

## Methods for Analyzing Temporally Entangled Multimodal Data

Hanall Sung, University of Wisconsin—Madison, hanall.sung@wisc.edu  
Michael I. Swart, University of Wisconsin—Madison, mswart@wisc.edu  
Mitchell J. Nathan, University of Wisconsin—Madison, mnathan@wisc.edu

**Abstract:** While the challenges of collecting multimodal data are becoming surmountable with the help of the rapid development of sensor technologies, the challenges of *analyzing* multimodality remain substantial. It is imperative that researchers explore ways to successfully integrate theoretical and methodological frameworks for analyzing multimodal interactions in CSCL contexts. We identified two main approaches to analyzing multimodal data in CSCL settings—triangulating and interleaving—and highlighted the remaining challenges to unfolding the dynamic interplay between different modes with the consideration of temporality. To tackle these challenges, we presented an empirical example of multimodal learning analysis that practically employed the multimodal matrix and ENA for operationalizing and visualizing temporally entangled multimodal interactions. This paper (1) extends the theoretical underpinnings of temporality in studies of learning processes in CSCL settings, and (2) provides empirical evidence that indicates the potential of the interleaving approach to capture the core of complex meaning-making processes.

### Introduction

As the phenomena of interest in CSCL has extended to include wider sources of data, including verbal and non-verbal interactions, multimodality in CSCL (Blikstein & Worsley, 2016) is garnering increasing attention. While the challenges of collecting multimodal data are becoming surmountable with the help of the rapid development of sensor technologies, the challenges of *analyzing* multimodal data remain substantial. One current challenge is the complexity of multimodal interactions. Though different modes have varying temporal characteristics, they are often intertwined and occur in parallel during learning processes. For instance, imagine several students collaboratively engaged in a face-to-face discussion while using computer-supported tools. Over the course of the discussion, students will access learning materials on a laptop and subsequently incorporate information they encounter into their speech. While verbally articulating their understandings and questions, students will also spontaneously gesture while they communicate their conceptualizations. In such instances, the streams of log data, verbal speech, and gesture, each of which have different temporal characteristics but together contribute to meaning-making. In addition to examining the modes which directly associated with the meaning-making, other studies may also include surveys responses, interviews, text message correspondences, and other sources to better understand the learning phenomena. Such diverse sources of multimodal data can often create difficulties for strategically combining and interpreting data all together. Thus, it is imperative that researchers explore ways to integrate theoretical and methodological frameworks for analyzing multimodality in CSCL contexts. How can we systematically address multimodal interactions to clarify the meaning-making process in CSCL environments?

We identify two approaches to *multimodal learning analysis*<sup>1</sup> in CSCL contexts, triangulating and interleaving, and examine the affordances and challenges of both. When we wish to compare how each accumulative accounts of modes differently or similarly depicts the meaning-making process, the triangulating approach is useful to enrich the understanding of the learning process. However, if the meaning-making arises from the dynamic interplay between multiple modes, the interleaving approach is needed to accurately capture the essence of the learning process. We start by providing an empirical example of multimodal data containing *temporally entangled multimodal interactions*, and present how we conduct multimodal learning analysis that employs interleaving. In the example, we quantitatively and qualitatively analyze pre-service math teachers' multimodal discourse data (e.g., speech and gestures) collected from the interviews conducted before and after an intervention designed to help teachers improve their understanding of students' embodied mathematical thinking (Sung et al., 2022). Then, we devise a multimodal matrix (Shum et al., 2019) to operationalize theoretical constructs of a multimodal discourse and visualize its patterns using epistemic network analysis (ENA; Shaffer et al., 2016). The contributions of this paper are: it (1) extends the theoretical underpinnings of temporality in studies of learning processes in CSCL settings, and (2) provides empirical evidence that indicates the potential of the interleaving approach to capture the core of complex meaning-making processes.

<sup>1</sup> Throughout the document, a new term, *multimodal learning analysis* is introduced and used to capture the broader encompassing set of methods practices in the research areas analyzing multimodal data in learning contexts.

## Approaches to Analyzing Multimodal Data

### The triangulating approach to analyzing multimodal data

The “triangulation” metaphor originates from old navigation techniques that used multiple references to pinpoint an object’s location. Similarly, researchers use multiple points of multimodal data to better understand similar phenomena in the learning process, often triangulating from mixed-methods qualitative and quantitative analytic techniques. *Triangulating* allows researchers to obtain a more complete account of learning phenomena than could be produced from a single analysis.

For example, Rodríguez-Triana et al. (2016) collected log data, questionnaires, text-based interaction data (i.e., messages and comments), and video recordings to examine the effective use of a social media app in a co-located classroom assisted by technology. Each data source was used to explain distinct aspects of students’ learning during the lesson with the social media app; log data were used to cluster students by the level of engagement and video recordings were used to document students’ interactions. Similarly, Starr et al. (2018) collected multimodal data as proxy measures of students’ collaboration skills and applied a triangulating approach to validate their analysis. Researchers were able to quantitatively analyze dyads of students’ collaboration by using the Kinect™ to detect students’ movements and video recordings to transcribe their speech. Next, they combined these results with qualitative evaluations of students’ collaboration that they manually coded. Their triangulated analysis showed that the amount of speech and certain upper-body movements produced were significant predictors of the quality of students’ collaborations.

The triangulating approach is effective when we want to explore how accumulative accounts of interactions in learning represent the learning phenomena in similar or different ways: (1) complimentary findings explain different aspects of the phenomena, (2) findings that converge help validate phenomena, and (3) findings that diverge highlight possible alternative explanations (Tashakkori & Teddlie, 2010).

### The interleaving approach to analyzing temporally entangled multimodal data

While triangulating can be powerful, in isolation, this approach does not take into *temporal* structure of the learning process. As an inherently cumulative process, the *sequence* of events over the course of students’ learning is a direct representation of *how* and *what* one learns (Reimann, 2009). In CSCL settings, where learning occurs through multimodal interactions between people and digital tools, sources like log data, verbal data, and gestures often co-occur *temporally* in specific *sequence* that intertwines the meaning-making process. Typical approaches, however, tend to rely on comparing and correlating cumulative accounts of interactions (Knight et al., 2017), which fail to account for temporal and therefore omit contextual information of the interactions. Excluding such information from an analysis may impede truly capturing the most accurate account of the learning process (Mercer, 2008). For example, a student’s clicks to review some curriculum material may significantly influence their subsequent speech or co-speech gestures in a group discussion. However, documenting the mere accumulation of information from each mode, as is done using triangulation, cannot elucidate the interaction between the modes (see Figure 1).

**Figure 1**

Examples of the triangulating and interleaving approaches to analyzing multimodal interactions (gesture, speech, and log data) in a CSCL context; S1 and S2 stand for Student 1 and Student 2.



In this paper, I define *Interleaving* as an analytical approach that explores the interplay between events occurring in different modes during the learning process while also considering the temporal sequence of the events. As a result, researchers apply the interleaving approach in analyzing multimodal interactions to fully leverage both the timing and the order of events in the learning process. For instance, Bridges et al. (2020) conducted a micro-ethnographic analysis of learning activities in problem-based learning (PBL). While engaging

in PBL activities, students' interacted with facilitators as well as multiple learning resources (e.g., documents, diagrams, webpages, and videos). By tracing moment-by-moment actions in multimodal discourse, the authors documented how an event from one mode (e.g., an utterance) was temporally and sequentially interleaved with an event in another mode (e.g., a gesture). Bridges et al. (2020) claim that this micro-analysis of the discourse revealed complex interrelationships among the multimodal elements in collaborative learning and its cumulative effect on the learning process overall.

## Challenges for analyzing temporally entangled multimodal data

Given the benefits of triangulating and interleaving, the processes for integrating the temporal structures of multimodal data continues to present considerable challenges. First, the interleaving approach is designed to deal with different temporal granularities of each mode in the data (Ochoa, 2017). For example, a "click event" in log data occurs in a fraction of a second, whereas a verbal response can take much longer. These differences can complicate the process of synchronizing and fusing multimodal data. To address this issue, researchers can establish one common analytic time unit by aggregating up to the maximum size of granularity from the collected data (Knight et al., 2017). For example, if the largest temporal granularity is a 30-second-long verbal event, then the analytic time can be set to 30 seconds. Choosing a unit of common analytic time has a significant impact on the patterns of the data even though such decisions are seldom based on a solid methodological foundation (Knight et al., 2017).

Another methodological challenge of the interleaving is that the *temporal proximity* of events should be taken into account. Prior actions or events do not always have impacts on the entire history of interactions, but they can have substantial impacts within the *recent temporal contexts* (Shaffer, 2017). For instance, an utterance made 30 minutes prior would not have the same influence on a subsequent gesture as an utterance made in the preceding 5 seconds. Considerations of temporal proximity are crucial to accurately unraveling the complex interactions in CSEL environments. While some research on verbal data implies that modeling recent temporal context in discourse is important (e.g., Siebert-Evenstone et al., 2017), this approach has yet to be fully applied in the analysis of multimodal data.

In addition to these challenges, interleaving also presents analytical challenges that it can be time- and labor-intensive to scrutinize the interrelations between events in different modes in consideration of temporality. While machine-augmented computational techniques can assist with some analyses requiring human effort, this approach is still relatively taxing to conduct and complicated to interpret the holistic meaning of the data.

Thus, much work remains to be done to completely overcome the methodological and analytical challenges of the interleaving approach. Nonetheless, in order to accurately modeling complex interactions in CSEL settings, it is critical to take into account of temporal structure of discourse when people's meaning-making is multimodally intertwined throughout the discourse.

While the examples of triangulation in multimodal learning analysis abound, there has been relatively little research employing the interleaving approach. In this paper, we present an example of multimodal learning analysis applying a novel way of the interleaving approach to an existing dataset to illustrate how one can operationalize multimodal data and construe meaning across multiple modalities while considering the temporality of events. Our approach inspired by Echeverría et al. (2019)'s work based on the methods of *quantitative ethnography* (Shaffer, 2017) that combine qualitative and quantitative analysis.

## Methods

### Research Contexts

As a pilot study, we recruited K-12 math pre-service teachers (N=16) from a large Midwestern research university in the United States and provided an online, embodied learning environment that we developed, *The Hidden Village* (THV). In this environment, the teacher-participants engaged in an augmented embodied geometry curriculum that enables teachers to perform mathematically related body movements (action-based gameplay activity). Then they worked in collaborative design teams with other teachers to design and develop new body-based actions that could be used to facilitate students' embodied geometric reasoning (embodied co-design activity). The entire study took place online using Zoom and all activities were video recorded.

We hypothesized that this action-based intervention with THV would improve teachers' awareness of their own embodied mathematical knowledge and also enhance their abilities to interpret students' gesture use when evaluating students' verbal and non-verbal mathematical thinking. To test these claims, we conducted semi-structured interviews with each teacher-participant before and after the intervention. During the pre- and post-interviews, teachers watched 1-minute-long videos of a student reasoning about geometric conjectures. The videos contained verbal and non-verbal expressions of students' mathematical reasoning. To maintain privacy, the

original student videos were re-enacted by an adult actor. Teachers were prompted to assess the students' mathematical understanding and explicitly draw evidence from their observations of the videos.

## Data analysis

### Discourse Coding

To examine whether the action-based intervention affected teachers' awareness and abilities to interpret students' gestures, we first transcribed teachers' speech and gestures during pre- and post-intervention interviews into an event-based data log. The automatic audio transcription tool supplied by Zoom provided initial transcriptions of teachers' verbal speech and then a human coder corrected any inaccuracies and also inserted descriptions of teachers' gestures temporally linked to instances of speech (e.g., co-temporal, before, during or after the speech).

The completed multimodal transcripts were segmented by utterance, defined as when a teacher made a statement without pausing (1,130 segments total). The transcripts of teachers' verbal speech and gestures were coded in two different ways by each mode: (1) three verbal codes derived from the speech transcripts were employed using an automated process based on regular expression matching techniques (nCoder; Marquart et al., 2018); (2) two gestural codes were coded by human coders (see Table 1). Consequently, researchers validated three verbal codes via comparisons between two human raters and nCoder, while the two gestural codes were validated between two human raters. Inter-rater reliability was obtained between the two human coders and the automated nCoder; pairwise Cohen's kappa scores ranged between  $0.81 \leq \kappa \leq 0.93$  for each code and all kappa values have Shaffer's rho values  $\rho < 0.05$  (Shaffer, 2017).

**Table 1**

*Coding scheme and inter-rater reliability statistics (\*indicates  $p(0.65) < 0.05$ ; \*\*indicates  $p(0.65) < 0.01$ )*

Code Name	Description	Example	Kappa R1 v. R2	Kappa R1 v. ncodeR	Kappa R2 v. ncodeR
MATHEMATICAL THINKING	Discussion of mathematical concepts such as angles, lines, and conjectures.	"... she's not hitting on the fact on why the congruent angles are equal."	0.92**	0.92*	0.92**
ASSESSMENT	Judgment on the level of students' understanding	"I think she has a pretty decent understanding of it."	0.89*	0.92*	0.89*
VERBAL EVIDENCE	Teachers' use of students' utterances or reference to students' verbal reasoning as evidence	"... I feel like she's not able to describe what's going on because she's not using the keywords exactly in the correct form."	0.83*	0.96*	0.83*
RE-GESTURING	Teachers' actions of mimicking students' gestures.	"... she started by showing like [showing a pose of crossing arms and holding the pose] a cross with her arms to indicate different angles."	0.93**	-	-
EMBODIED COMMUNICATION	Teachers' embodied ways of communication beyond re-gesturing such as producing gestures while speaking or interpreting the meaning of gestures.	"I'd employ more of that, but then I would also make it clear that, like an arm, [showing a line diagonally with an arm], your arm is a line."	0.81*	-	-

Note: Quotation indicates verbal speech and brackets [...] indicate gestures

### Multimodal matrix

To structure the multimodal data for analysis, we tabulated the *multimodal matrix*, a conceptual data representation introduced by Echeverría et al. (2019) (see Figure 2). This approach to grounding quantitative data in the qualitative interpretation of the contexts in which the multimodal discourse emerges is inspired by the

methods of *quantitative ethnography* (Shaffer, 2017). Here, we discuss the details of the multimodal matrix to understand how we operationalized multimodal data taking into account temporality.

**Figure 2**

*An example of the multimodal matrix inspired by Echeverría et al. (2019)*

		Dimensions of discourse					Multimodal observation
		Verbal			Gestural		
Stanzas	Time	Verbal Evidence	Math. thinking	Assessment	Re-gesturing	Embodied. Com.	Segment
	00:00:02 - 00:00:09	0	0	1	0	1	
	00:00:11 - 00:00:18	1	1	0	1	0	
	00:00:22 - 00:00:32	0	1	1	0	1	
	00:00:36 - 00:00:50	0	0	1	0	0	
	....						
	00:05:03 - 00:05:12	0	0	1	0	0	
00:05:15 - 00:05:31	1	1	0	1	0		

In this multimodal matrix (Figure 2), each type of multimodal data can be coded into *multimodal observations* (i.e., kinds of information derived from multimodal traces) that identify the *dimensions of discourse* (e.g., verbal, gestural). Codes of individuals' utterances are associated with verbal dimensions while codes of individuals' gesture use are associated with gestural dimensions. As shown in Figure 2, each dimension of discourse can include several multimodal observations in a series of columns, where each column represents a different code. If there are three different epistemic codes for verbal dimension, then there is a different column for each to be scored. Each row in the matrix represents *segments*, the smallest unit of meaning considered for analysis (Shaffer, 2017). In the current matrix, a segment contains all the information relevant to an event (e.g., utterances, gestures) during the semi-structured interviews. Time windows (i.e., analytic time unit) for each segmentation were not pre-defined but instead segmented by the occurrence of natural pauses in speech taken by the participants. Next, these segments grouped into *stanzas*, representing collections of rows considered to be within the same recent temporal context. We applied the method of moving stanza windows (Siebert-Evenstone et al., 2017) to construct a network model that captured the natural interactions between observations occurring within recent temporal context. Using this method, we explain interactions between temporally and contextually entangled multimodal observations. In essence, we wanted to capture how the events in one mode (e.g., an utterance in the verbal dimension) influence events in the other mode (e.g., hand shape in the gestural dimension) in a recent temporal context. Based on a grounded analysis driven by the data, we set the size of a moving stanza window to 6 rows (the current row plus the 5 previous rows) within each teacher's interview.

### ENA Discourse Model

Epistemic network models were constructed based on the multimodal matrix using ENA (Shaffer et al., 2016), a discourse analysis technique for identifying and quantifying the connections among cognitive elements in a discourse. ENA creates dynamic nodal networks of discourse around a computed mean centroid weighting interconnections between codes (Shaffer, 2017). The codes in ENA models correspond to the epistemic elements that characterize the discourse where the edges of the network represent the relative frequency of co-occurrence between two codes. To investigate differences (if any) between the ENA networks of the pre- and post-interviews, we applied a two-tailed paired-sample t-test, assuming unequal variance, to compare the positions of the plotted points in the projected ENA space. Finally, corresponding ENA network graphs were created to visually interpret the connections and account for the differences between pairings.

## Results

### Qualitative Results

#### Pre-Interviews

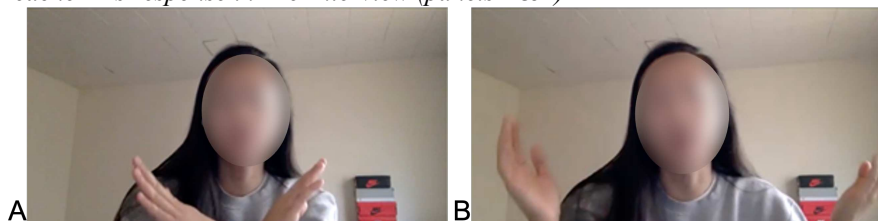
In pre-interviews, teachers frequently made hasty inferences between students' gestures and students' mathematical understanding. For example, when a student in the videos produced a static gesture representing geometric objects relevant to the conjecture, teachers concluded that the student's conceptual understanding was accurate, overlooking the relevance of the *function* of the student's gesture in their reasoning process.



Figure 3 presents an example of a pre-interview with teachers; screenshots of Teacher 1 (T1, panels A & B) and an excerpt of their interpretation of the student's mathematical understanding. T1's re-gesturing mimics the student's gesture expressing vertical angles (Panel A), indicating that they recognize what the student's gesture depicts (Multimodal transcript, line 1). T1, however, assumes prematurely that the appearance of the student's gestures translates to a solid mathematical understanding. Instead of considering the more precise role of gestures, T1's attention quickly moves to the student's utterance ("It either adds up to 180 or 360", lines 1-2) and infers the student's level of mathematical understanding based on how the student stated it without assurance ("[mimicking 'I don't know' pose] at the end that just tells me ...", Panel B, lines 2-3).

**Figure 3**

*An example of Teacher 1's response in Pre-Interview (panels A&B)*



Timeline	Multimodal transcript
00:02:00	She understood the opposite rule [mimicking the student's X pose], but when she said, "it either
-	adds up to 180 or 360" and then was like 'I don't know really' [making a pose to portray 'I don't
00:02:17	know'] at the end that just tells me that she doesn't fully understand the rule.

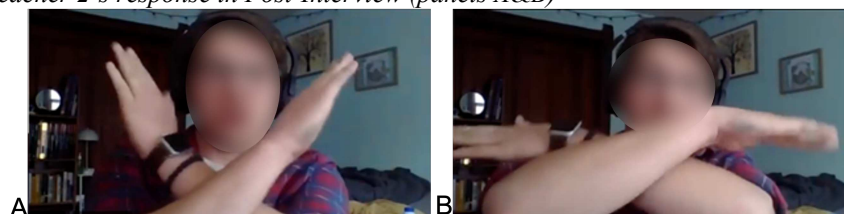
### Post-Interviews

Teachers in post-interviews, on the other hand, were more likely to pay specific attention to the relationships between the student's gestures and verbal speech, especially how student's gestures complemented student reasoning about the geometric conjectures.

For instance, Teacher 2 (T2) in Figure 4 evaluates how a student's explanations of an underlying mathematical idea were deficient. T2 does so by purposefully integrating the information provided by the student's speech and gestures. First, T2 highlights the core logic of the student's proof ("The logic she was trying to use was almost like a contradiction, like a proof by contradiction"; Multimodal transcript, lines 2-3). Next, T2 demonstrates how the student's bent hand gesture (Panel A & B) is unavailing in their proof by saying "but she didn't really do anything with it" (line 6). T2's assessment of the student's mathematical thinking is an interpretation of what the student's gestures mean as well as how the gestures contribute to the reasoning process.

**Figure 4**

*An example of Teacher 2's response in Post-Interview (panels A&B)*



Timeline	Multimodal transcript
00:02:24 - 00:02:34	I thought she did a kind of poor job on that one. You very much could see the problems.
00:02:36 - 00:02:42	The logic she was trying to use was almost like a contradiction, like a proof by contradiction I felt like.
00:02:42 - 00:02:49	She was trying to say "okay, well, if the lines aren't straight than the angles won't work".
00:02:50 - 00:02:55	And the gesture, she kept using which is this [mimicking the student's sequence of poses], but like she didn't really do anything with it.

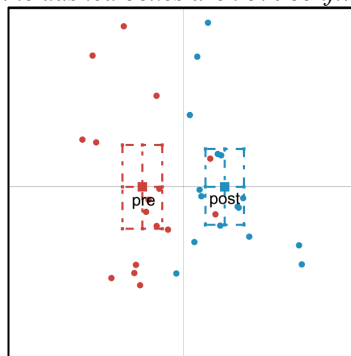
### Quantitative Results

Using ENA, we analyzed the multimodal discourse data from pre- and post-interviews. The ENA scatter plots (Figure 5) revealed that there were statistically significant differences in discourse patterns between pre- and

post-interviews that corroborated the aforementioned qualitative findings ( $\bar{x}_{Pre} = -0.68$ ,  $\bar{x}_{Post} = 0.68$ ,  $t(16) = 7.32$ ,  $p < 0.01$ , Cohen's  $d = 1.77$ ).

**Figure 5**

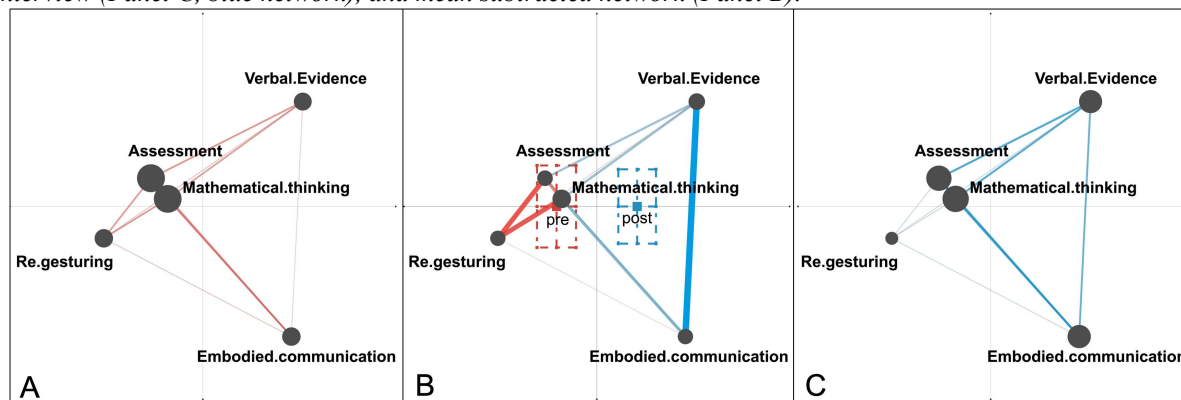
ENA scatterplot showing teachers in pre-interviews (red) and post-interviews (blue). Each point is a single teacher; the squares are group means; the dashed boxes are 95% confidence intervals (t-distribution).



Next, we constructed the mean epistemic networks to examine which connections account for the differences in teachers' responses between pre- and post-interviews (see Figure 6). When we subtract pre from post to identify the differences (Panel B), we can see that the teachers in the pre-interviews (red network, left) made more links between RE-GESTURING, ASSESSMENT, and MATHEMATICAL THINKING, whereas teachers in the post-interviews (blue network, right) made more links between VERBAL EVIDENCE, EMBODIED COMMUNICATION, and MATHEMATICAL THINKING. This means initially, teachers were more likely to be engrossed with drawing simple connection between representing gestures and having mathematical knowledge. After the intervention, teachers were more likely to incorporate the information from students' gestures with speech and focus on how they work together in the reasoning process while assessing students' mathematical understanding. These results also align with the qualitative results.

**Figure 6**

Mean ENA network diagrams showing the connections made in pre-interview (Panel A, red network) and post-interview (Panel C, blue network), and mean subtracted network (Panel B).



## Discussion

In this study, we identified two main approaches to analyzing multimodal data in CSCL contexts—triangulating and interleaving—and highlighted the remaining challenges to unfolding the dynamic interplay between different modes with the consideration of temporality. To tackle these challenges, we presented an empirical example of multimodal learning analysis that practically employed the multimodal matrix and ENA for operationalizing and visualizing temporally entangled multimodal interactions. In this paper, we: (1) reflected the contextual and temporal context in structuring discourse data instead of arbitrarily setting analytic time units (Knight et al., 2017), (2) modeled the interactions between multimodal events in light of temporal proximity, and (3) fully utilized machine-augmented analytic techniques (e.g., automatic transcription, nCoder and ENA) to make the process more accurate and less laborious (Blikstein & Worsley, 2016). Through quantitative and qualitative analysis, we

demonstrated the potential of the interleaving approach to capture the complexities of the meaning-making processes in CSCL settings.

Despite these promising results, this study still has several limitations. While the temporal proximity of the events in multimodal data was considered, it did not employ a more Bayesian model that updates different weights of the influences that prior and subsequent events would have on a given event. Furthermore, it did not completely account for ranges of temporal granularities that vary across the different modes for each segmentation (albeit it came close). Further research is needed to develop and deploy more advanced methods for modeling temporality with greater precision.

Nonetheless, this study offers practical application of methodologies for operationalizing multimodal data and interpreting its meaning while accounting for temporality of contexts. Moreover, it contributes to a growing line of research addressing of the importance of temporality of learning processes in CSCL environments.

## References

- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining : Using Computational Technologies to Measure Complex Learning Tasks. *Journal of Learning Analytics*, 3(2), 220–238.
- Bridges, S. M., Hmelo-Silver, C. E., Chan, L. K., Green, J. L., & Saleh, A. (2020). Dialogic intervisualizing in multimodal inquiry. *International Journal of Computer-Supported Collaborative Learning*, 15(3), 283–318. <https://doi.org/10.1007/s11412-020-09328-0>
- Echeverría, V., Martínez-Maldonado, R., & Shum, S. B. (2019). Towards collaboration translucence: Giving meaning to multimodal group data. *Conference on Human Factors in Computing Systems - Proceedings*, 1–16. <https://doi.org/10.1145/3290605.3300269>
- Knight, S., Wise, A. F., & Chen, B. (2017). Time for Change : Why Learning Analytics Needs Temporal Analysis. *Journal of Learning Analytics*, 4(3), 7–17.
- Marquart, C. L., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). *ncodeR: Techniques for Automated Classifiers. R package version 0.1.2*. <https://cran.r-project.org/package=ncodeR>
- Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, 17(1), 33–59. <https://doi.org/10.1080/10508400701793182>
- Ochoa, X. (2017). Multimodal Learning Analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 129–141). Society for Learning Analytics Research. <https://doi.org/10.18608/hla17>
- Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 239–257. <https://doi.org/10.1007/s11412-009-9070-z>
- Rodríguez-Triana, M. J., Holzer, A., Prieto, L. P., & Gillet, D. (2016). Examining the effects of social media in co-located classrooms: A case study based on speakup. *European Conference on Technology Enhanced Learning*, 247–262. [https://doi.org/10.1007/978-3-319-45153-4\\_19](https://doi.org/10.1007/978-3-319-45153-4_19)
- Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45.
- Shum, S. B., Echeverría, V., & Martínez-Maldonado, R. (2019). The Multimodal Matrix as a Quantitative Ethnography Methodology. *International Conference on Quantitative Ethnography*, 26–40. [https://doi.org/10.1007/978-3-030-33232-7\\_3](https://doi.org/10.1007/978-3-030-33232-7_3)
- Siebert-Evenstone, A. L., Collier, W., Ruis, A. R., & Shaffer, D. W. (2017). In Search of Conversational Grain Size : Modelling Semantic Structure Using Moving Stanza Windows. *Journal of Learning Analytics*, 4(3), 123–139.
- Starr, E. L., Reilly, J. M., & Schneider, B. (2018). Toward using multi-modal learning analytics to support and measure collaboration in co-located dyads. *Proceedings of International Conference of the Learning Sciences, ICLS, I*, 448–455.
- Sung, H., Swart, M. I., & Nathan, M. J. (2022). Teaching teachers teaching students: How embodied cognition can help pre-service teachers assess students' mathematical thinking. *Proceedings of the Annual Meeting of the American Educational Research Association (AERA)*.
- Tashakkori, A., & Teddlie, C. (2010). *Sage handbook of mixed methods in social & behavioral research* (2nd ed.). Sage.