Text-Based Visualizations for Learning Analytics: Perspectives from STEM Education Researchers

Huyen N. Nguyen* Los Alamos National Laboratory
Kathleen A. Bowe† University of New Hampshire
Minh-Huyen Nguyen‡ Hanoi University of Science and Technology
Kit Thompson§ University of Washington Bothell
Caleb M. Trujillo¶ University of Washington Bothell

ABSTRACT

Visualizing student reflections collected throughout a course can help instructors track changes in student language and inform teaching decisions. Interpreting these changes, however, requires tools that connect overall patterns with individual responses, and how to design such visualizations for learning analytics remains an open question. To address this gap, we conducted a mixed-methods study with 10 STEM education researchers who offered *dual* perspectives as both instructors and qualitative researchers. We used two visualization platforms as probes to elicit expert perspectives: the Journal Data Dashboard, for analyzing formative assessments, and WordStream Maker, for authoring custom visualizations. Participants evaluated both platforms by analyzing weekly student journal responses from a data visualization course. Findings revealed shared prioritization of context and control, discipline-rooted differences in perceived usefulness, and, notably, *epistemological* disagreement about frequency as viable evidence in qualitative work. Some participants regarded frequency-based visualization as a productive entry point to qualitative analysis, while others argued that it conflicts with the values of qualitative data analysis and may obscure rare but critical responses. These findings support the argument that perceived ease of use and usefulness, context of use or role, and epistemic frames are relevant to users of qualitative data visualization and learning analytics tools. We discuss design implications for future tools to better integrate quantitative techniques with qualitative interpretation. Transcripts, tutorials, and all supplementary materials are available at https://osf.io/z2f8d.

Index Terms: User study, Learning analytics, Data visualization

1 INTRODUCTION

Data-driven learning analytics can reveal emerging patterns within a student cohort and provide concrete evidence for improving the learning environment. For example, tracking how students' language shifts over time can give instructors insight into where students are and how their thinking is developing. Much of the richest evidence for this purpose is in written form, such as reflections, open responses, and formative assessments. However, the qualitative and open-ended nature of this format makes it highly context-laden, requiring substantial time and effort to identify overarching trends without losing the nuances of individual contexts [45].

We use the term *qualitative learning analytics* to describe this kind of analysis. Situated within broader efforts to use qualitative studies to provide richer context for the visualization community [35], a growing body of work characterizes visualization for qualitative data exploration [9, 21, 37] and, in particular, for educational settings [22]. However, little of this work focuses specifically on qualitative learning analytics, which demands approaches that track the temporal evolution of student language while preserving the specific contexts in which that language appears.V

To design qualitative learning analytics tools, designers must account for how qualitative data are used in education and research. Visualizations should connect high-level trends with close reading, supporting instructors as they assess student work and provide feedback, and researchers as they analyze student writing and other textual data. Visualizing qualitative data may incur a risk in striping away important contexts. Aggregation and emphasis can conceal important information from a viewer, raising ethical concerns [13] and oversimplifying rich qualitative data [7, 2]. These concerns parallel critical approaches to information visualization that critique *whose* questions an encoding ultimately serves [24].

Natural language processing [52] and, more recently, assistive AI tools [19] are more recent approaches to support qualitative data visualization. As computational approaches they analyze, codify and display textual patterns. These approaches can scale and automate through pattern recognition. However, these approaches often run counter to the researcher-as-instrument view espoused by many qualitative researchers. Similarly, analyses of computationally-based text visualizations seldom assess trustworthiness or credibility [53, 31], two major values of qualitative researchers. Therefore, textual learning analytics tools are more likely to be accepted if they scale computational analysis while preserving the unique benefits of qualitative approaches.

Despite their demonstrated utility, we still lack a broad understanding of how well these tools fit educational qualitative workflows, especially for people who both design analyses and use them in their own teaching. **Instructors who are also qualitative education researchers** sit at this intersection: they bring disciplinary expectations about qualitative rigor while facing the practical constraints of classroom assessment. Their reflections on how well the tools align with these *dual* roles can inform better designs. Seeing this opportunity, we ask what these instructor–researchers can teach us about designing visualization tools for qualitative learning analytics. We pose three research questions, in order of increasing abstraction:

RQ1. How do these instructor-researchers evaluate the two platforms with respect to ease of use, usefulness, and alignment with their qualitative education research practices when analyzing student assessment data?

RQ2. How do these experts judge the appropriateness of frequency-based quantitative encoding for representing qualitative data?

RQ3. How might these judgments inform the design of visualizations for qualitative educational data?

In this study we investigate these questions through a mixed-methods user study with STEM education researchers who also teach courses. We originally designed a comparative evaluation of two qualitative learning analytics platforms, recognizing that they serve distinct primary purposes. The comparison, however, surfaced an unanticipated tension: beneath the numerical usability ratings lay a genuine disagreement about whether frequency-based

*e-mail: huyen@lanl.gov
†e-mail: kathleen.jeffery7@gmail.com
‡e-mail: huyennm@soict.hust.edu.vn
§e-mail: kitty.alice.t@gmail.com
¶e-mail: calebtru@uw.edu

encoding can adequately represent qualitative data [34] . Taken together, this study offers the following contributions:

- a mixed-methods evaluation, with STEM instructor–researchers, of two qualitative learning analytics platforms, using real-world student journals to assess ease of use, usefulness, and alignment with qualitative education research practices;
- an emergent characterization of epistemological stances among these participants regarding the use of frequency-based encodings for qualitative educational data; and
- design implications for visualization tools that seek to integrate quantitative techniques into the qualitative analysis of educational data.

2 Related Work

2.1 Bridging Visualization and Qualitative Research

Visualization and qualitative approaches connect in several directions in the service of research, evaluation, and education [35, 23, 21, 37, 22]. For instance, Grounded theory, a qualitative research method, has been used within visualization research [21, 29] and while developing text-analytic interfaces [9].

Similar qualitative methods are often used to evaluate visualizations. Input from feedback sessions often takes written form [49]. Text data is often collected as words directly through surveys, interviews, video transcripts, notes, or written feedback. Researchers code this text into themes that become concrete feedback and design implications, informing new tools and recommendations. However, qualitative reporting has tended to lean on quotes, tables, process models, and conceptual diagrams more than on data displays [50, 30]. These forms suggest an opportunity for innovation to support users in their qualitative analysis and reporting.

Quantitative ethnography is one of the most direct attempts to bridge this divide, developing statistical representations of coded discourse [43, 1]. Its network-based methods illustrate that quantification can move beyond counts to focus on relationships. Epistemic network analysis (ENA) and ordered network analysis (ONA) quantify and visualize patterns of association among qualitative codes as weighted network models, with ONA additionally modeling the order in which those connections occur [45].

In summary, qualitative approaches are increasingly found in visualization research and evaluation but reported with only a limited set of visual representations, and ongoing work with coded textual data further integrates these areas.

2.2 Temporal Text Visualization

We focus on text data with a time dimension, the case pertinent to qualitative learning analytics where an instructor tracks student writing over a semester. Foundational work established the streamgraph idiom for this task: ThemeRiver [27] visualized thematic change across a document collection over time, and later systems such as TIARA [54] and TextFlow [15] extended it to track the evolution, birth, and merging of topics. WordStream [16] refined this line of work by keeping the extracted keywords legible on screen, and the idiom has since been embedded in larger interactive dashboards [55], frequently for text analysis [38, 48, 17]. Application examples span domain-specific tasks: CovidStream [3] traces the evolution of emotions throughout the COVID-19 pandemic in Peru, and VisDmk [8] represents the evolving sentiment and pop-up commentary of a video. The two platforms we evaluate, the Journal Data Dashboard [40] and WordStream Maker [39], put this idiom in the service of education. Despite WordStream's growing adoption, it has rarely been subjected to rigorous evaluation with domain experts who would actually use it for qualitative analysis, a gap our study aims to address.

3 The Probes: Journal Data Dashboard and WordStream Maker

Qualitative learning analytics require visualizations that connect high-level trends with close reading, supporting both instructional feedback and in-depth research on student writing. WordStream [16] provides a temporal overview by displaying changes in term prevalence. It combines a word cloud [51] with the temporal layout of a streamgraph [26], keeping extracted terms legible along a timeline. Section 3.3 details the encoding.

We use two platforms built on this idiom as *probes* [28] to elicit expert perspectives on visualization for qualitative learning analytics: the Journal Data Dashboard [40] for analyzing formative assessments and WordStream Maker [39] for authoring custom visualizations. Both are open-source, browser-based applications. As summarized in Table 1, the probes differ in scope, data input, access to source text, and customization. Their distinct purposes make a direct feature-by-feature comparison *inherently asymmetric*. We use these contrasts to examine what expert users value, trust, and critique at different stages of an educational data visualization workflow.

3.1 Journal Data Dashboard

Journal Data Dashboard [40], hereafter **Dashboard**, is a multi-panel tool for exploring a fixed corpus of formative assessments. It links a WordStream to the weekly prompts that students answered and to their full responses. As shown Figure 1, selecting a word retrieves the responses that contain it. These links support *insight extraction* by allowing users to move from aggregate patterns to individual student writing.

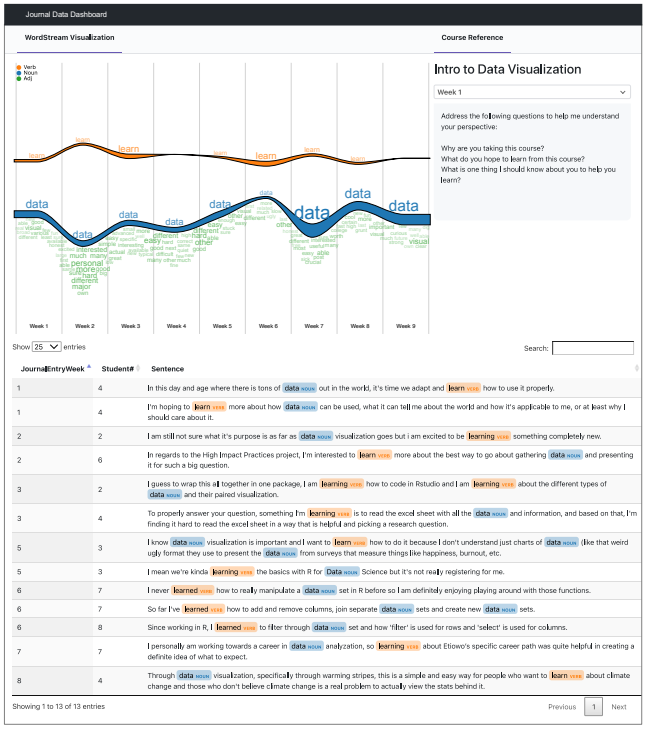


Figure 1: Journal Data Dashboard [40] user interface for analyzing formative assessments

The Dashboard is available as a web-based application at huyen-nguyen.github.io/journal-dashboard/wordstream, with an accompanying tutorial[1] and a demonstration video.[2]

[1] huyen-nguyen.github.io/journal-dashboard/tutorials
[2] huyen-nguyen.github.io/journal-dashboard/video

3.2 WordStream Maker

WordStream Maker [39], hereafter **Maker**, is a general-purpose, no-code tool for *visualization creation*, as shown in Panel B of Figure 2. Its graphical interface accepts user-provided tabular text data and combines data wrangling, natural language processing (NLP), and WordStream rendering in a single pipeline. Users can select part-of-speech categories and word-scoring methods. The resulting aggregate views, however, do not provide linked access to the source responses.

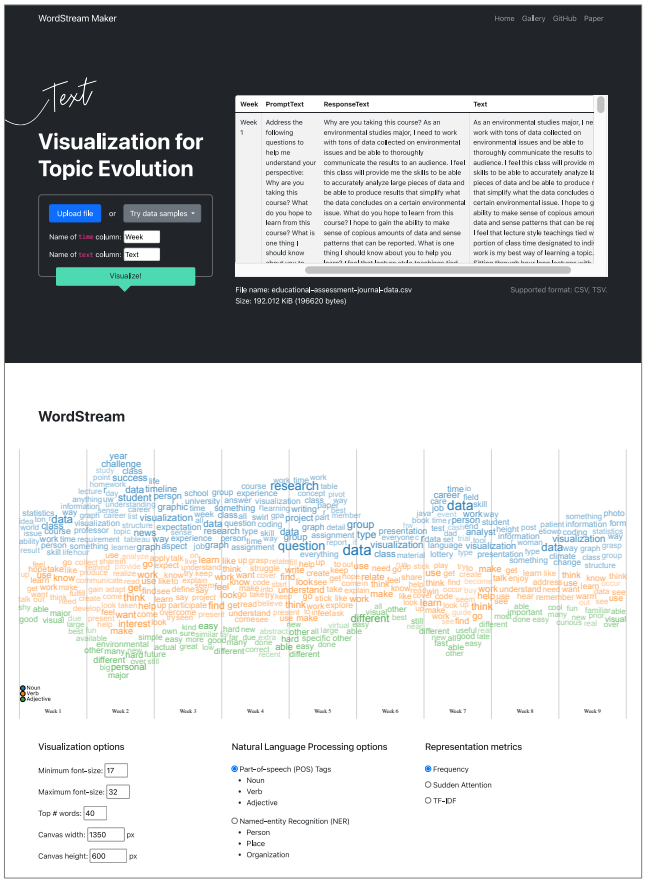


Figure 2: WordStream Maker [39], an end-to-end tool for authoring custom visualizations.

Maker includes built-in examples and is available online at `huyen-nguyen.github.io/maker`. In this study, participants used the "Education Assessment" sample, which contains the same student reflection corpus presented in the Dashboard.

Table 1: Key design contrasts between the two probes, Dashboard [40] and Maker [39]. Each tool supports a different activity within an analytical workflow.

	Dashboard [40]	**Maker [39]**
Scope	Domain-specific	General-purpose
Data	Fixed assessment corpus	User-uploaded at runtime
Linked text	Verbatim responses	None
Primary role	Insight extraction	Visualization creation
Customization	Limited, preset	User-selectable parameters

3.3 Shared Visualization Idiom: WordStream

Within the a WordStream [16] view, which are embedded in both Figures 1 and 2, words and phrases are sized by a computed score at each time point. Both probes use word frequency by default; Maker also offers sudden attention and TF–IDF as alternative scoring methods. The appropriateness of frequency-based encoding for qualitative educational data is the focus of this research study.

The shared idiom provides a common basis for exploring temporal patterns, while the platforms offer different ways to interpret those patterns. In particular, keeping extracted terms visible in WordStream differs from providing access to their full context: the Dashboard supplies that context through linked responses, whereas Maker presents an aggregate view.

4 Methods

4.1 Conceptual Framing

This work is guided by the Technology Acceptance Model (TAM) [18], which suggests that users are more likely to adopt a technology based on its perceived ease of use and perceived usefuless. Alongside TAM, we incorporate the concept of *context of use* [29, 46] to better understand the circumstances surrounding the use or potential use of such tools. Specifically, we frame the visualization platforms as technology probes [28] through which participants can reflect on existing workflows and possible changes to their practice. Together, these perspectives guide our use of both platforms to elicit experts' perception on their usefulness and ease of use, in relation to their current work contexts and potentially explore how the tools might support their practices.

4.2 Study Design

A mixed-methods design was used following an initial pilot study [14, 40]. We included two qualitative-quantitative cycles with confirmatory checking. This design offers flexibility and supports emergent findings, appropriate for the research pursuit. The study procedure overview is shown in Figure 3.

A quantitative component compared the two platforms, Dashboard and Maker, on perceived ease of use and usefulness. Two platforms serve distinct primary purposes, so we positioned them not as direct competitors but as probes into two stages of an analytical workflow: creating a visualization (Maker) from raw data and extracting insight from an interactive multi-view with linked text excerpts (Dashboard). The quantitative approach primarily addresses RQ1, providing an evaluation by expert users.

A concurrent qualitative component used thematic analysis to explore participants' experiences and contexts of use beyond what an agreement scale can capture. A qualitative approach addresses RQ1 and RQ2, providing judgment of appropriateness.

The quantitative and qualitative findings were integrated in a final interpretation research stage. This stage supports evidence integration, providing the basis of design recommendations, addressing RQ3.

4.3 Participants

We recruited ten STEM education researchers with expertise in both qualitative research methods and classroom assessment. These instructor-researchers were sampled purposefully for their dual expertise. We refer to participants as Experts 1–10; this numbering reflects our internal labeling and is not ordered meaningfully. The first analysis cycle included four participants (Experts 1, 7, 8, and 9), and the second cycle comprised all ten participants, with six newly recruited (Experts 2, 3, 4, 5, 6, and 10). Participants received a $100 (USD) gift card for their time.

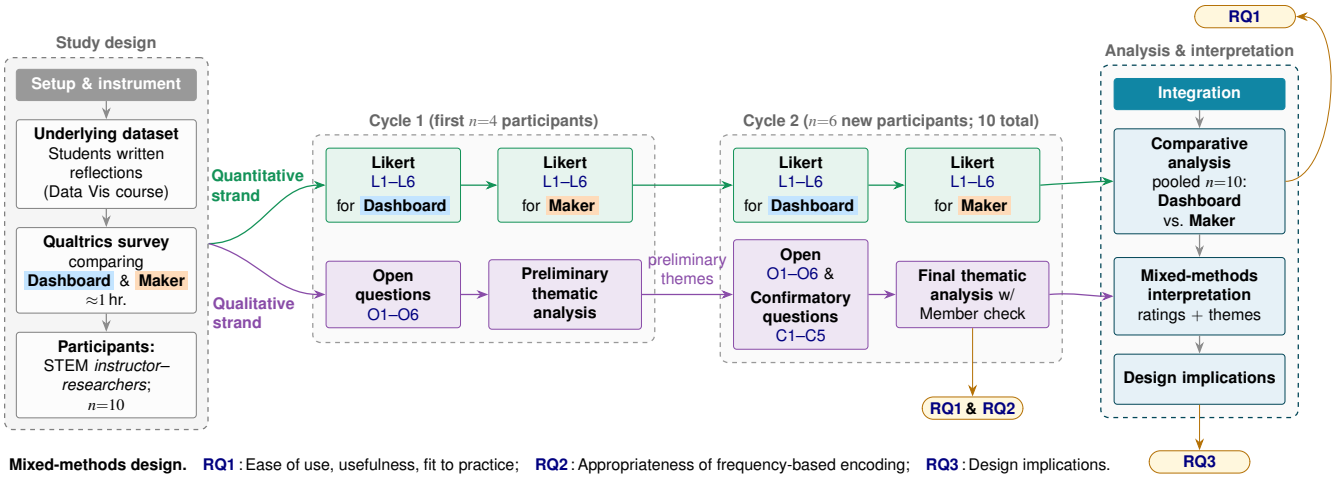


Figure 3: Study procedure with a two-phase mixed-methods design with ten STEM instructor-researchers. All participants worked on the same student reflection corpus within a survey, which included a quantitative strand (Likert ratings L1–7 for Dashboard & Maker tool) and a qualitative strand (open-ended responses O1–O6, plus confirmatory items C1–C5 in Cycle 2). Themes from Cycle 1 (n=4) were checked and refined in Cycle 2 (n=6), producing four themes (T1–T4) that were interpreted together with tool comparisons (n=10) to answer RQ1, RQ2, and RQ3.

4.4 Tasks and Apparatus

Each participant used both platforms, performed the assigned tasks, and completed a Qualtrics survey in approximately one hour. During the task, participants first explored the Dashboard and then explored the Maker, both of which accessed the same underlying "Education Assessment" sample dataset, consisting of 62 weekly reflections written by students in a data visualization course. Both the task tutorial and the complete instruments are provided in the Supplementary Material.

The survey questions were grounded in the TAM and context of use, measuring perceived usefulness and easy to use [18]. The survey included seven rating statements for each platform and six open-response questions, and for six participants in the second cycle, five additional confirmatory questions were included. All participants rated each platform on the same seven agreement statements (five-point Likert scale: *strongly disagree*, *disagree*, *neither agree nor disagree*, *agree*, *strongly agree*). The seven Likert-rating statements for evaluating both platforms were as follows:

- **L1 Details.** The app allows me to explore the details of the students' responses.
- **L2 Trends.** The app allows me to identify broad trends among the students' responses.
- **L3 Progress.** The app helps me draw conclusions related to student progress.
- **L4 Ease of use.** It is easy to use the app.
- **L5 Exploration.** The app makes it easier to explore student ideas.
- **L6 Development.** The app gives me insight into the development of student ideas over time.
- **L7 Diversity.** The app lets me see the diversity in students' responses and experiences.

The six open-response questions were as follows:

- **O1 Preference.** Which of the two tools did you prefer? Why?
- **O2 Trust.** Compare the data presented by both tools. To what extent do you trust the visualization(s) provided? Explain.
- **O3 Concerns.** What concerns do you have about how the visualizations represent the data? Explain.
- **O4 Workflow change.** Thinking about how you usually analyze and present qualitative data, how would these tools change the way you work, if at all?
- **O5 Desired features.** What features (current or additional) would make you more likely to use these tools? Explain.
- **O6 Other thoughts.** Please share any other thoughts you have about the tools.

The seven rating statements (L1–7) and the preference, trust, and workflow change questions (O1,O2,O4) target RQ1 to understand the participants' evaluation of the tools; the concerns and desired features questions (O3,O5) elicit their judgments about frequency-based encoding central to RQ2, relating the platforms to the representational format of qualitative data. In parallel, the workflow change and the desired features questions (O4,O5) motivate RQ3.

To test the preliminary themes emerging from Cycle 1, the six Cycle 2 participants additionally responded to five *confirmatory* statements, each rated on the same five-point scale and accompanied by a free-text explanation. The five confirmation themes were as follows:

- **C1 Rich context.** I valued functions that helped me explore the rich context of relationships in the qualitative data.
- **C2 Compare in context.** I valued functions for comparing and viewing data in context.
- **C3 More control.** More functions would allow me to increase my control to analyze data in context.
- **C4 Distrust under uncertainty.** I distrusted the tool when the functions evoked uncertainty.
- **C5 Specific contexts.** I considered specific contexts (e.g., teaching, assessment, research) when I evaluated the tools.

These confirmatory items functioned as a member-checking step, returning the candidate themes underlying RQ1 and RQ2 to participants for agreement and explanation. The complete instruments are provided in the Supplementary Material.

4.5 Quantitative Analysis

Likert responses were summarized per statement (L1–L7) and per tool (Dashboard vs. Maker) for visual analysis (Figure 4, Table 2); the small purposeful sample (n=10) precludes inferential statistics [10]. We report group-level means per statement and per tool, treating the ratings as ordinal indicators of perceived ease of use and perceived usefulness rather than as a basis for hypothesis testing. This component provides the comparative evaluation that answers RQ1.

4.6 Thematic Analysis

A thematic analysis was selected to conduct the concurrent qualitative study due to its ability to explore insightful patterns of sentiments and experiences not captured on an agreement scale. We performed a thematic analysis of written responses to the open-format questions following a six-step approach of coding data to develop themes iteratively over two cycles [5, 6] with the aid of Dedoose [20] (see Supplementary Material for details). To ensure the credibility of the interpretation, the last author (CMT) and a research student met weekly to discuss the coding process, emerging themes, and the meanings of the definitions. Before proceeding with preliminary findings, the first author (HNN) and second author (KAB) triangulated the analysis by verifying, modifying, or denying the defined themes with specific examples from the experts' writing.

After the second cycle, the six phases of thematic analysis were conducted again with the new participant data. During this iteration, three new team members, including the fourth author (KAT), were involved before seeking consensus with the full group. The second cycle led to the removal of one preliminary theme from consideration and the addition of a new theme, Theme 4, based on ample evidence (detailed in the Supplementary Material). The full team then debriefed about the themes and agreement-data visuals produced, providing additional insights, clarification, and limitations for the final themes, and integrating the two data sets for interpretation. As with other qualitative approaches, we did not intend to make generalizable claims; instead, we aimed to produce highly contextualized, generative knowledge from our sample in order to further the design and provide deep insights for future studies [25].

5 Results

5.1 Quantitative Findings on Perceived Ease of Use and Usefulness (RQ1)

Across all ten participants, the Dashboard was rated at or above the Maker on every Likert item. Table 2 reports per-item means derived directly from the study data, and Figure 4 shows detailed agreement ratings on each of the seven Likert items. Given the small sample n=10, we overlay each participant's individual rating (jittered to prevent overplotting) on the box plots, keeping the per-participant distribution as the primary basis for interpretation.

Table 2: Per-item mean ratings (1–5 scale) for the Dashboard and the Maker across all n=10 participants. The mean of Dashboard is at or above the Maker's on every item.

Item	Description	Dashboard	Maker	Δ
L1	Details	4.10	2.80	+1.30
L2	Trends	4.10	3.90	+0.20
L3	Progress	3.20	2.90	+0.30
L4	Ease of use	4.40	4.00	+0.40
L5	Idea exploration	4.00	3.30	+0.70
L6	Idea development	3.80	3.70	+0.10
L7	Diversity	4.00	3.70	+0.30
—	**Overall mean**	**3.94**	**3.47**	**+0.47**

The overall mean was 3.94 for the Dashboard and 3.47 for the Maker; expressed as a share of *agree* or *strongly agree* ratings, 77% of Dashboard ratings (54/70) were positive versus 60% for the Maker (42/70). The largest gap is on exploring details (L1, $\Delta = 1.30$) and almost no difference on broad trends (L2, $\Delta = 0.20$) and development over time (L6, $\Delta = 0.10$). Both platforms received their highest ratings on ease of use (L4: Dashboard 4.40, Maker 4.00) and their lowest on student progress (L3: Dashboard 3.20, Maker 2.90). Variance across participants was substantial on several items, reflecting the diversity of disciplinary stances discussed in the following sections.

At the per-participant level (Figure 5), the same pattern holds when ratings are collapsed across items: nine of the ten participants gave the Dashboard a mean agreement rating at or above their Maker rating, yielding the predominantly downward-sloping connecting lines, with Expert 3 the sole exception (red line). The descriptive medians of the per-participant means are 4.07 for the Dashboard and 3.57 for the Maker, and the Maker's distribution extends further toward the lower end of the scale, where two participants sit near or below the midpoint.

Given the purposeful but small sample, we read these as descriptive indicators rather than evidence of a reliable effect. In the following section, tool preference (O1) reflects the ratings: seven participants preferred the Dashboard, two preferred the Maker (Experts 3 and 4, both citing its side-by-side, simultaneous temporal display), and one (Expert 8) was mixed, valuing the Maker for quantitative trend-spotting while preferring the Dashboard's in-context word search.

5.2 Themes 1–3: Tool Experience and Disciplinary Context of Use (RQ1)

Three themes characterize how participants experienced the tools and situated them in their own practice. The five confirmatory items (C1–C5), answered by the six Cycle 2 participants, served as a member check on these themes; agreement was strong for the items underlying Themes 1–3 (C1,C2,C3,C5). We report Theme 4 in a dedicated Section 5.3 to answer RQ2.

Theme 1: Enhancing navigation and exploration of qualitative patterns across scales of context. Users valued features that enhanced ease of use by making it easier to compare and examine qualitative data in context at different scales of analysis. Specifically, they appreciated being able to search responses for words and phrases through filters, view trends and patterns through visualizations, and, for some, associate those words and phrases with quantitative trends. Expert 1 wrote:

> I preferred [Dashboard] because it provided both a traditional wordcloud in addition to the weekly visualization. I also like the feature of being able to **select words in combination** with one another to view the responses where they were used together. (Cycle 1, O1)

Expert 7 similarly valued the navigability from trend to text:

> The [Dashboard] is easier to navigate, I can easily see the **pattern of words per questions**, I can navigate from the summary to the wordstream and see changes in trends. (Cycle 1, O1)

Their feedback focused on the specific functionality, interactions, and visual elements of each platform's existing design. The member check reinforced this; explaining strong agreement with C2, Expert 6 (Cycle 2) noted, "Context gives a great deal of meaning to these topics/phrases." Across responses, a common thread was the desire for autonomy where participants wanted to explore the dataset on their own terms using a variety of representations.

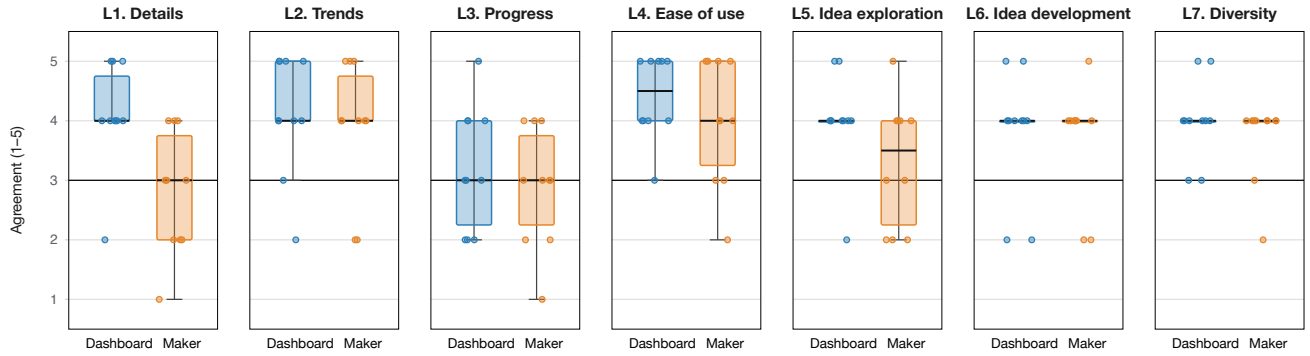


Figure 4: Agreement ratings (1–5, 5 = strongly agree) for the Dashboard (blue) and Maker (orange) on seven Likert items (L1–L7), across all ten participants. Boxes show the median and interquartile range; points represent individual participants with horizontal jitter to reduce overplotting. Given the small sample (n = 10), individual data points are shown on each box plot, and quartiles are descriptive rather than inferential.

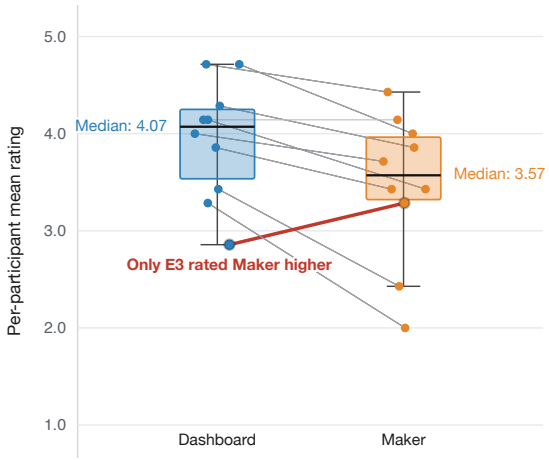


Figure 5: Per-participant mean agreement ratings (**averaged across the 7 Likert items**) for the Dashboard (blue) and Maker (orange). Each gray line represents a participant, linking their paired scores. Most lines slope downward, showing a general preference for the Dashboard; the red line marks the only exception (E3). Similar to Figure 4, with n = 10 the boxplots only show descriptive medians rather than inferential.

This autonomy extended to the unit of analysis itself. Participants wanted to recompose the same corpus around units other than the whole class. Expert 1 wanted "to select the responses of only one student over time" (Cycle 1, O5); Expert 3 wanted students' responses "grouped by question" (Cycle 2, O3); Expert 2 wanted "a way to look at how many students are saying particular phrases (not just how common the words are overall)" (Cycle 2, O5); and Expert 10 wanted to track "multiple words within the same category" together (Cycle 2, O5).

Ultimately, the perceived ease of searching, comparing, organizing, and otherwise manipulating the visualization to make sense of patterns was the primary basis on which users valued the WordStream platforms' features.

Theme 2: Extending analytic workflows through user control. Users evaluated the tools' usefulness largely through the features they wanted added that would extend the visualizations into a fuller analytic pipeline, from data ingestion and pre-processing through coding, quantification, and integration with established software. Several wanted the tools to ingest their own data. Expert 10 wrote:

> The functionality where I can **upload my own data** into this dashboard interface would be the icing on the cake for rapidly understanding my data in the early stages. (Cycle 2, O1)

Others asked for preprocessing capabilities that would clean inputs before visualization; Expert 1 (Cycle 1, O5) wanted "the ability to filter out some words," noting that prompt words tend to repeat across student responses and obscure the substantive vocabulary worth examining.

A second strand of suggestions pushed the tools downstream, toward existing analytic practice. Expert 2 saw the visualizations as a starting point for qualitative coding:

> These tools would be an excellent way to quickly look for common codes for further analysis... helpful for open coding and for looking for trends in how particular topics are discussed over time. (Cycle 2, O4)

Expert 5 asked for direct interoperability with established software:

> I'd almost want it to merge with a tool like SPSS text analysis — so analyze the text, tag it, and display the frequency of those tags. (Cycle 2, O5)

Expert 10 likewise wanted quantitative readouts attached to the qualitative view, suggesting instance counts for each noun, verb, and adjective; and Expert 1 called for the platforms themselves to be consolidated, asking that "the customization of Maker [be] embedded within Journal [Dashboard]."

The two tools' strengths were complementary: the Dashboard led precisely on the in-context items the Maker lacks (details L1, exploration L5), while the Maker offered the customization the Dashboard lacks. Participants asked for the combination, as in Expert 1's request to consolidate the platforms and Expert 10's request to upload their own data into the Dashboard.

Across these suggestions, users considered the tools as candidate components of their existing analytic practice rather than as a standalone artifacts. Usefulness, in this case, was a function of how readily the visualizations could be folded into the ingestion, preprocessing, coding, quantification, and reporting routines users already had.

In the member check, participants framed this desire as one of more control. This was evident in ratings of C3, which drew agreement or strong agreement from five of six Cycle 2 participants. From written responses, 'more control' was specified as control over the analytic workflow surrounding the tools rather than the

exploration within the visualization that was noted in Theme 1.

Theme 3: Evaluating features through a deep context-of-use perspective across disciplinary domains. Participants evaluated the tools within their specific disciplinary activities: teaching, assessment, and research, demonstrated through their comments closely linked to these disciplinary uses. They heavily reference their practices in examining classroom assessments or conducting qualitative research when evaluating the tools, as shown in an excerpt from Expert 6:

> There is certainly a need for tools like these. However, the demands placed on a tool meant for qualitative data analysis are complex and vary greatly from project to project. I'm inclined to think increasing user customization is going to be key for tools like this. (Cycle 2, O6)

This comment reflects familiarity with qualitative research practice and recognition that the highly variable demands placed on technology to accommodate diverse methods pose major limitations for any visualization tool.

Our participants shared the strengths and weaknesses of the platforms related to their various roles and uses. For example, when asked to comment on an initial draft of this theme, Expert 5 made the **dual role** of educator and researcher explicit:

> I always have this in mind - would this help me in teaching (e.g., give me fast feedback on my students' thinking), in research ([...] looking at students' self-reported study skills - this type of tool could be really helpful here, giving us some initial ideas of study approaches students are using). (Cycle 2, C5)

Envisioned research uses clustered around early-stage analysis; Expert 4 (Cycle 2, O4) saw a fit for "a great addition to initial qualitative coding to get an idea of what responses contain. Could be useful in building a codebook." Confirmatory item C5 drew agreement or strong agreement from all six Cycle 2 participants, each of whom named teaching, assessment, or research as the lens through which they judged the tools.

5.3 Theme T4: An Epistemological Dissensus over Frequency-Based Encoding (RQ2)

This Section is dedicated to **Theme T4**, which emerged only in Cycle 2 and prompted the addition of RQ2. **Theme T4** was a genuine disagreement about whether a frequency-based encoding can carry qualitative meaning. Participants ranged from skepticism to enthusiasm, with several occupying a pragmatic stance in the middle. At the more skeptical end, Expert 8 elaborated on shifting toward purpose-dependent uses:

> Word frequency is not a qualitative research tool, but can direct researchers to transcripts or videos to help identify patterns. (Cycle 1, O3)

They later add:

> I'm struggling to see how the frequency of a word can be used for qualitative data analysis where the researcher wants to know how or why, or even a common theme. (O6)

Expert 5 (Cycle 2, O3) made the loss-of-context argument concretely: "Word count/frequency doesn't actually tell us a lot. Words in context have a lot more meaning." Expert 2 named the long-tail risk:

> While I think the tool would be very helpful in looking for general trends, sometimes **the most important factors are those that are not as common.** (Cycle 2, O3)

Expressing a similar concern on rare but critical responses, Expert 3 wrote:

> I also worry about losing critical responses that are less frequent [...] I might expect to see a diversity of responses, and I would wonder about the bias introduced in this methodology. (Cycle 2, O4)

Alongside the Diversity item (L7), both tools scored only moderately, which shows a limitation in how they represent data and affects how fairly we can analyze student voices.

A pragmatic middle treated frequency as helpful but partial. Expert 7 (Cycle 1, O2) observed: "Frequency of words can be helpful, but in my experience, short phrases can be more informative to capture students' ideas." Expert 10 located the tension in disciplinary reception rather than in the encoding itself:

> I would love for us to be operating in a world where visualizations in combination with student response data is sufficient for our colleagues to accept the significance of our findings, however, many still want a number. (Cycle 2, O3)

Even skeptical viewpoints found value in frequency as a *locator* that routes attention back to close reading. Expert 8, quoted above, saw frequency as a way to direct researchers to the underlying transcripts, and described it as "locator tools to direct me to where I will apply other methodologies" (Cycle 1, O4). Experts 2 and 4 likewise saw a role in early open coding and codebook building (Section 5.2).

At the enthusiastic end, Expert 9 stated:

> I think you all have just revolutionized the future of qualitative coding, thematic analysis, and just how qualitative data is viewed and shared in general! I wish I had access to this tool when I was a graduate student! (Cycle 1, O6)

Questions also examined trustworthiness in the visual encodings. Trustworthiness is a value of qualitative research methods. For the Wordstreams, user's trust depended on being able to see how encodings were derived and how they connected to the source text. We initially identified this as a preliminary theme of uncertainty. For instance, Expert 1 did not know the inclusion cutoffs or "what sudden attention or TF-IDF mean" (Cycle 1, O3). Similarly, Expert 2 found it "unclear to me how the visualization chooses which words to include or exclude" (Cycle 2, O3). They also requested features to "customize or modify the topic/phrasing list" (Cycle 2, O5), suggesting that distrust, transparency and control were linked in participants' minds.

Notably, the member check did *not* resolve the skepticism-enthusiasm into a single stance related to trust. Confirmatory item C4 (distrust when functions evoke uncertainty) was the weakest-endorsed of the five: among the six Cycle 2 respondents, only one agreed, one disagreed, and four were neutral, with several reframing what "distrust" meant. This weak endorsement explains why a preliminary 'uncertainty/confidence' theme was not fully explanatory. Participants redefined the concept of 'distrust' instead of confirming it, which highlighted the deeper disagreement reported in Theme 4. Expert 6 **tied trust to traceability**, "Without seeing clear connections to the response text, I had a hard time trusting the visualizations" (Cycle 2, C4), whereas Expert 3 located the issue **not in uncertainty but in frequency bias itself** (Cycle 2, C4). The dissensus is therefore a meaningful disagreement about *what the encoding is—and should be—for*, instead of a simple deviation from consensus. In other words, the visualization had hidden operations that evoked users' uncertainty and did not align with the values they associated with qualitative data.

6 Discussion

Integration of findings. STEM instructor–researchers assessed education-facing tools through their own qualitative practice. The tools functioned as probes to elicit feedback about functionality and ideas about their deeper epistemological assumptions underlying the visual encoding. We found the Dashboard rates higher because it preserves in-context detail (L1, L5), addressing RQ1. The two tools are comparable for trend- and development-level reading (L2, L6). However, the substantive finding (RQ2) is that perceived value, and therefore the likelihood of technology acceptance, was dependent on participants underlying epistemic values about the qualitative-quantitative approaches. In other words, some participant accepted a frequency-based view as a legitimate qualitative instrument. Others remained skeptical, considering them, at most, an entry point into qualitative analysis practices.

6.1 Findings upon the dissensus

The disagreement over frequency encoding's value (Theme 4) did not resolve, and we read this as informative to our pursuit rather than as a failure to reach consensus. It resonates with long-standing concerns about what aggregation and emphasis can hide [13] and about whose questions an encoding serves [24]. The dissensus here may be due to competing role demands due to **our participants' multiple roles and perspectives**: as instructors making classroom decisions and as researchers committed to particular methodological traditions. Participants' comments often followed their research identity: a self-described grounded theorist valued the Dashboard's transparency (Expert 10, C1), whereas a scholar attentive to underrepresented voices raised concerns about frequency bias on equity grounds (Expert 3, O4). They rated the distrust item (C4) as the weakest-endorsed and several participants re-framed what "distrust" meant, relating it to uncertainty in the representation.

6.2 Implications

Extensive research has established previously TAM as a reliable model to explain the adoption and actual use of educational technology [18, 42, 41]. However, the basic TAM model fell short of explaining the expert user's judgments of appropriateness for qualitative learning analytics. A conceptual model of TAM may need to consider not only the context of use associated with participants' envisioned **role and task**, but also the **epistemic frames** that users bring. An updated model to design visual platforms better is shown in Figure 6).

Incorporating user roles and epistemic frames could help designers to consider the functions, methods, and values that are deeply debated and grounded in the qualitative research traditions. To distinguish roles, designers, evaluators and researchers could be define users within an institutional context, their professional identity, the functions they perform, or their credentials [42].

However, investigating researcher perspectives and epistemic beliefs about qualitative data may require greater effort. This is because each method aligns with a sets of paradigms: positivism, post-positivism, critical theory, and constructivism [25]. These paradigms have come from different traditions and have deeply embedded assumptions. To add to the challenge, the distinctions among qualitative research paradigms may be implicit, contradictory, partially formed or still evolving through multidecadal debates [34, 25, 33].

Alternatively to understand epistemic frames, researchers and designers may benefit from learning general practices that span multiple traditions of the qualitative research. These values include trustworthiness, authenticity, credibility, dependability, confirmability, plausibility, and transferability [32, 47]. These values are evident in long standing qualitative reporting norms as well (e.g., thick descriptions, data crystallization, and multivocality) [47]. Thus, tools and functions that support in support users aligned with these values may be perceived as more valuable for qualitative data exploration than frequency-centered perspectives. We therefore aim to design with these components in mind, motivate our design recommendations.

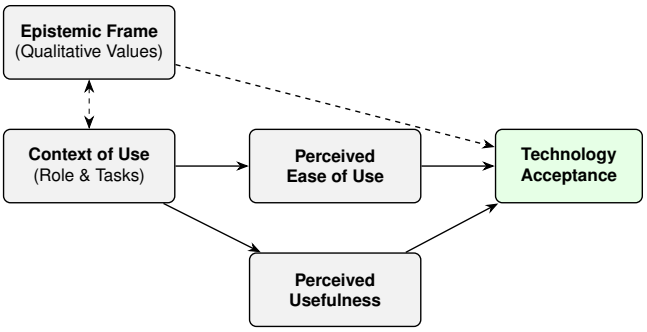


Figure 6: Proposed extension of the Technology Acceptance Model (TAM) for qualitative learning analytics. User's context of use (e.g., instructor, researcher) and epistemic frame (e.g., assumptions about valid qualitative practices) may affect the perceived ease of use and usefulness. Solid arrows indicate established research relationships. Dashed arrows represent novel hypotheses.

6.3 Design Recommendations for Qualitative Learning Analytics

Consideration of the integrated results informs design implications for tools that bring quantitative techniques into qualitative learning analytics (RQ3). Below, we relate our recommendations to prior methodological arguments from qualitative research and text-analysis systems.

6.3.1 Enable custom data uploads with insight extraction

A unified tool should let users import their own corpus, adjust the Maker's display parameters while preserving the Dashboard's context retrieval path from any stream element back to the verbatim responses. This design suggestion aligns with previous work in supporting users to transition between aggregate and context-specific views [44]. The priority of user control by increasing interactivity and access to detail comes with tradeoffs to representation, dashboard structure, and page layout simplicity [4].

6.3.2 Make encoding choices transparent & controllable

Tools should (1) provide plain-language annotations explaining all thresholds and scoring methods, (2) allow users to adjust these thresholds and methods, and (3) offer an editable list of topics and phrases, with stop words removed automatically by default but adjustable by the user. These suggestions align with Chuang et al. [11], who recommend exposing adjustable parameters and enabling progressive disclosure back to source data to make text visualizations interpretable and trustworthy. Our findings extend these principles to qualitative educational analysis by showing how educators valued such controls.

6.3.3 Preserve rare but critical responses

A frequency encoding surfaces the common and can bury the infrequent, which participants flagged as a risk to validity. Tools intended for student data should offer complementary views that surface the long tail, e.g., a low-frequency or outlier panel, or a per-response distribution, so that diversity of voice is not lost to aggregation. Such features would support qualitative research values of credibility, trustworthiness, and confirmability [32] and support critical research traditions [34, 25, 33]. For instance in a thematic analysis, this design would realizes Braun and Clarke's distinction

between thematic prevalence and analytic significance [5, 6] where a theme can be important to the research question without being frequent. We translate such methodological concerns from participants and the literature into a visualization requirement to keep infrequent responses accessible alongside dominant patterns. This argument supports why rare responses matter; however, it does not validate a particular solution such as an outlier panel.

6.3.4 Support flexible units of analysis

Participants wanted to recompose the same corpus around units other than the whole class. Tools should let users pivot the corpus by student, by prompt, and by distinct-speaker count. Functions that support these groupings—e.g., by overlaying several terms on one timeline—would let one tool serve the several disciplinary uses named under Theme 3. In a recent work, ThemeClouds [12] developed participant-level weighting in qualitative interview visualizations. Our participants additionally requested regrouping by student, prompt, and conceptual category. Such methods could be directly relevant to grounded theory methods and support values of data triangulation, transparency, and trustworthiness [21, 34]. Furthermore, the suggestion by Expert 5 to merge with pre-existing text analysis tools suggest that mixed-methods research may also benefit from a flexible unit of analysis.

6.3.5 Use frequency as an entry point instead of endpoint

We recommend designing the encoding explicitly as a bridge: the first step of a pipeline that provides one-click navigation from a stream element to its matching, relevant excerpts, plus export of selected words, phrases, and responses into coding workflows. This positioning is consistent with McNaught and Lam's [36] characterization of word clouds as supplementary tools for preliminary analysis. As noted by Expert 1, 2, 7, and 10, designs that support early stages to get familiar with a data set would be extremely valuable and are often a first step for qualitative analyses and grounded theory [6, 23]. The tool can then align the encoding with how these researchers actually work, complementing existing qualitative analysis practices. This would help address the strongest objections raised in Theme 4.

6.4 Limitations

The sample is small (n=10), chosen for dual expertise in qualitative methods and educational assessment, so the ratings are descriptive indicators reported without inferential tests [10]. This instrument is grounded in the constructs of technology acceptance but reliability and validity measures are needed to scale findings and applications of the assessment. Instead, we attempted triangulated findings through the open-ended and confirmatory items. Our claims are therefore interpretive, not generalizable. This also resonates with Losev et al. [35]; scalability is not usually a goal in qualitative studies. Instead, we strive for transferability, so this model and design suggestions can be explored in other setting. The question designs should be revisited, recognizing that participants may envision the application in several roles as uncovered in our analysis of instructor-researchers.

Limitations are also related to the WordStream platforms. We selected Dashboard and Maker due to our historic familiarity. The two tools are deliberately designed to be asymmetric, expanding the initial purpose-built design of the Dashboard into lightweight, web-based platform Maker. We used the asymmetry as probes under the assumption that it would give us insights into the stages of a workflow, not as matched competitors, so the comparison reflects context of use as much as the artifacts themselves. Readers should be cautioned from generalizing since findings and recommendations come from a user study of two specialized platforms.

All participants analyzed the same single corpus from one Data Visualization course; several anticipated that working with their own data would change their judgments. Finally, the data are self-reported responses, but do not investigate the fidelity between participant's stated and enacted practices. Given the variety of roles and qualitative methods, future work should consider case-specific uses, comparisons, and personas.

7 Conclusion

We set out to learn what qualitative education researchers can teach visualization designers, using two WordStream-based platforms: the Journal Data Dashboard and WordStream Maker, as probes into creating and reading visualizations of student journal writing for learning analytics. Ten STEM instructor-researchers rated the in-context Dashboard higher on perceived ease of use and usefulness, especially for exploring response detail, and were largely comparable on trend- and development-level reading. Alongside those ratings, our thematic analysis surfaced a sustained valuing of control and context (Themes 1–3) and, most importantly, an unresolved epistemological dissensus over whether a frequency-based encoding can meaningfully represent qualitative data (Theme 4). These findings support an argument that (1) perceived ease of use, (2) perceived usefulness, (3) context of use or role, and (4) epistemic frames are relevant to users of qualitative data visualization and learning analytic tools.

We translated these findings into five design implications: (1) unify creation and extraction; (2) make encoding transparent and controllable; (3) preserve rare but critical responses; (4) support flexible units of analysis; and (5) treat frequency as an entry point rather than an endpoint. These principles are an essential step to integrate quantitative technique with qualitative learning analytics inquiry. We hope these recommendations inform tools that let instructors and researchers move flexibly between the shape of a cohort's language and the individual voices that make it up.

Acknowledgments

This work was supported in part by the National Aeronautics and Space Administration (NASA) and the Gordon Research Conferences under NASA Award #17-TWSC17-0055. This work was also supported in part by the National Nuclear Security Administration (NNSA) Advanced Simulation and Computing (ASC) Program under grant number DE-SCL0000089. This work is released under LA-UR-26-29665.

The authors would like to thank Riyansha Goyal (RG), Jaime Serrano (JS), and Jayrylle Jaylo (JJ) for their support of this study. Funding support for research students came from the Scholarship, Research, and Creative Practice (SRCP) Seed Award and School of Interdisciplinary Arts and Sciences at the University of Washington Bothell.

References

[1] C. M. Abramson and D. Dohan. Beyond text: Using arrays to represent and analyze ethnographic data. *Sociological methodology*, 45(1):272–319, 2015. 2

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*, vol. 4. Springer, 2011. 1

[3] H. A. H. Baca, F. d. Luz Palomino Valdivia, Y. P. Atencio, M. J. Ibarra, M. A. Cruz, and M. E. H. Baca. Covidstream: Interactive visualization of emotions evolution associated with covid-19. In *Annual International Conference on Information Management and Big Data*, pp. 540–551. Springer, 2020. 2

[4] B. Bach, E. Freeman, A. Abdul-Rahman, C. Turkay, S. Khan, Y. Fan, and M. Chen. Dashboard design patterns. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):342–352, 2023. doi: 10.1109/TVCG.2022.3209448 8

[5] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006. 5, 9

[6] V. Braun and V. Clarke. Toward good practice in thematic analysis: Avoiding common problems and be (com) ing a knowing researcher. *International journal of transgender health*, 24(1):1–6, 2023. 5, 9

[7] K. Börner, A. Bueckle, and M. Ginda. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, 116(6):1857–1864, 2019. 1

[8] S. Cao, D. Guo, L. Cao, S. Li, J. Nie, A. K. Singh, and H. Lv. Visdmk: visual analysis of massive emotional danmaku in online videos. *The Visual Computer*, pp. 1–18, 2022. 2

[9] S. Chandrasegaran, S. K. Badam, L. Kisselburgh, K. Ramani, and N. Elmqvist. Integrating visual analytics support for grounded theory practice in qualitative text analysis. *Computer Graphics Forum*, 36(3):201–212, 2017. doi: 10.1111/cgf.13180 1, 2

[10] L.-T. Chen and L. Liu. Methods to analyze likert-type data in educational technology research. *Journal of Educational Technology Development and Exchange (JETDE)*, 13(2):3, 2021. 5, 9

[11] J. Chuang, D. Ramage, C. Manning, and J. Heer. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 443–452, 2012. 8

[12] J. T. Colonel and B. Lin. Word clouds as common voices: Llm-assisted visualization of participant-weighted themes in qualitative interviews. In *Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+ NLP)*, pp. 169–177, 2025. 9

[13] M. Correll. Ethical dimensions of visualization research. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–13, 2019. 1, 8

[14] J. W. Creswell and V. L. P. Clark. *Designing and conducting mixed methods research*. Sage publications, 2017. 3

[15] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. Textflow: Towards better understanding of evolving topics in text. *IEEE transactions on visualization and computer graphics*, 17(12):2412–2421, 2011. 2

[16] T. Dang, H. N. Nguyen, and V. Pham. WordStream: Interactive Visualization for Topic Evolution. In J. Johansson, F. Sadlo, and G. E. Marai, eds., *EuroVis 2019 - Short Papers*. The Eurographics Association, 2019. 2, 3

[17] T. Dang, V. Pham, H. N. Nguyen, and N. V. Nguyen. Agasedviz: visualizing groundwater availability of ogallala aquifer, usa. *Environmental Earth Sciences*, 79(5):1–12, 2020. 2

[18] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pp. 319–340, 1989. 3, 4, 8

[19] R. M. Davison, H. Chughtai, P. Nielsen, M. Marabelli, F. Iannacci, M. van Offenbeek, M. Tarafdar, M. Trenz, A. A. Techatassanasoontorn, A. Díaz Andrade, and N. Panteli. The ethics of using generative ai for qualitative data analysis. *Information Systems Journal*, n/a(n/a):1—-35, 2024. doi: 10.1111/isj.12504 1

[20] L. Dedoose. Dedoose version 9.2.007 cloud application for managing, analyzing, and presenting qualitative and mixed method research data. *Los Angeles, CA: SocioCultural Research Consultants, LLC*, 2021. 5

[21] A. Diehl, A. Abdul-Rahman, B. Bach, M. El-Assady, M. Kraus, R. S. Laramee, D. A. Keim, and M. Chen. An analysis of the interplay and mutual benefits of grounded theory and visualization. *IEEE Transactions on Visualization and Computer Graphics*, 31(9):5462–5479, 2025. doi: 10.1109/TVCG.2024.3452985 1, 2, 9

[22] A. Diehl, E. E. Firat, T. Torsney-Weir, A. Abdul-Rahman, B. Bach, R. Laramee, R. Pajarola, and M. Chen. Visguided: A community-driven approach for education in visualization. In *Eurographics*. The Eurographics Association, 2021. 1, 2

[23] A. Diehl, M. Kraus, A. Abdul-Rahman, M. El-Assady, B. Bach, R. S. Laramee, D. Keim, and M. Chen. Studying visualization guidelines according to grounded theory. *arXiv preprint arXiv:2010.09040*, 2020. 2, 9

[24] M. Dörk, P. Feng, C. Collins, and S. Carpendale. Critical infovis: exploring the politics of visualization. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, p. 2189–2198. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2468356.2468739 1, 8

[25] E. G. Guba and Y. S. Lincoln. Paradigmatic controversies, contradictions, and emerging confluences. *The Sage handbook of qualitative research*, 2005. 5, 8

[26] R. L. Harris. *Information graphics: A comprehensive illustrated reference*. Oxford University Press, USA, 1999. 2

[27] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: Visualizing thematic changes in large document collections. *IEEE transactions on visualization and computer graphics*, 8(1):9–20, 2002. 2

[28] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, et al. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 17–24, 2003. 2, 3

[29] P. Isenberg, T. Zuk, C. Collins, and S. Carpendale. Grounded evaluation of information visualizations. In *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, pp. 1–8, 2008. 2, 3

[30] J. R. Jaylo, M. Chastain, A. Nemec, C. S. Ouch, Y. Asefa, M. Li, A. Ung, and C. M. Trujillo. Visualization use in qualitative research reports: Evolving media types and competing epistemologies. *arXiv preprint arXiv:2603.27849*, 2026. 2

[31] K. Kucher, N. Sultanum, A. Daza, V. Simaki, M. Skeppstedt, B. Plank, J.-D. Fekete, and N. Mahyar. An interdisciplinary perspective on evaluation and experimental design for visual text analytics: Position paper. In *IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*, pp. 28–37, 2022. 1

[32] Y. S. Lincoln. Emerging criteria for quality in qualitative and interpretive research. *Qualitative inquiry*, 1(3):275–289, 1995. 8

[33] Y. S. Lincoln, S. A. Lynham, E. G. Guba, et al. Paradigmatic controversies, contradictions, and emerging confluences, revisited. *The Sage handbook of qualitative research*, 4(2):97–128, 2011. 8

[34] Y. Liu. Paradigmatic compatibility matters: A critical review of qualitative-quantitative debate in mixed methods research. *Sage Open*, 12(1):21582440221079922, 2022. doi: 10.1177/21582440221079922 2, 8, 9

[35] T. Losev, J. Raynor, S. Carpendale, and M. Tory. Embracing disciplinary diversity in visualization. *IEEE Computer Graphics and Applications*, 42(6):64–71, 2022. 1, 2, 9

[36] C. McNaught and P. Lam. Using wordle as a supplementary research tool. *Qualitative Report*, 15(3):630–643, 2010. 9

[37] H. Meng, Q. Lyu, P. Qin, Y. Yang, R. Zhang, W.-C. Lin, and Y.-C. Lee. Designing computational tools for exploring causal relationships in qualitative data. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2026. 1, 2

[38] H. N. Nguyen and T. Dang. Eqsa: Earthquake situational analytics from social media. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 142–143, 2019. 2

[39] H. N. Nguyen, T. Dang, and K. A. Bowe. WordStream Maker: A Lightweight End-to-end Visualization Platform for Qualitative Time-series Data. *IEEE VIS 2022 NLVIZ Workshop: Exploring Research Opportunities for Natural Language, Text, and Data Visualization*, 2022. 2, 3

[40] H. N. Nguyen, C. M. Trujillo, K. Wee, and K. A. Bowe. Interactive qualitative data visualization for educational assessment. In *Proceedings of the 12th International Conference on Advances in Information Technology*, pp. 1–9, 2021. 2, 3

[41] C. Or et al. Thirty-five years of the technology acceptance model: insights from meta-analytic structural equation modelling. *The Open/Technology in Education, Society, and Scholarship Association Journal*, 4(3):1–26, 2024. 8

[42] R. Scherer, F. Siddiq, and J. Tondeur. The technology acceptance model (tam): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, 128:13–35, 2019. doi: 10.1016/j.compedu.2018.09.009 8

[43] D. W. Shaffer. Epistemic network analysis: Understanding learning by using big data for thick description. In *International handbook of the learning sciences*, pp. 520–531. Routledge, 2018. 2

[44] S. Sinclair and G. Rockwell. *Text Analysis and Visualization*, chap. 19,

pp. 274–290. John Wiley & Sons, Ltd, 2015. doi: 10.1002/9781118680605.ch19 8

[45] Y. Tan, Z. Swiecki, A. R. Ruis, and D. Shaffer. Epistemic network analysis and ordered network analysis in learning analytics. In *Learning analytics methods and tutorials: A practical guide using R*, pp. 569–636. Springer, 2024. 1, 2

[46] M. Tory. User studies in visualization: A reflection on methods. *Handbook of Human Centric Visualization*, pp. 411–426, 2014. 3

[47] S. J. Tracy and M. M. Hinrichs. Big tent criteria for qualitative quality. *The international encyclopedia of communication research methods*, pp. 1–10, 2017. 8

[48] H. Van, H. N. Nguyen, R. Hewett, and T. Dang. Hackernets: Visualizing media conversations on internet of things, big data, and cybersecurity. In *2019 IEEE International Conference on Big Data (Big Data)*, pp. 3293–3302. IEEE, 2019. 2

[49] A. van den Brandt, S. L'Yi, H. N. Nguyen, A. Vilanova, and N. Gehlenborg. Understanding visualization authoring techniques for genomics data in the context of personas and tasks. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):1180–1190, 2025. doi: 10.1109/TVCG.2024.3456298 2

[50] S. Verdinelli and N. I. Scagnoli. Data display in qualitative research. *International Journal of Qualitative Methods*, 12(1):359–381, 2013. 2

[51] F. B. Viegas, M. Wattenberg, and J. Feinberg. Participatory visualization with wordle. *IEEE transactions on visualization and computer graphics*, 15(6):1137–1144, 2009. 2

[52] X. Wang, S. Liu, Y. Chen, T.-Q. Peng, J. Su, J. Yang, and B. Guo. How ideas flow across multiple social groups. In *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 51–60. IEEE, 2016. 1

[53] F. Wanner, A. Stoffel, D. Jäckle, B. C. Kwon, A. Weiler, D. A. Keim, K. E. Isaacs, A. Giménez, I. Jusufi, T. Gamblin, et al. State-of-the-art report of visual analysis for event detection in text data streams. In *EuroVis (STARs)*, 2014. 1

[54] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. Tiara: a visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 153–162, 2010. 2

[55] S. Zhang, Y. Wang, H. Li, and W. Zhang. Who will support my project? interactive search of potential crowdfunding investors through insearch. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–6, 2022. 2