0.11, −0.02]	*		0	[−0.05, 0.05]		
VLHigh	**0.13**	**[0.06, 0.19]**	***	**0.64†††**	0	[−0.07, 0.07]		0

Effect sizes: †: 0.10–0.3 (small effect), ††: 0.30–0.5 (moderate effect) and †††: ≥ 0.5 (large effect). [a]
*P-values *: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1* [*]

4.2 RQ2: Differences in Interpretation Depending on VL and Visualization Guidance

The results of an exploratory visual analysis are presented in Figure 5. These show the average depth of teachers' interpretation for each task, based on their VL and the type of dashboard used (no guidance—NG—and with visualization guidance—VG). This revealed that participants with both low and high VL engaged in a similar depth of interpretation when using the NG dashboard, with task 6 being an exception. However, when using dashboards with visualization guidance, participants adapted their depth of interpretation depending on the task, with larger shifts between horizontal lines. Specifically, for tasks T3, T4, and T6 (*comprehension category*), participants typically engaged in deep interpretations compared to tasks T1, T2, and T5 (*retrieval category*). This suggests that visualization guidance enabled participants to adjust their level of interpretation according to the task type.

However, our statistical analysis did not confirm this observation. The final model selected, based on the AIC, favoured a simpler model without a three-way interaction ($AIC_{VL+TaskLevel+Vis.Guidance} = -194$ and $AIC_{VL*TaskLevel*Vis.Guidance} = -184$, $x^2(df = 7) = 3.86$, $p = 0.8$). Similar to the results in RQ1, VL had a large effect size on interpretation depth, while tasks requiring comprehension skills had a medium effect size on the depth of interpretation compared to tasks requiring simple information retrieval skills (Table 4, MR1; the full model table is available in Appendix B). Yet, there was no significant main effect of visualization guidance on the interpretation depth. Lastly, in regards to effects of the use of single and multiple visualizations (Table 4, SVD2), none of the variables had a significant effect.

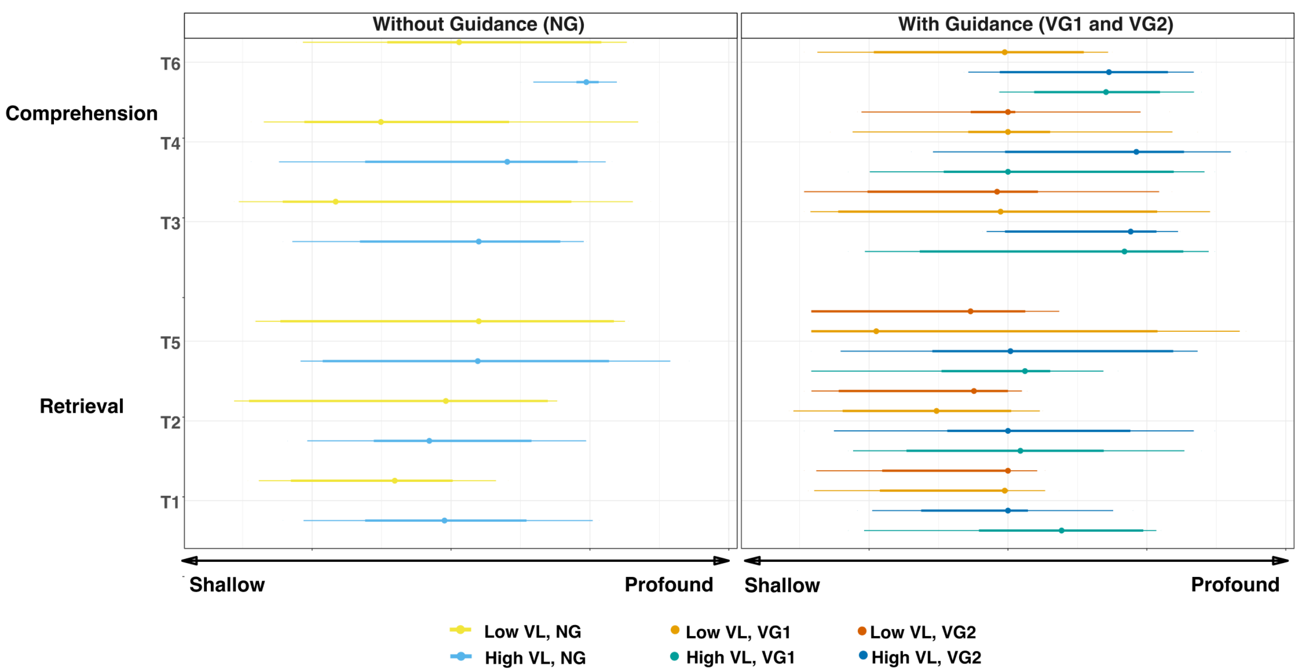

Figure 5. Teachers' interpretation depth using dashboards with and without guidance (NG and VG). The horizontal lines show the average depth, with the centre dot as the median and bold line segments for the first and third quartiles. Line colour represents the levels of VL. The *X* axis ranges from shallow (left, code 3, "Depth1") to profound (right, code 3, "Depth3"). The *Y* axis indicates tasks (T1–T5), with tasks 3, 4, and 6 focusing on advanced comprehension and tasks 1, 2, and 5 on basic information retrieval.

Table 4. LMMs for MR1 and SVD2.

	MR1 ~ VL + TaskLevel + Vis.Guidance + $(1\|prt)$				SVD2 ~ VL + TaskLevel + Vis.Guidance + $(1\|prt)$			
Variables	Estimates	CI	p	eff.size	Estimates	CI	p	eff.size
(Intercept)	-0.11	$[-0.16,-0.06]$	***		-0.02	$[-0.08,0.05]$		
VLHigh	**0.13**	**[0.06, 0.19]**	***	**0.66†††**	0	$[-0.07,0.07]$		0
Task Level	**0.09**	**[0.06, 0.13]**	***	**0.46††**	0	$[-0.05,0.05]$		0.02
Comprehension								
VG1	0	$[-0.04,0.05]$			0.04	$[-0.02,0.1]$		
VG2	0.01	$[-0.04,0.05]$			0	$[-0.06,0.07]$		

Effect sizes and significance intervals have the same intervals as in Table 3.

4.3 RQ3: Differences in Correct Task Completion Depending on VL and Visualization Guidance

We examined the association between VL and the type of visualization guidance used, and the odds ratio of correctly answering a task. Our findings showed no association between participants' VL and changes in the odds ratio, indicating equal chances of task completion for participants with both low and high VL. Significant differences were observed between a dashboard without visualization guidance (NG) and dashboard VG1, which included visualization guidance ($OR = 1.9$, $CI = [1.04, 3.52]$, $p = 0.039$, $d = 0.35$). No significant difference was found between the odds ratio of the NG dashboard and dashboard VG2 ($OR = 1.8$, $CI = [0.99, 3.32]$, $p = 0.55$). The predicted probabilities to correctly complete the task for a given VL and dashboard type are presented in Figure 6.

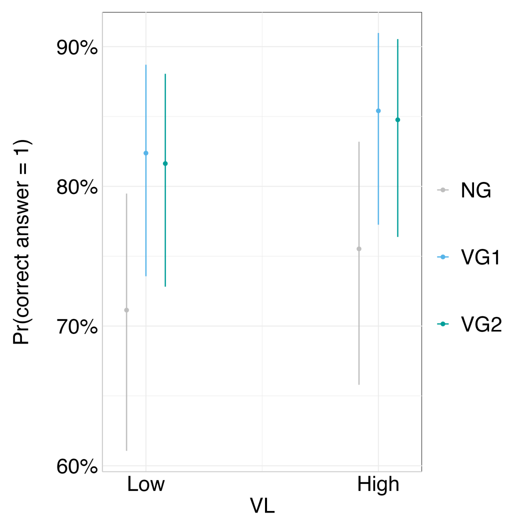

Figure 6. Predicted probabilities for participants to correctly complete the task given a particular combination of VL and dashboards with or without visualization guidance. There was a significant difference between the NG dashboard and dashboard VG1.

5. Discussion

5.1 Findings

RQ1. Interpretation depth gap between low- and high-VL teachers Our research reveals that teachers with both low VL and high VL referred to specific indicators of students' activity during the interpretation (*dimension 1*). The main difference, however, was that teachers with low VL had a higher tendency to refer to specific student names. While prior research emphasized that analytics users tended to use reference points to scaffold their interpretation (Wise & Jung, 2019; Li et al., 2021), our results might indicate that it is easier for teachers with low VL to refer to data points on visualizations mapped to individual students during the interpretation process. This result also complements the findings outlined by van Leeuwen et al. (2019), who found that teachers use more individual indicators of student activity when using dashboards with a low degree of visualization guidance. Our results indicate that this is associated with teachers' VL.

We found that teachers with low VL tend to interpret student data using a single visualization, especially for deep interpretations (*dimension 2*). Low-VL teachers used visualization elements in their interpretation when focusing on a single visualization. In contrast, high-VL teachers often incorporated insights from multiple visualizations for deep interpretations, aligning with Pozdniakov, Martinez-Maldonado, Tsai, Echeverria, and colleagues (2023)'s findings. This is consistent with literature reporting that experienced instructors are more inclined to use multiple visualizations, make complex comparisons, and draw more useful insights from dashboard-depicted student data (Li et al., 2021). This focused approach allows them to reach a relatively deep level of interpretation, primarily using colour emphasis to anchor their interpretation. Due to the limitations of the rotated ENA model, we could not verify whether the associations between teachers' VL and the use of single or multiple visualizations (*dimension 2*) were statistically significant.

Our findings show that differences in strategies between low- and high-VL teachers are associated with the interpretation depth (*dimension 3*). *Low-VL participants* typically engaged in *shallow* interpretations, while high-VL participants tended to engage in deep interpretations. Our findings add to the literature on teacher-facing dashboard use. Prior research highlighted variations in dashboard use by teachers but did not identify factors explaining these variations (Molenaar & Knoop-van Campen, 2019). We discovered that differences in interpretation depth can be linked to teachers' VL. Our modelling results indicated significant differences in interpretation depth between low- and high-VL groups.

Lastly, we found that high-VL teachers had a higher tendency to engage in metacognitive reflections than teachers with low VL (*dimension 4*). This finding emphasizes that teachers with high VL have a more pronounced understanding of the visualizations and could verbalize and identify whether their interpretation succeeded or not (Kirsh, 2014). This somewhat complements Auerbach and colleagues (2018)'s findings, suggesting that instructors with high teaching expertise had superior diagnostic capabilities and deeper reflections. Our findings emphasize the role that VL literacy could play in these differences.

RQ2. Differences between teachers' interpretations depending on their VL, tasks, and visualization guidance In contrast to van Leeuwen and colleagues (2019), who found that the depth of teachers' interpretation is contingent on visualization guidance, our findings indicated that the relationship between teachers' interpretation depth and visualization guidance is rather subtle. In general, the depth of the interpretation was mainly dependent on VL. However, the depth of the interpretation could also have varied according to the task difficulty for teachers with both low and high VL. Future research is required to confirm whether visualization guidance could assist teachers in better adapting the depth of their interpretations depending on the task difficulty.

RQ3. Association between teachers' VL, visualization guidance, and correct task completion We found a minor yet significant positive association between the use of a guided dashboard and the rate of correct task completion. This suggests that more explicit visualization guidance had a limited effect on task completion accuracy. Specifically, the inclusion of textual summaries in dashboard VG2 did not improve accuracy. This contrasts with literature suggesting that visualization guidance can improve task completion (Lalle et al., 2021; van Leeuwen & Rummel, 2022). A possible explanation is that as teachers familiarize themselves with a dashboard, the benefits of explicit visualization guidance decrease (Willett et al., 2007). We see the resemblance between this finding and the broader implications outlined by Nazaretsky and colleagues (2022), who suggested that explicit guidance is not necessarily beneficial for teachers' use of LA. Interestingly, we found no association between VL and task completion accuracy, contradicting theories that posit that VL enhances data interpretation skills, leading to correct answers (Lee et al., 2017; Börner et al., 2019). Controlling for time on task as a covariate in a regression model could yield more accurate estimates of the relationship between VL and task completion accuracy and yield different results for the association between VL and task completion accuracy.

5.2 Implications for Design

Our findings led to the following recommendations for LA designers:

- Visualization guidance can aid teachers' interpretation during difficult tasks. Designers should provide visualization guidance by (i) highlighting key visualization elements using heuristics and consistent colour, (ii) integrating visualizations with descriptive text, (iii) decluttering visuals by eliminating non-essential elements, and (iv) framing titles as potential questions teachers may have when using LA (Pozdniakov et al., 2022; Echeverria, Martinez-Maldonado, Granda, et al., 2018).

- Given the strong link between the use of multiple visualizations and deep interpretations, designers could build upon this finding. When context permits, designers could group several visualizations under one title, presenting diverse data for a more comprehensive view of group progression (Pozdniakov et al., 2022). Alternatively, these visualizations could be combined in different dashboard views (Martinez-Maldonado et al., 2020).

- To address the interpretation depth gap between low- and high-VL teachers, designers could offer explicit examples of correct data interpretation via dashboards. These could be presented in an on-demand pop-up window or a separate interface view, containing text and exemplary visualizations (Lalle et al., 2021). These examples, based on real-world scenarios, could be tailored to the specific data the teacher is viewing.

5.3 Implications for Research

Currently, there is emerging evidence regarding the benefits of visualization guidance in teacher-facing dashboards. Previous collective research efforts have shown promise in directing teachers' attention to crucial elements (Chen et al., 2019; Martinez-Maldonado et al., 2020; G. Fernandez-Nieto et al., 2022; Fernandez Nieto et al., 2022; G. M. Fernandez-Nieto et al., 2024) and assisting with teachers' interpretation (Echeverria, Martinez-Maldonado, Granda, et al., 2018; Echeverria et al., 2024). However, visualization guidance alone may not be sufficient to support teachers' interpretation. This is because the interpretation of visual data is a complex process that requires not only the ability to focus on key elements but also a certain level of skill to process information visualization. Therefore, future research in this area might need to explore whether the strategies to upskill teachers' VL could be more beneficial than providing visualization guidance only. Future research could also explore whether visualization guidance is more effective in scaffolding teachers' interpretation when using analytics presented in mediums different from the dashboards, e.g., interactive reports (Fernandez Nieto et al., 2022).

The present study focused on visualization guidance aimed to *support teachers' interpretation of students' data* with a set of well-formulated questions. Previous research on factors explaining the effectiveness of dashboards with varying visualization guidance found no association between teachers' characteristics and the interpretation of the learning situation depicted on the dashboards (van Leeuwen et al., 2021). However, a follow-up study by van Leeuwen and Rummel (2022) showed that factors such as time pressure in the real classroom affected the effectiveness of guidance, such that visualization guidance in the dashboards helped teachers detect groups requiring attention. The present study contributes to this emerging evidence of the effectiveness of teacher guidance by emphasizing the role of teachers' VL in detecting and interpreting students' progress depicted via dashboards with varying visualization guidance. The present study represents specific sensemaking conditions confined by a specific activity system and situation and operationalization of noticing, perceiving, and interpretation. Future research could investigate whether the gap in interpretation depth between teachers with varying VL persists in sensemaking settings different from the one described in the current study.

5.4 Limitations and Future Work

The present study has some limitations. The first limitation relates to the binary categorization of teachers' VL into low and high, which may oversimplify skill diversity and affect result generalizability. Future research could explore a more nuanced operationalization of VL to verify our results. A second limitation is in the use of concurrent think-aloud protocols to study teacher interpretation via dashboards. This method captures immediate thoughts but may not fully encapsulate the complexity of teachers' thought processes, especially if teachers are not familiar with this technique (S. McDonald et al., 2013). Future research could address this by combining think-aloud protocols with stimulated recall to capture teachers' hindsight rationalization. Recording participants' eye movements while using dashboards with and without visualization guidance, and then playing these recordings back with eye-tracking data overlaid on the dashboards, could deepen understanding of teachers' interpretation and the role of visualization guidance (McIntyre et al., 2022).

Regarding RQ1, a limitation is that our analysis only showed associations between dimensions of teachers' interpretation, such as specificity of referring to visualization elements and interpretation depth. We could not identify conditions for enabling deep interpretations for teachers with low VL. Future research could replicate our design and use methods like qualitative comparative analysis for causal conclusions (Hanckel et al., 2021). For RQ2, we used exploratory visual analysis and statistical analysis. Exploratory analysis suggested that visualization guidance could help teachers adapt their interpretations to task difficulty levels. However, we used a simple model without an interaction effect, which might not fully validate this finding. We could not include a model with the interaction effect between VL, visualization guidance, and task type as it did not fit the data better than a model without interaction. Future research could consider a three-way interaction effect using power analysis and collect sufficient data. More advanced statistical modelling could also be used. For RQ3, we did not have enough variance in our data to converge multilevel logistic regression, limiting our ability to use the modelling type used in RQs 1 and 2. This could result in less robust estimates of the effect of visualization guidance and VL on correct task completion. Future work could address this by conducting the study with a larger sample.

6. Concluding Remarks

Our findings reveal a disparity in the interpretation depth of student data on LA dashboards between teachers with low and high VL. Specifically, those with low VL generally exhibited a shallower interpretation depth. No association was found between visualization guidance and interpretation depth. Medium guidance was linked to improved task completion, but this link vanished with explicit guidance. Overall, this paper elucidates the relationship between VL, visualization guidance, and teachers' data interpretation on LA dashboards.

Acknowledgements

The authors thank Vitomir Kovanović for providing the source code for ENA computations and visualizations.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was at least in part funded by The Australian Research Council's Industrial Transformation Training Centre for Information Resilience (CIRES) (IC200100022).

References

Ahn, J., Campos, F., Hays, M., & DiGiacombo, D. (2019). Designing in context: Reaching beyond usability in learning analytics dashboard design. *Journal of Learning Analytics*, *6*(2), 70–85. https://doi.org/10.18608/jla.2019.62.5

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, *6*, 100215. https://doi.org/10.1016/j.caeai.2024.100215

Auerbach, A. J., Higgins, M., Brickman, P., & Andrews, T. C. (2018). Teacher knowledge for active-learning instruction: Expert–novice comparison reveals differences. *CBE—Life Sciences Education*, *17*. https://doi.org/10.1187/cbe.17-07-0149

Bao, H., Li, Y., Su, Y., Xing, S., Chen, N.-S., & Rosé, C. P. (2021). The effects of a learning analytics dashboard on teachers' diagnosis and intervention in computer-supported collaborative learning. *Technology, Pedagogy and Education*, *30*(2), 287–303. https://doi.org/10.1080/1475939x.2021.1902383

Barral, O., Lalle, S., Iranpour, A., & Conati, C. (2021). Effect of adaptive guidance and visualization literacy on gaze attentive behaviors and sequential patterns on magazine-style narrative visualizations. *ACM Transactions on Interactive Intelligent Systems*, *11*(3-4), 1–46. https://doi.org/10.1145/3447992

Barral, O., Lallé, S., Guz, G., Iranpour, A., & Conati, C. (2020). Eye-tracking to predict user cognitive abilities and performance for user-adaptive narrative visualizations. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (ICMI 2020), 25–29 October 2020, online (pp. 163–173). ACM. https://doi.org/10.1145/3382507.3418884

Bartindale, T., Chen, P., Marshall, H., Pozdniakov, S., & Richardson, D. (2021). ZoomSense: A scalable infrastructure for augmenting Zoom. In *Proceedings of the 29th ACM International Conference on Multimedia* (MM 2021), 20–24 October 2021, online (pp. 3771–3774). ACM. https://doi.org/10.1145/3474085.3478332

Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, *116*(6), 1857–1864. https://doi.org/10.1073/pnas.1807180116

Bowles, M. A. (2018). Introspective verbal reports: Think-alouds and stimulated recall. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 339–357). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-59900-1_16

Bressler, D. M., Annetta, L. A., Dunekack, A., Lamb, R. L., & Vallett, D. B. (2022). How STEM game design participants discuss their project goals and their success differently. In B. Wasson & S. Zörgő (Eds.), *Advances in quantitative ethnography. ICQE 2021. Communications in computer and information science* (pp. 176–190, Vol. 1522). Springer International. https://doi.org/10.1007/978-3-030-93859-8_12

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics*, *6*(2), 1–9. https://doi.org/10.18608/jla.2019.62.1

Burns, A., Xiong, C., Franconeri, S., Cairo, A., & Mahyar, N. (2020). How to evaluate data visualizations across different levels of understanding. In A. Bezerianos, K. Hall, S. Huron, M. Kay, M. Meyer, & M. Sedlmair (Eds.), *2020 IEEE Workshop on Evaluation and Beyond—Methodological Approaches to Visualization* (BELIV), 25–30 October 2020, Salt Lake City, Utah, USA (pp. 19–28). IEEE. https://doi.org/10.1109/beliv51497.2020.00010

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, *42*(3). https://doi.org/10.1177/0049124113500475

Campos, F., Ahn, J., DiGiacomo, D. K., Nguyen, H., & Hays, M. (2021). Making sense of sensemaking: Understanding how K–12 teachers and coaches react to visual analytics. *Journal of Learning Analytics*, *8*(3), 1–21. https://doi.org/10.18608/jla.2021.7113

Ceneda, D., Gschwandtner, T., May, T., Miksch, S., Schulz, H.-J., Streit, M., & Tominski, C. (2017). Characterizing guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 111–120. https://doi.org/10.1109/TVCG.2016.2598468

Chavan, P., & Mitra, R. (2022). Tcherly: A teacher-facing dashboard for online video lectures. *Journal of Learning Analytics*, *9*(3), 125–151. https://doi.org/10.18608/jla.2022.7555

Chen, Q., Li, Z., Pong, T.-C., & Qu, H. (2019, April). Designing narrative slideshows for learning analytics. In *Proceedings of the 2019 IEEE Pacific Visualization Symposium* (PacificVis 2019), 23–26 April 2019, Bangkok, Thailand (pp. 237–246). IEEE. https://doi.org/10.1109/pacificvis.2019.00036

Choudhry, A., Sharma, M., Chundury, P., Kapler, T., Gray, D. W. S., Ramakrishnan, N., & Elmqvist, N. (2021). Once upon a time in visualization: Understanding the use of textual narratives for causality. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), 1332–1342. https://doi.org/10.1109/tvcg.2020.3030358

Deakin Crick, R. E., Knight, S., & Barr, S. (2017). Towards analytics for wholistic school improvement: Hierarchical process modelling and evidence visualization. *Journal of Learning Analytics*, *4*(2), 160–188. https://doi.org/10.18608/jla.2017.42.13

Dickler, R., Gobert, J., & Sao Pedro, M. (2021). Using innovative methods to explore the potential of an alerting dashboard for science inquiry. *Journal of Learning Analytics*, *8*(2), 105–122. https://doi.org/10.18608/jla.2021.7153

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://doi.org/10.3991/ijet.v15i17.15041

Dourado, R. A., Rodrigues, R. L., Ferreira, N., Mello, R. F., Gomes, A. S., & Verbert, K. (2021). A teacher-facing learning analytics dashboard for process-oriented feedback in online learning. In *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK 2021), 12–16 April 2021, Irvine, California, USA (pp. 482–489). ACM. https://doi.org/10.1145/3448139.3448187

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 *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (LAK 2018), 7–9 March 2018, Sydney, New South Wales, Australia (pp. 131–140). ACM. https://doi.org/10.1145/3170358.3170380

Echeverria, V., Yan, L., Zhao, L., Abel, S., Alfredo, R., Dix, S., Jaggard, H., Wotherspoon, R., Osborne, A., Buckingham Shum, S., Gasevic, D., & Martinez-Maldonado, R. (2024). TeamSlides: A multimodal teamwork analytics dashboard for teacher-guided reflection in a physical learning space. In *Proceedings of the 14th International Conference on Learning Analytics and Knowledge* (LAK 2024), 18–22 March 2024, Kyoto, Japan (pp. 112–122). ACM. https://doi.org/10.1145/3636555.3636857

Elmoazen, R., Saqr, M., Tedre, M., & Hirsto, L. (2022). A systematic literature review of empirical research on epistemic network analysis in education. *IEEE Access*, *10*, 17330–17348. https://doi.org/10.1109/access.2022.3149812

Fernandez Nieto, G. M., Kitto, K., Buckingham Shum, S., & Martinez-Maldonado, R. (2022). Beyond the learning analytics dashboard: Alternative ways to communicate student data insights combining visualisation, narrative and storytelling. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 219–229). ACM. https://doi.org/10.1145/3506860.3506895

Fernandez-Nieto, G., An, P., Zhao, J., Buckingham Shum, S., & Martinez-Maldonado, R. (2022). Classroom dandelions: Visualising participant position, trajectory and body orientation augments teachers' sensemaking. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (CHI 2022), 29 April–5 May 2022, New Orleans, Louisiana, USA (pp. 1–17). ACM. https://doi.org/10.1145/3491102.3517736

Fernandez-Nieto, G. M., Martinez-Maldonado, R., Echeverria, V., Kitto, K., Gašević, D., & Buckingham Shum, S. (2024). Data storytelling editor: A teacher-centred tool for customising learning analytics dashboard narratives. In *Proceedings of the 14th International Conference on Learning Analytics and Knowledge* (LAK 2024), 18–22 March 2024, Kyoto, Japan (pp. 678–689). ACM. https://doi.org/10.1145/3636555.3636930

Gamoran Sherin, M., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, *60*(1), 20–37. https://doi.org/10.1177/0022487108328155

Germanakos, P., Kasinidou, M., Constantinides, M., & Samaras, G. (2019). A metacognitive perspective of InfoVis in education. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (UMAP'19 Adjunct), 9–12 June 2019, Larnaca, Cyprus (pp. 331–336). ACM. https://doi.org/10.1145/3314183.3323674

Hanckel, B., Petticrew, M., Thomas, J., & Green, J. (2021). The use of qualitative comparative analysis (QCA) to address causality in complex systems: A systematic review of research on public health interventions. *BMC Public Health*, *21*(1). https://doi.org/10.1186/s12889-021-10926-2

Hernández-Leo, D., Martinez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J. A., & Rodríguez-Triana, M. J. (2018). 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

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., & Hlosta, M. (2019). A large-scale implementation of predictive learning analytics in higher education: The teachers' role and perspective. *Educational Technology Research and Development*, *67*(5), 1273–1306. https://doi.org/10.1007/s11423-019-09685-0

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (LAK 2018), 7–9 March 2018, Sydney, New South Wales, Australia (pp. 31–40). ACM. https://doi.org/10.1145/3170358.3170421

Jørnø, R. L., & Gynther, K. (2018). What constitutes an "actionable insight" in learning analytics? *Journal of Learning Analytics*, *5*(3), 198–221. https://doi.org/10.18608/jla.2018.53.13

Kaliisa, R., Jivet, I., & Prinsloo, P. (2023). A checklist to guide the planning, designing, implementation, and evaluation of learning analytics dashboards. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00394-6

Kaliisa, R., Mørch, A. I., & Kluge, A. (2022). "My point of departure for analytics is extreme skepticism": Implications derived from an investigation of university teachers' learning analytics perspectives and design practices. *Technology, Knowledge and Learning*, *27*(2), 505–527. https://doi.org/10.1007/s10758-020-09488-w

Kasepalu, R., Chejara, P., Prieto, L. P., & Ley, T. (2023). Studying teacher withitness in the wild: Comparing a mirroring and an alerting and guiding dashboard for collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *18*(4), 575–606. https://doi.org/10.1007/s11412-023-09414-z

Khosravi, H., Shabaninejad, S., Bakharia, A., Sadiq, S., Indulska, M., & Gašević, D. (2021). Intelligent learning analytics dashboards: Automated drill-down recommendations to support teacher data exploration. *Journal of Learning Analytics*, *8*(3), 133–154. https://doi.org/10.18608/jla.2021.7279

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074. https://doi.org/10.1016/j.caeai.2022.100074

Kirsh, D. (2014). Metacognition, distributed cognition, and visual design. In P. Gardenfors & P. Johansson (Eds.), *Cognition, education, and communication technology* (pp. 147–179). Routledge.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lalle, S., Toker, D., & Conati, C. (2021). Gaze-driven adaptive interventions for magazine-style narrative visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *27*(6), 2941–2952. https://doi.org/10.1109/TVCG.2019.2958540

Lee, S., Kim, S.-H., & Kwon, B. C. (2017). 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 *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK 2021), 12–16 April 2021, Irvine, California, USA (pp. 344–353). ACM. https://doi.org/10.1145/3448139.3448172

Liu, Y., Pozdniakov, S., & Martinez-Maldonado, R. (2024). The effects of visualisation literacy and data storytelling dashboards on teachers' cognitive load. *Australasian Journal of Educational Technology*, *40*(1), 78–93. https://doi.org/10.14742/ajet.8988

Marquart, C. L., Swiecki, Z., Collier, W., Eagan, B., Woodward, R., & Shaffer, D. W. (2017). rENA: Epistemic network analysis. https://doi.org/10.32614/cran.package.rena

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* (CHI 2020), 25–30 April 2020, Honolulu, Hawaii, USA (pp. 1–15). ACM. https://doi.org/10.1145/3313831.3376148

Martinez-Maldonado, R., Echeverria, V., Fernandez-Nieto, G., Yan, L., Zhao, L., Alfredo, R., Li, X., Dix, S., Jaggard, H., Wotherspoon, R., Osborne, A., Buckingham Shum, S., & Gašević, D. (2023). Lessons learnt from a multimodal learning analytics deployment in-the-wild. *ACM Transactions on Computer-Human Interaction*, *31*(1), 1–41. https://doi.org/10.1145/3622784

McDonald, N., Schoenebeck, S., & Forte, A. (2019). Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–23. https://doi.org/10.1145/3359174

McDonald, S., Zhao, T., & Edwards, H. M. (2013). Dual verbal elicitation: The complementary use of concurrent and retrospective reporting within a usability test. *International Journal of Human-Computer Interaction*, *29*(10), 647–660. https://doi.org/10.1080/10447318.2012.758529

McIntyre, N. A., Draycott, B., & Wolff, C. E. (2022). Keeping track of expert teachers: Comparing the affordances of think-aloud elicited by two different video perspectives. *Learning and Instruction*, *80*, 101563. https://doi.org/10.1016/j.learninstruc.2021.101563

Milesi, M. E., & Martinez-Maldonado, R. (2024). Data storytelling in learning analytics? A qualitative investigation into educators' perceptions of benefits and risks. In *Proceedings of the 14th International Conference on Learning Analytics and Knowledge* (LAK 2024), 18–22 March 2024, Kyoto, Japan (pp. 167–177). ACM. https://doi.org/10.1145/3636555.3636865

Molenaar, I., & Knoop-van Campen, C. A. N. (2019). How teachers make dashboard information actionable. *IEEE Transactions on Learning Technologies*, *12*(3), 347–355. https://doi.org/10.1109/TLT.2018.2851585

Molenaar, I., Horvers, A., Dijkstra, R., & Baker, R. S. (2020). Personalized visualizations to promote young learners' SRL: The learning path app. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany. ACM. https://doi.org/10.1145/3375462.3375465

Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, *53*(4), 914–931. https://doi.org/10.1111/bjet.13232

Nazaretsky, T., Cukurova, M., & Alexandron, G. (2022). An instrument for measuring teachers' trust in AI-based educational technology. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 56–66). ACM. https://doi.org/10.1145/3506860.3506866

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). https://doi.org/10.1186/s41239-020-00201-6

Pandey, S., & Ottley, A. (2023). Mini-VLAT: A short and effective measure of visualization literacy. *Computer Graphics Forum*, *42*(3), 1–11. https://doi.org/10.1111/cgf.14809

Poquet, O. (2024). A shared lens around sensemaking in learning analytics: What activity theory, definition of a situation and affordances can offer. *British Journal of Educational Technology*, *55*(4), 1811–1831. https://doi.org/10.1111/bjet.13435

Pozdniakov, S., Martinez-Maldonado, R., Singh, S., Khosravi, H., & Gašević, D. (2023). Using learning analytics to support teachers. In B. du Boulay, A. Mitrovic, & K. Yacef (Eds.), *Handbook of artificial intelligence in education* (pp. 322–349). Edward Elgar Publishing. https://doi.org/10.4337/9781800375413.00026

Pozdniakov, S., Martinez-Maldonado, R., Tsai, Y.-S., Cukurova, M., Bartindale, T., Chen, P., Marshall, H., Richardson, D., & Gasevic, D. (2022). The question-driven dashboard: How can we design analytics interfaces aligned to teachers' inquiry? In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 175–185). ACM. 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 *Proceedings of the 13th International Conference on Learning Analytics and Knowledge* (LAK 2023), 13–17 March 2023, Arlington, Texas, USA (pp. 89–99). ACM. https://doi.org/10.1145/3576050.3576063

Pozdniakov, S., Martinez-Maldonado, R., Tsai, Y.-S., Srivastava, N., Liu, Y., & Gasevic, D. (2023). Single or multi-page learning analytics dashboards? Relationships between teachers' cognitive load and visualisation literacy. In O. iberg, I. Jivet, P. Muñoz-Merino, M. Perifanou, & T. Papathoma (Eds.), *Responsive and sustainable educational futures. EC-TEL 2023. Lecture notes in computer science* (pp. 339–355, Vol. 14200). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-42682-7_23

Rets, I., Herodotou, C., & Gillespie, A. (2023). Six practical recommendations enabling ethical use of predictive learning analytics in distance education. *Journal of Learning Analytics*, *10*(1), 149–167. https://doi.org/10.18608/jla.2023.7743

Sahin, M., & Ifenthaler, D. (2021). Visualizations and dashboards for learning analytics: A systematic literature review. In M. Sahin & D. Ifenthaler (Eds.), *Visualizations and Dashboards for Learning Analytics* (pp. 3–22). Springer International Publishing. https://doi.org/10.1007/978-3-030-81222-5_1

Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, *12*(8), 569. https://doi.org/10.3390/educsci12080569

Saldaña, J. (2015). *The coding manual for qualitative researchers*. SAGE.

Sampson, D. (2017). Teaching and learning analytics to support teacher inquiry. In *Proceedings of the 2017 IEEE Global Engineering Education Conference* (EDUCON 2017), 25–28 April 2017, Athens, Greece (pp. 1881–1882). IEEE. https://doi.org/10.1109/EDUCON.2017.7943109

Sarmiento, J. P., & Wise, A. F. (2022). Participatory and co-design of learning analytics: An initial review of the literature. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 535–541). ACM. https://doi.org/10.1145/3506860.3506910

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). 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

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media*, *1*(2), 33–53. https://doi.org/10.1162/ijlm.2009.0013

Sloan-Lynch, J., & Morse, R. (2024). Equity-forward learning analytics: Designing a dashboard to support marginalized student success. In *Proceedings of the 14th International Conference on Learning Analytics and Knowledge* (LAK 2024), 18–22 March 2024, Kyoto, Japan (pp. 1–11). ACM. https://doi.org/10.1145/3636555.3636844

Sousa, E. B. d., Alexandre, B., Ferreira Mello, R., Pontual Falcão, T., Vesin, B., & Gašević, D. (2021). Applications of learning analytics in high schools: A systematic literature review. *Frontiers in Artificial Intelligence*, *4*, 737891. https://doi.org/10.3389/frai.2021.737891

Tsai, Y.-S., Kovanović, V., & Gašević, D. (2021). Connecting the dots: An exploratory study on learning analytics adoption factors, experience, and priorities. *The Internet and Higher Education*, *50*, 100794. https://doi.org/10.1016/j.iheduc.2021.100794

Van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, *10*(4), 571–596. https://www.learntechlib.org/primary/p/9171/

van Leeuwen, A., van Wermeskerken, M., Erkens, G., & Rummel, N. (2017). Measuring teacher sense making strategies of learning analytics: A case study. *Learning: Research and Practice*, *3*(1), 42–58. https://doi.org/10.1080/23735082.2017.1284252

van Leeuwen, A. (2019). Teachers' perceptions of the usability of learning analytics reports in a flipped university course: When and how does information become actionable knowledge? *Educational Technology Research and Development*, *67*(5), 1043–1064. https://doi.org/10.1007/s11423-018-09639-y

van Leeuwen, A., Knoop-van Campen, C. A. N., 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 Leeuwen, A., & Rummel, N. (2022). The function of teacher dashboards depends on the amount of time pressure in the classroom situation: Results from teacher interviews and an experimental study. *Unterrichtswissenschaft*, *50*(4), 561–588. https://doi.org/10.1007/s42010-022-00156-9

van Leeuwen, A., Rummel, N., & van Gog, T. (2019). What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? *International Journal of Computer-Supported Collaborative Learning*, *14*(3), 261–289. https://doi.org/10.1007/s11412-019-09299-x

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 *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 35–40). ACM. https://doi.org/10.1145/3375462.3375504

Willett, W., Heer, J., & Agrawala, M. (2007). Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1129–1136. https://doi.org/10.1109/TVCG.2007.70589

Wise, A. F., & Jung, Y. (2019). Teaching with analytics: Towards a situated model of instructional decision-making. *Journal of Learning Analytics*, *6*(2), 53–69. https://doi.org/10.18608/jla.2019.62.4

Appendices

A: RQ1—LMMs for MR1 and SVD2— Full Model

Table 5. *LMMs for MR1 and SVD2.*

	MR1 ~ VL + (1|*prt*)				SVD2 ~ VL + (1|*prt*)			
Variables	Estimates	CI	p	eff.size	Estimates	CI	p	eff.size
(Intercept)	−0.06	[−0.11, −0.02]	*		0	[−0.05, 0.05]		
VLHigh	**0.13**	**[0.06, 0.19]**	***	**0.64†††**	0	[−0.07, 0.07]		0
Random Effects								
σ^2	0.04				0.07			
τ_{00prt}	0.04				0.07			
ICC	0.11				0.05			
N_{prt}	23				23			
Observations	414				414			
df	21				21			
Marginal R^2 / Conditional R^2	0.094 / 0.195				0 / 0.047			

Effect sizes: †: 0.10–0.3 (small effect), ††: 0.30–0.5 (moderate effect) and †††: ≥ 0.5 (large effect). [a]
*P-values *:* *** < 0.001, ** < 0.01, * < 0.05, . < 0.1 [*]

B: RQ2—LMMs for MR1 and SVD2—Full Model

Table 6. *LMMs for MR1 and SVD2.*

	MR1 ~ VL + TaskLevel + Vis.Guidance + (1|*prt*)				SVD2 ~ VL + TaskLevel + Vis.Guidance + (1|*prt*)			
Variables	Estimates	CI	p	eff.size	Estimates	CI	p	eff.size
(Intercept)	−0.11	[−0.16, −0.06]	***		−0.02	[−0.08, 0.05]		
VLHigh	**0.13**	**[0.06, 0.19]**	***	**0.66†††**	0	[−0.07, 0.07]		0
Task Level Comprehension	**0.09**	**[0.06, 0.13]**	***	**0.46††**	0	[−0.05, 0.05]		0.02
VG1	0	[−0.04, 0.05]			0.04	[−0.02, 0.1]		
VG2	0.01	[−0.04, 0.05]			0	[−0.06, 0.07]		
Random Effects								
σ^2	0.03				0.07			
τ_{00prt}	0.03				0.07			
ICC	0.12				0.05			
N_{prt}	23				23			
Observations	414				414			
df	43.9				68.1			
Marginal R^2 / Conditional R^2	0.14 / 0.243				0.004 / 0.05			

Effect sizes and significance intervals have the same intervals as in Table 3.