JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

Volume 12(1), 271–292. https://doi.org/10.18608/jla.2025.8423

# Transmodal Analysis

David Williamson Shaffer[1]*, Yeyu Wang[2] and Andrew Ruis[3]

**Abstract**

Learning is a multimodal process, and learning analytics (LA) researchers can readily access rich learning process data from multiple modalities, including audio-video recordings or transcripts of in-person interactions; logfiles and messages from online activities; and biometric measurements such as eye-tracking, movement, and galvanic skin response. While many techniques are used in LA to model different types of learning process data—most of which are state-dependent (or state-space) approaches that model a learning process at any given time as a function of the preceding events—constructing multimodal models has so far relied on fusion of different data streams, which converts multimodal data into a unimodal format. This creates a number of problems for multimodal modelling, the most important of which is that it treats different data modalities as equivalent. That is, existing state-dependent models of fused data cannot easily account for (a) events that may have different impacts on future events based on what those future events are and the context in which they are occurring; (b) how events may influence some groups of learners differently; and (c) which events are visible (and thus potentially impactful) to which students. In this paper, we propose *transmodal analysis* (TMA), a mathematical and computational framework designed to address these challenges. TMA is not a data analysis method but rather an approach to modelling that can augment existing state-dependent models of learning processes to account for multimodal data without data fusion. We present a conceptual and methodological description of TMA, and we include an appendix with a detailed worked example as a proof of concept. While this approach is in the early stages of development, it has the potential to significantly improve the ease, efficiency, and fairness of multimodal analyses of learning processes.

**Notes for Research**

- Existing multimodal models of learning processes typically rely on data fusion, which converts multimodal data into a unimodal format. While this enables many analytic techniques to operate on multimodal data, it also makes it difficult if not impossible to account for important differences across modes, such as different temporal scales, different impacts of events on different learner populations, and different horizons of observation.
- *Transmodal analysis* (TMA) is a methodology for augmenting existing learning process models to fully account for multimodality without requiring data fusion.
- TMA provides a framework for making principled decisions (or hypotheses) about how different data streams interact that could ultimately improve the speed, efficiency, transparency, and fairness of analyses of learning processes.

**Keywords**

Multimodal data, transmodal analysis, data fusion, data transfusion, temporal influence functions, horizon functions, learner impact functions.

**Submitted:** 20/03/2024 — **Accepted:** 25/11/2024 — **Published:** 23/01/2025

[1*] *Corresponding author Email: dws@education.wisc.edu Address: Wisconsin Center for Education Research, University of Wisconsin–Madison, 1025 W. Johnson St., Madison, WI, 53706, USA. ORCID iD: https://orcid.org/0000-0001-9613-5740*

[2] *Email: ywang2466@wisc.edu Address: Wisconsin Center for Education Research, University of Wisconsin–Madison, 1025 W. Johnson St., Madison, WI, 53706, USA. ORCID iD: https://orcid.org/0000-0003-1978-5453*

[3] *Email: arruis@wisc.edu Address: Wisconsin Center for Education Research, University of Wisconsin–Madison, 1025 W. Johnson St., Madison, WI, 53706, USA. ORCID iD: https://orcid.org/0000-0003-1382-4677*

## 1. Introduction

Learning is a multimodal process (K.-s. Tang et al., 2014). Online or face to face, students interact with peers, teachers, parents, tutors (machine and human), games, simulations, calendar organizers, online tools, videos, images, and texts. Because many learning processes now take place online—and especially as a result of the COVID-19 pandemic—education researchers have

access to rich data about the learning process across multiple modalities: logfiles of student actions in simulations, chat or discussion board messages, records of resources accessed, and transcripts or videos of discussions, as well as data such as eye tracking, gestures, geospatial location, and galvanic skin response in more experimental settings.

The field of *learning analytics* (LA) uses many different models of learning processes to analyze this data, and different methods highlight and investigate different facets of learning. The prevailing approaches to analyzing learning among LA researchers are *state-dependent and state-space models of learning* (SMLs), which include (but are not limited to) techniques such as Markov chains (Gupta et al., 2022), process mining (Huang et al., 2023), lag sequential models (Jeng & Chung-Nien, 2022), Bayesian knowledge tracing (Lee et al., 2023), and epistemic network analysis (Teasley et al., 2023). These SMLs assume that learning at some time point is influenced by the events that precede it, and thus they use the same data streams as both predictors and outcomes in models of learning processes. That is, these approaches model learning at some point in time as a function of the events that came before it (Rahmani & Fay, 2022).

With a few exceptions, however, the SMLs that are currently used to analyze learning build models based on one data modality at a time. That is, they model dialogue, or a logfile, or gestures, but not dialogue and gestures or dialogue, logfile, and gestures. A *multimodal* model of learning that would include all of these data types has to account for complexities that are either not accounted for or difficult to account for in a unimodal model. For example, a multimodal model has to provide a mechanism to address how

1. different *types of events* (questions from a teacher, chats with a peer, views of a resource) and different *properties of events* (gender of a person gesturing, linguistic fluency of a speaker, reading level of a person reading a document) might influence future events with more or less impact over time;
2. different *characteristics of students* (age, cultural or ethnic background, gender identification, whether instruction is in their native language [L1] or a non-native language [L2]) might lead students to respond to events in different ways; and
3. different students in a learning environment have different *horizons of observation* (Hutchins, 1995), making some events (e.g., a conversation in another group of students) *visible or invisible* (or less visible) to a given learner at a particular point in time.

To understand multimodal learning processes, it is thus critical to *extend* existing SMLs to accommodate (a) multiple data types and streams, (b) how different types and properties of events influence different learners over time, and (c) how these differences function given the structure of a particular learning environment.

Here, we propose an approach to constructing SMLs that accounts for these complexities in modelling multimodal data. We describe *transmodal analysis* (TMA) as a *mathematical and computational framework* that enables widely used models of learning processes to analyze multiple modalities of data. In doing so, we stress two things:

1. TMA is not a data analysis *method*; rather, it is an approach to modelling that can *augment existing models* of learning processes—each with its own strengths—to account for multimodal data.

2. While we present a worked example of a TMA analysis, our intent is to provide this as a proof of concept rather than to argue that these results, by themselves, definitively establish the utility of TMA. That is, this is a *conceptual* and *methodological* paper rather than an empirical paper, and we hope that readers will be willing to indulge us in this approach.

## 2. Background

### 2.1 Multimodality

*Multimodality* refers to the use of multiple modes or channels of communication within a single activity. Multimodal communication can be simultaneously textual, verbal, aural, spatial, tactile, and imagistic—and it may also include measurable characteristics like emotional arousal, attention, and the range of human sensory apparatus. Di Mitri and colleagues (2018) argue that these different modalities *interact* to convey and deepen meaning, and thus from a theoretical perspective, multimodal LA is better aligned with the nature of human communication than unimodal approaches.

Critical to the concept of multimodality is the idea that different modes or channels may require different skills and competencies. For example, learning in science, technology, engineering, and mathematics (STEM) is a multimodal phenomenon because STEM domains rely heavily on *representational artifacts* that express scientific ideas and concepts in a range of modes, including verbal explanations, text, diagrams, graphs, and simulations (K.-s. Tang et al., 2014).

Expressing ideas in these multiple modes is an integral part of the language of science. Indeed, much of STEM learning involves developing the ability to interpret and translate between different representations of scientific information. As a result, much of the research on multimodal STEM learning investigates how students develop scientific, mathematical, and technical understanding by simultaneously using different modalities within and across multiple representations (Jewitt et al., 2001).

And, of course, every domain of learning shares this basic property: learning is multimodal because it involves developing the ability to work across representations using a range of communicative channels.

Analyses of learning that focus on only one modality at a time, such as classroom dialogue, have made significant contributions to our understanding of learning processes. However, analyses of learning that integrate multiple modes of data are potentially more accurate and more equitable—and therefore better able to influence both current practice and future research (Worsley, 2022). (Those who wish to see a more comprehensive view of the literature on multimodal learning as it relates to TMA can refer to Appendix B.)

### 2.2 Multimodal Analysis of Learning

The field of *multimodal learning analytics* (MMLA) uses multimodal data to model how students develop understanding. Work on such models has grown steadily over the last decade (Figure 1). For example, studies have looked at how data on talk, motion, location, gaze, emotion, and self-reports can be used to model how students learn from a lecture (Raca & Dillenbourg, 2013; Sümer et al., 2021; Alkabbany et al., 2023) or how data on talk, writing, drawing, movement, and galvanic skin response can be used to model how students learn from problem-solving exercises (H. Tang et al., 2022).

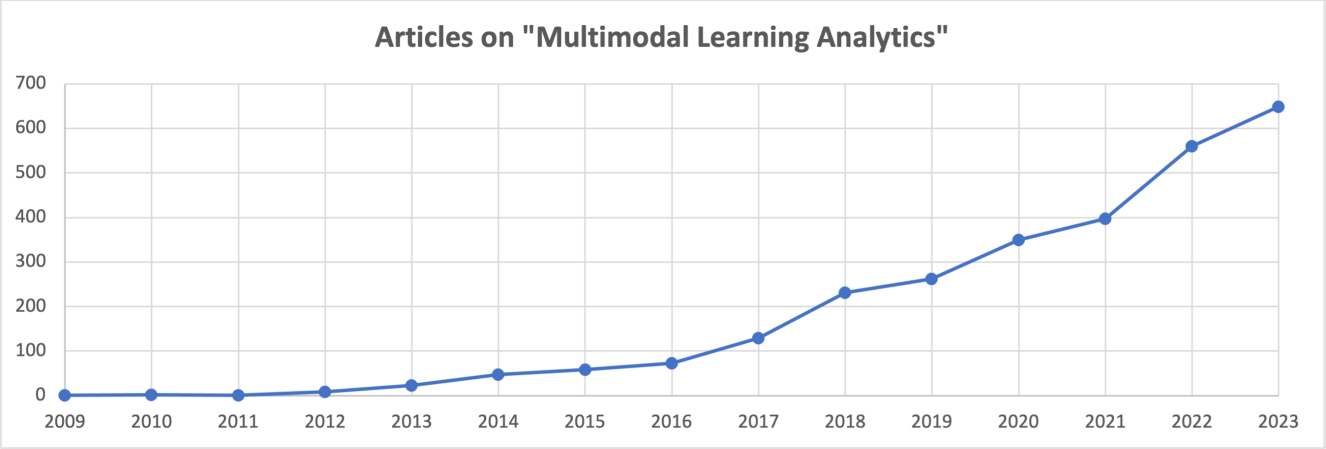


**Figure 1.** Number of articles on Google Scholar referencing "multimodal learning analytics" over time.

Ochoa and Worsley (2016) argue that different modes of data in a multimodal context are produced by different physical, psychological, and social processes and therefore influence learning through different mechanisms. For example, Hung and Higgins (2015) compared communications between learners using synchronous video conferencing and synchronous chat and found that video communication involved faster turn-taking, more rephrasing, and more confirmatory revoicing or repetition; thus, video conversations involved more turns of talk per topic (and more topics in the same period of time) than conversations in chat.

Moreover, data collected from different modalities uses different *extraction processes*, which often have different characteristics (Yusuf et al., 2023). For example, data collected from galvanic skin response sensors may be sampled at up to 2000 Hz (although rates of 1–10 Hz are more typical), whereas functional magnetic resonance imaging data is typically collected at a rate of 0.5 Hz, and both need to be corrected to account for phasic activity—that is, for cyclical biological rhythms such as heartbeat (Lahat et al., 2015). Gesture data may be video-recorded at 30 Hz, but it needs to be interpreted as gestural units that may be several seconds in duration. Eye movements are typically recorded on the order of 30 Hz but need to be translated into *regions of interest* in order to be interpreted. Chat data may be recorded in bursts of text every few seconds. Events in a computer logfile may be recorded anywhere from every second to one every few minutes depending on the activity involved.

Consider, for example, a learning setting in which children read aloud from a science textbook to a *reading companion robot* (Figure 2; Michaelis & Mutlu, 2018, 2019; Cagiltay et al., 2020). When a child sees an ID tag in the text, they show the tag to the robot. The robot selects from several available comments about a short activity in the tagged section of the book. The child reads the activity directions, works on the activity, and then describes what they observed during the activity to the robot.

Seven data streams are collected: (1) audio recording of the child's speech, (2) the child's eye gaze, (3) video of the child's gestures, (4) the tags that the child shows to the robot, (5) the robot's replies, (6) the robot's eye gaze, and (7) any written work the child does. Critically, these events take place on *different time scales* and are *interpreted in different ways*. For example, showing a tag to the robot may be the result of a combination of the child reading, gesturing, and speaking to the robot or to themself over several minutes. Eye gaze is interpreted differently in relation to reading than to speaking.

A critical challenge in multimodal data analysis is representing, segmenting, and integrating such different data streams in a way that preserves the important properties of each type of data (Sharma & Giannakos, 2020). Worsley (2014) suggests that

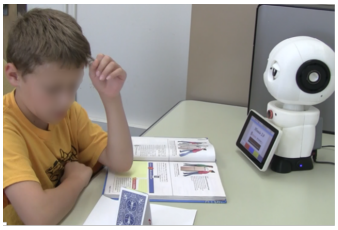

**Figure 2.** A child reading to a robot generates multiple modes of data that need to be coordinated, including speech, eye gaze, gestures, and pages that the child shows to the robot.

integration of complex, multimodal data can take place at different levels:

1. In *naive fusion*, researchers analyze data from each stream independently and then use the results of those analyses in some combined model of learning (Emerson et al., 2020).
2. In *low-level fusion*, researchers put the raw data into a common data format, usually by aggregating data within equal segments of time, for example, aggregating data on eye gaze, gesture, and student speech into 1-second segments (Ma et al., 2022).
3. In *high-level fusion*, the raw data is coded based on its relationship to some hypothesized learning process, and then codes for each data stream are placed in a data matrix in equal segments of time (Sung et al., 2022).

These fusion methods are all problematic, however. Some allow researchers to analyze each data stream in ways that respect individual characteristics (naive and high-level fusion). Others allow researchers to align or coordinate different data streams in time (low- and high-level fusion). However, all of these approaches *convert multimodal data into a unimodal format*, and all of these approaches are *bespoke* in the sense that each fusion process has to be invented de novo for datasets with different combinations of data modalities.

### 2.3 Challenges for Existing Fusion Methods

As a result, all of these existing approaches (1) ignore the moment-by-moment influence of one learning modality on another, (2) ask researchers to perform complex transformations on multiple data streams without providing general conceptual and procedural guidance, and/or (3) ask researchers to transform large datasets in complex ways without adequate technological and theoretical support. As a result, none of these data fusion methods makes it easy for researchers to integrate multiple data modalities in ways that Di Mitri and colleagues (2018) argue are central to understanding multimodal learning.

First, there is no standard MMLA approach that accounts for the fact that the same kind of event may have different impacts depending on what future event is being influenced and what the learning context is. For example, high levels of galvanic skin response (indicating stress or excitation) may have a different impact on how students answer test questions than on how students read (Nourbakhsh et al., 2012)—that is, excitation from test anxiety is qualitatively different than excitation from immersion in a compelling narrative. As anyone who watches horror movies or rides roller coasters knows, not every form of discomposure is the same.

Second, there is no standard MMLA approach to account for how *events may differentially influence some groups of students*. For example, research suggests that bilingual students read faster in L2 than monolingual students in L1 with equivalent comprehension (Spätgens & Schoonen, 2019), and math anxiety affects test performance differently for female students than for male students (Devine et al., 2012). This poses a problem for the development and validation of multimodal SMLs because learning datasets typically contain subgroups: populations of students defined by demographics (e.g., race, native language, disability, income) or other metadata (e.g., test scores, attendance) that have relatively low numerical representation in a dataset. Models that researchers develop and validate on such data may be biased toward majority groups and thus ultimately unfair to subgroups (Mehrabi et al., 2021; Chouldechova & Roth, 2018), with the potential to reify and even augment existing inequities (Gardner et al., 2019; Mayfield et al., 2019). But despite broad attention to issues of equity in education, model bias in MMLA has received little systematic attention (Yan et al., 2022).

Third, no extant MMLA method makes it possible to *distinguish which data streams are visible to a given student* within the structure of the learning environment, for example, a setting where students can hear one another talk but cannot see what is on each other's phones or computers.

Of course, we recognize the challenges in *processing* multimodal data. As described by Fayyad and colleagues (1996), processing steps include (1) extracting relevant information from raw data, (2) cleaning the data (for example, handling missing data or removing noise), and (3) transforming the data into a format that can be used by analysis tools. However, we argue that an equally fundamental problem is to develop a method for modelling learning processes that *integrates* multiple streams of data once those streams have been collected and cleaned.

Further, we argue that while it may be possible to accomplish this with some existing SMLs by reformatting and fusing data by hand, such bespoke approaches (a) are prohibitive for large and complex datasets, (b) make it difficult or impossible to systematically explore the impact of multimodality on learning processes, and (c) make it difficult to systematically address questions of equity and bias.

These are the specific challenges in MMLA that TMA is designed to address.

### 2.4 State-Dependent Models of Learning

From an *evidence-centred design* perspective (Almond et al., 2013), any model of learning has four components: (1) a *context determination process*, (2) an *evidence capture process*, (3) an *evidence identification process*, and (4) an *evidence accumulation system*. Broadly speaking, the context determination process specifies instances in which evidence is collected, the evidence capture process records activity in those instances, the evidence identification process converts those recordings into variables that describe relevant features, and the evidence accumulation system uses those variables across multiple instances to construct a model of the learning that was taking place.

Reimann (2009) argues that much research on learning is *variable centred*: (a) concepts are operationalized as constructs and measured as numerical variables, (b) variables are measured at the same time points, and (c) the underlying assumption is that independent variables act continuously on dependent variables. However, a growing body of evidence shows the importance of process models to describe the moment-by-moment progression of learning as it unfolds in time (Knight et al., 2017). Csanadi and colleagues (2018), for example, showed that argumentation is better explained by the temporal ordering of evidence, reasoning, and hypothesizing than the total amounts of such argumentative moves.

Reimann and colleagues (2014) and Alonso-Fernández and colleagues (2019) provide useful overviews of analytic methods that model learning processes as temporal sequences of events, that is, models that provide information about the relationship between some prior event(s) in the learning process and some subsequent learning event. SMLs commonly used in LA include methods such as Petri nets for process mining (Huang et al., 2023), Markov chains (Gupta et al., 2022), sequential analysis (Akçapınar & Hasnine, 2022), epistemic network analysis (Teasley et al., 2023), group communication analysis (Dowell et al., 2018), Bayesian networks (Jiang et al., 2023), Bayesian knowledge tracing (Lee et al., 2023), autoregression (Ahmad et al., 2023), vector autoregression (Y. Zhou & Kang, 2023), additive factors models (F. Chen & Lu, 2022), time series analysis (Dorodchi et al., 2020), and structural equation models (X. Liu, 2022).

These SMLs used in LA are all *event-based* models—they represent learning processes as sequences of actions (Reimann et al., 2011)—and they are based on three core principles: (1) student actions in learning are *responses* to the conditions in the setting, known as the *ground* for those actions; (2) the ground for an action can include both conditions when that action takes place and prior conditions leading up to it; and (3) student actions are in general more influenced by things that happened recently.

In other words, using the language of evidence-centred design, the *contexts* are individual events. Evidence is *captured* in a single modality, *identified* in the ground for each event, and *accumulated* across multiple events and their ground to estimate some latent variable or variables for students (or groups, or other units of analysis associated with events).

More precisely, any SML is a modelling function $\Psi$ that takes a set of events $E$ and applies a grounding function $G$ and an estimation function $\Phi$:

$$\Psi(E) = \Phi(E, G(E)) \tag{1}$$

Formally, event $x$ is represented as a vector $\boldsymbol{e}_x$, which is a concatenation of a metadata vector $\boldsymbol{m}_x$ and a content or code vector $\boldsymbol{c}_x$. That is, $\boldsymbol{e}_x = (\boldsymbol{m}_x, \boldsymbol{c}_x)$. So, for example, $m_{xt}$ could indicate the *time* that event $\boldsymbol{e}_x$ occurred and $m_{x\delta}$ could indicate the *duration* of the event. Similarly, $m_{xs}$ could indicate which student's action is being recorded and $m_{xd}$ which *discussion group* the student is in, and similarly for school, class, demographic information, and so on. (Properties are ascribed to *events* rather than *students* because characteristics of a student might change over time.)

For example, Chiu (2018) constructed a simple autoregressive model based on data from ninth-grade students in an algebra class. Eighty students were video-recorded working in mixed-gender groups of four. Each turn of talk in the videos was coded by two research assistants using social moderation.

The data was coded as

$c_{i,evaluation} = 1$ if there was a CORRECT EVALUATION (a student agreed with a previous speaker's correct idea or disagreed with an incorrect idea), and

$c_{i,content} = 1$ if there was NEW CORRECT CONTENT (a student proposed a correct mathematical statement that was new in the discussion).

Both codes were zero otherwise.

The *grounding function* Chiu used was

$$G_{Chiu}(\boldsymbol{e}_x) = \boldsymbol{c}_{(x-1),evaluation} \tag{2}$$

That is, the ground for any turn of talk, $x$, was a function of the content of the event immediately preceding it—in this case, whether the previous turn had a CORRECT EVALUATION.

The *estimation function* was a regression:

$$c_{x,content} = \beta\ G_{Chiu}(\boldsymbol{e}_x) + \varepsilon \tag{3}$$

So in this model, whether the current event had NEW CORRECT CONTENT was predicted by its ground, that is, by whether the previous turn of talk had a CORRECT EVALUATION.

Within such models, the content $\boldsymbol{c}_x$ of any event $x$ may have different *weighting*. For example, a gesture could indicate SURPRISE at an intensity level of 0.4 or 40%, and saying "I think you're wrong" could indicate DISAGREEMENT with an intensity level of 1.0 or 100%. But SMLs are unimodal: the algorithms and formulas treat the gesture and the statement the same way. All that matters is that at some time $\boldsymbol{m}_{xt}$ there was SURPRISE (intensity 0.4) and/or DISAGREEMENT (intensity 1.0), regardless of the source of those inferences.

### 2.5 Challenges in Adapting State-Dependent Models of Learning to Multimodal Data

This core assumption of SMLs is problematic from a multimodal perspective. For example, consider Table 1, which shows a sequence of events that might take place over 90 seconds in an imagined mathematics classroom.

**Table 1.** Events in a mathematics classroom as two students complete a worksheet on solving quadratic equations.

| Event | Time | Actor | Role | Modality | Action |
|---|---|---|---|---|---|
| $\boldsymbol{e}_1$ | 10:02:03 | Lin | Student | Reading | Opens textbook to sample problem |
| $\boldsymbol{e}_2$ | 10:02:04 | Ms. Gomez | Teacher | Reading | Opens lesson plan |
| $\boldsymbol{e}_3$ | 10:02:15 | Ms. Gomez | Teacher | Speech | "Ok, class: Go ahead and start the worksheet with your partner." |
| $\boldsymbol{e}_4$ | 10:02:20 | Desda | Student | Writing | Starts working on first problem |
| $\boldsymbol{e}_5$ | 10:02:21 | Lin | Student | Writing | Starts working on first problem |
| $\boldsymbol{e}_6$ | 10:02:56 | Desda | Student | Reading | Opens textbook to sample problem |
| $\boldsymbol{e}_7$ | 10:02:59 | Lin | Student | Speech | "Desda, how do you factor a quadratic again?" |
| $\boldsymbol{e}_8$ | 10:03:01 | Lin | Student | Reading | Opens chapter on quadratics |
| $\boldsymbol{e}_9$ | 10:03:02 | Desda | Student | Speech | "They want us to do it by completing the square." |
| $\boldsymbol{e}_{10}$ | 10:03:07 | Lin | Student | Speech | "Oh, yeah, right." |
| $\boldsymbol{e}_{11}$ | 10:03:14 | Lin | Student | Reading | Turns to textbook page for completing the square |
| $\boldsymbol{e}_{12}$ | 10:03:33 | Desda | Student | Speech | "Do you need help getting started?" |

In this short excerpt, the teacher, Ms. Gomez, looks at her lesson plan ($\boldsymbol{e}_2$) and then asks the class to complete a worksheet on quadratic equations in pairs ($\boldsymbol{e}_3$).

Desda and Lin are a mixed-ability pair of students: Desda is a native English speaker who is doing well in the class. Lin is a multilingual language learner (not a native English speaker) who has been struggling in math.

Both Desda ($\boldsymbol{e}_4$) and Lin ($\boldsymbol{e}_5$) take out the worksheet and start working on the problems.

Almost as soon as they start working (36 seconds later), Desda opens their textbook to a sample problem ($\boldsymbol{e}_6$). Lin's book was already open to the example ($\boldsymbol{e}_1$).

A moment later ($\boldsymbol{e}_7$), Lin becomes confused and asks Desda for help on the first problem, which is "factoring a quadratic." At the same time, they open their text to the chapter on quadratics ($\boldsymbol{e}_8$).

Desda replies ($\boldsymbol{e}_9$) that "They want us to do it by completing the square." Lin says, "Oh, yeah, right" ($\boldsymbol{e}_{10}$) and then turns to the section on completing the square in their text ($\boldsymbol{e}_{11}$). Desda waits for 20 seconds and then asks if Lin needs help ($\boldsymbol{e}_{12}$).

This short sequence shows the interplay of talk, writing, and reading as two students work to follow the instructions of the teacher and complete a math worksheet—what Sung and colleagues (2022) describe as the *interleaving* of data modalities, where events interact with one another across modalities. Desda's suggestion that they complete the square ($\boldsymbol{e}_9$) is a response to the teacher's instruction to work in pairs ($\boldsymbol{e}_3$), the fact that Desda read the problem they are working on ($\boldsymbol{e}_4$), and Lin's question

about how to factor a quadratic ($\boldsymbol{e}_7$). Of these, Lin's question that immediately precedes Desda's suggestion is almost certainly the most influential, even though it does not provide a complete explanation of the reasons for Desda's action.

Figure 3 illustrates how a lag sequential analysis of the kind used by Chiu (2018) might model these events. When Desda says to Lin that they should solve a problem by completing the square ($\boldsymbol{e}_9$), a lag sequential model might predict this based on a window containing the three events that immediately precede it: Desda opening the textbook ($\boldsymbol{e}_6$), Lin asking Desda how to factor a quadratic ($\boldsymbol{e}_7$), and Lin looking at the chapter on quadratics ($\boldsymbol{e}_8$)[1].

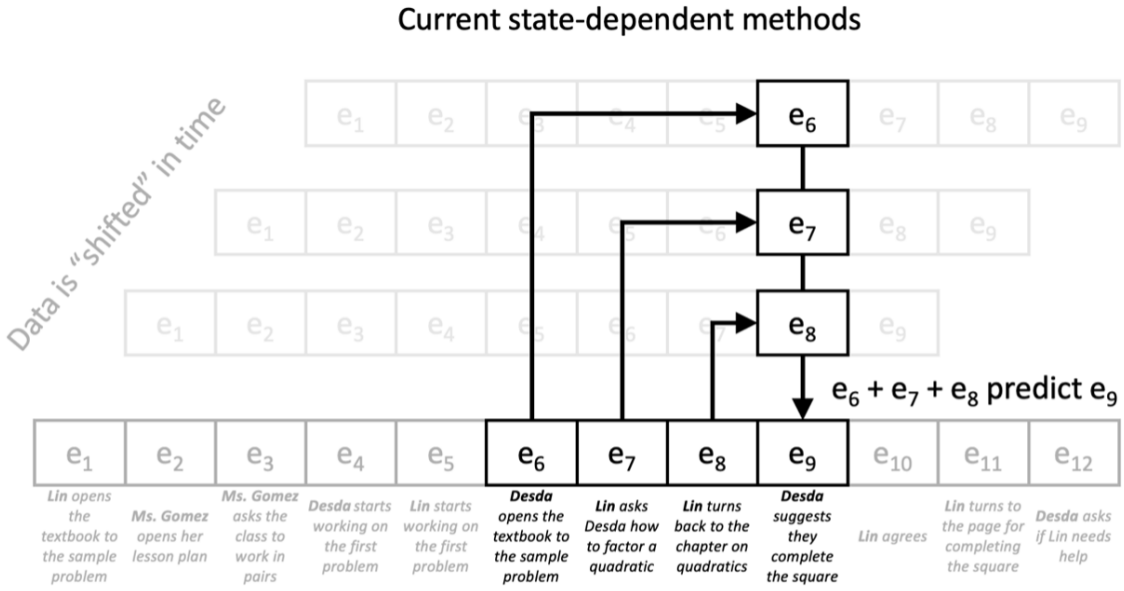


**Figure 3.** SMLs model learning at one point in time using previous events.

The size of the window determines which events are relevant predictors, and a different window size could include more or fewer events. Thus, in a lag sequential model (and in most existing SMLs), key values of *influence* and *recency* are fixed within the modelling framework. For example, every event might influence the next three or four events, or might influence events in the next 4 or 5 seconds in the model.

Applying such models across multiple data modalities is challenging because multimodal data is *heterogeneous*: attention span, visibility, and impact can vary between modes of data. To take a simple example, Lin opens their textbook to the chapter on quadratics ($\boldsymbol{e}_8$) immediately before Desda suggests completing the square ($\boldsymbol{e}_9$). But whether Desda could see what page the book was open to depends on the structure of the learning environment at that point in time. If they were working online, probably not. If they were working side by side, perhaps yes.

Thus, a multimodal model must account not only for the weight and content of events but also for the way that different events in different circumstances influence events in the future. Specifically, a multimodal SML needs to account for

1. *types of events*, which ground future events in different ways;
2. *characteristics of student populations*, allowing types of events to ground future events differently depending on characteristics of the learner involved; and
3. *horizons of observation*, allowing events to be included in the ground only for some students or groups of students.

In other words, discourse has *structure* (Gee, 2015; Shaffer, 2017): there are particular ways in which events unfold in a specific context, and the structure of discourse determines how one event influences another. This is, of course, true for any data and model of events, but we argue that in the context of *multimodal* data, the issue is more salient because the structure of discourse is often more complex than existing SMLs can account for.

[1]This description is a simplification of more complex models, such as those in state-space analyses that include both the influence of prior states and how observations provide evidence for latent variables (e.g., with a Q matrix). However, the central claim—that SML models depend on time-lag properties of prior information—remains true even for models that add additional components or more complex mathematical representations.

As Table 2 shows, with few exceptions, SMLs are not designed to address these features of multimodal data. Existing approaches that account for multiple modes of data (a) make assumptions that do not always hold in the context of LA and (b) do not account for all relevant features of multimodal heterogeneity in learning process data.

**Table 2.** Multimodal features of state-dependent models.

| | Multiple Data Sources | Multiple Event Types | Student Population & Event Type Interactions | Horizon of Observation |
|---|---|---|---|---|
| Additive Factors Model | ✗ | ✗ | ✗ | ✗ |
| Autoregression | ✗ | ✗ | ✗ | ✗ |
| Bayesian Knowledge Tracing | ✗ | ✗ | ✗ | ✗ |
| Group Conversation Analysis | ✗ | ✗ | ✗ | ✗ |
| Lag Sequential Models | ✗ | ✗ | ✗ | ✗ |
| Markov Chains | ✗ | ✗ | ✗ | ✗ |
| Time Series Analysis | ✗ | ✗ | ✗ | ✗ |
| Bayesian Networks | ✗ | ✗ | ✓ | ✗ |
| Petri Nets | ✗ | ✗ | ✓ | ✗ |
| Epistemic Network Analysis | ✗ | ✗ | ✓ | ✓ |
| Structural Models | ✓ | ✓ | ✗ | ✗ |
| Vector Autoregression | ✓ | ✓ | ✗ | ✗ |

### 2.6 Foundation of TMA

SMLs thus model learning by *lagging*, or shifting data in time, and then using some prior time point(s) to either (a) predict the next action in the learning process or (b) compute the value of some latent variables that describe the learning process.

TMA differs from this approach in three significant ways, as shown in Figure 4.

Specifically, a *TMA-augmented SML* (T/SML) models a current event in terms of all previous events, rather than only events within some potentially arbitrary fixed window (Siebert-Evenstone et al., 2017). For example, a T/SML model would account for the fact that the teacher, Ms. Gomez, asked the students to work in pairs ($\boldsymbol{e}_3$) even though this happened outside of the window of three events.

TMA accomplishes this by constructing three *transmodal grounding functions* that augment or replace the grounding function for an existing SML.

Recall that events are represented by vector $\boldsymbol{e}_x = (\boldsymbol{m}_x, \boldsymbol{c}_x)$, where

- $\boldsymbol{m}_x$ is a vector of metadata about the event, including its time $t_x$, type $\tau_x$, relevant learner characteristics $\boldsymbol{\lambda}_x$, and relevant information $\boldsymbol{\omega}_x$ about where the event fits in the structure of the environment—that is, $\boldsymbol{m}_x = (t_x, \tau_x, \boldsymbol{\lambda}_x, \boldsymbol{\omega}_x)$—and
- $\boldsymbol{c}_x$ is a vector of codes that represents the relevant content of the event.

Transmodal grounding functions include the following:

1. A set of temporal influence (TI) functions $t_\tau(\Delta t) \in \mathcal{T}$ that describe how *events of type* $\tau$ influence future events. Specifically, if $\Delta_{ix} = t_x - t_i$ is the difference in time for a pair of events $\boldsymbol{e}_i$, $\boldsymbol{e}_x$, then $t_{\tau_i}(\Delta t_{ix})$ is the influence of an event of type $\tau_i$ at time $t_i$ on a later event at time $t_x$.

   For example, when Ms. Gomez asks the class to work in pairs ($\boldsymbol{e}_3$), this influences future events differently for Desda than opening the textbook to the sample problem ($\boldsymbol{e}_6$). Thus, $t_{talk}$ would be applied to $\boldsymbol{e}_3$ and $t_{read}$ would be applied to $\boldsymbol{e}_6$ because the events are different event types, and $t_{talk}$ would be different than $t_{read}$ because these event types influence future events differently.

   TI functions are thus *hypotheses* about how one *type* of event grounds future events.

2. A set of *learner impact* (LI) functions $\ell_{\boldsymbol{\lambda}}(t_{\tau_i}, \dot{\boldsymbol{m}}_i, \Delta t_{ix}) \in \mathcal{L}$ that describe how the impact of events with characteristics $\dot{\boldsymbol{m}}_i$ differ from the usual function $t_{\tau_i}$ for learners with characteristics, $\boldsymbol{\lambda}$. Specifically,

$$\ell_{\boldsymbol{\lambda}_x}(t_{\tau_i}, \dot{\boldsymbol{m}}_i, \Delta t_{ix}) = v_{\dot{\boldsymbol{m}}_i \boldsymbol{\lambda}_x} t_{\tau_i}(\rho_{\dot{\boldsymbol{m}}_i \boldsymbol{\lambda}_x} \Delta t_{ix}) \tag{4}$$

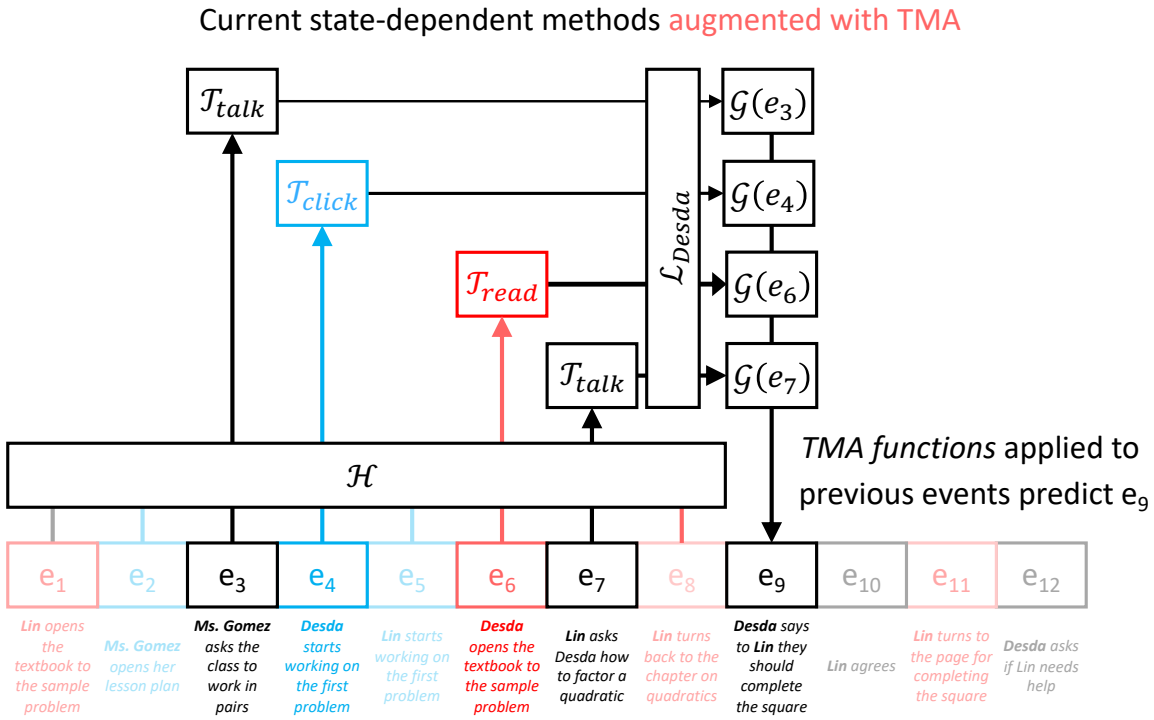


**Figure 4.** T/SMLs model learning at one point in time by applying temporal influence ($\mathcal{T}$), learner impact ($\mathcal{L}$), and horizon ($\mathcal{H}$) functions to construct ground functions based on previous events.

Here, $\dot{\boldsymbol{m}}_i \subset \boldsymbol{m}_i$ is a subset of the metadata relevant to the learner impact function, $v_{\dot{\boldsymbol{m}}_i \boldsymbol{\lambda}_x}$ is a parameter indicating the *strength* with which an event with properties $\dot{\boldsymbol{m}}_i$ influences a student with characteristics $\boldsymbol{\lambda}_x$, and $\rho_{\dot{\boldsymbol{m}}_i \boldsymbol{\lambda}_x}$ indicates the change in influence over time of events with properties $\dot{\boldsymbol{m}}_i$ on students with characteristics $\boldsymbol{\lambda}_x$.

For example, the teacher's request ($\boldsymbol{e}_3$) may influence Lin's question ($\boldsymbol{e}_7$) differently than it influences Desda opening the text ($\boldsymbol{e}_6$) because the class was taught in Desda's L1 and Lin's L2. Thus, $\ell_{L1}(\boldsymbol{e}_3, \boldsymbol{e}_6)$ would not be the same as $\ell_{L2}(\boldsymbol{e}_3, \boldsymbol{e}_7)$.

LI functions are thus *hypotheses* about how events influence future events *for students with different characteristics*.

3. A set of *horizon functions* $\hbar_{\boldsymbol{\omega}}(\ddot{\boldsymbol{m}}_i) \in \mathcal{H}$ that determine the impact of an event with properties $\ddot{\boldsymbol{m}}_i$ on events with properties $\boldsymbol{\omega}$ given the structure of the learning environment. Thus, for some $\ddot{\boldsymbol{m}}_i \subset \boldsymbol{m}_i$, $\hbar_{\boldsymbol{\omega}_x}(\ddot{\boldsymbol{m}}_i)$ indicates whether (or how much) event $\boldsymbol{e}_i$ influences event $\boldsymbol{e}_x$.

   For example, when Desda responds ($\boldsymbol{e}_9$) to Lin's question ($\boldsymbol{e}_7$), we assume that Desda cannot see what Lin is reading ($\boldsymbol{e}_8$) or where Ms. Gomez clicked her screen ($\boldsymbol{e}_5$). Thus the horizon function would exclude those events from the model of Desda's talk ($\boldsymbol{e}_9$), modelling $\boldsymbol{e}_9$ as a response only to the teacher's talk ($\boldsymbol{e}_3$), Lin's talk ($\boldsymbol{e}_7$), and Desda's own prior actions ($\boldsymbol{e}_4$ and $\boldsymbol{e}_6$).

   Horizon functions are thus *hypotheses* about how the *configuration* of the learning environment determines which events can influence others.

Thus, in TMA, actions and latent variables are not modelled on the *events themselves* shifted in time. Instead, they are model-based *functions applied to events* that represent the temporal characteristics of different event types (TI functions) and how those types of events impact different groups of students (LI functions) given the organization of the learning environment (horizon functions). This results in a *transmodal grounding function* with the general form

$$G^T(\boldsymbol{e}_x) = \sum_{i|t_i \leq t_x} \hbar_{\boldsymbol{\omega}_x}(\ddot{\boldsymbol{m}}_i)\ell_{\boldsymbol{\lambda}_x}(t_{\tau_i}, \dot{\boldsymbol{m}}_i, \Delta t_{ix})\boldsymbol{c}_i \tag{5}$$

Put another way, this TMA grounding function expresses hypotheses about the structure of discourse in terms of functions that can be applied to data rather than through data formatting.

Without such functions, no current SML technique accounts for all this information about multimodal learning processes. It is possible (in some cases) to reconfigure data and then combine multiple iterations of an existing SML to account for some of

the features of a TMA model. For example, it is possible to account for horizons of observations by manually constructing a dataset for each individual student, including only the events that can influence them. However, TMA provides a *systematic process for transmodal data fusion* based on explicit hypotheses about the influence of events, the characteristics of students, and the learning environment being modelled.

### 2.7 Parameter Derivation

Central in the process of defining a T/SML, then, is specifying the TI, LI, and horizon functions.

#### 2.7.1 TI Functions

In general, TI functions are created by choosing an appropriate function type (power, exponential, logistic, etc.) based on prior empirical work or theory. Specific parameters are determined using one of two *impact derivation processes* (IDPs):

1. In *a priori IDP*, parameters are specified based on either theory or prior empirical work.

2. In *empirical IDP*, an influence estimation is conducted by defining an influence function:

$$f(\boldsymbol{e}_i, \boldsymbol{e}_x) = \begin{cases} 1, & \text{information from event } \boldsymbol{e}_i \text{ is needed to interpret } \boldsymbol{e}_x \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

   Then, TI functions are defined by the *inverse cumulative distribution function*, $1 - F(t)$, where $F(t)$ is the cumulative distribution function computed on a sample of events in the dataset[2]; e.g., the probability that reading a resource influences a student's action 2 minutes after the resource was opened would be $\mathscr{t}_{reading}(2) = 1 - F(2)$.

#### 2.7.2 LI Functions

LI functions use similar approaches, except that in an empirical IDP for an LI function, two parameters ($\rho_{\dot{m}\boldsymbol{\lambda}}$ and $v_{\dot{m}\boldsymbol{\lambda}}$) are optimized such that if $f(\boldsymbol{e}_i, \boldsymbol{e}_x)$ is a function that indicates the level of impact of $\boldsymbol{e}_i$ on $\boldsymbol{e}_x$, then

$$f(\boldsymbol{e}_i, \boldsymbol{e}_x) \approx v_{\boldsymbol{m}_i\boldsymbol{\lambda}_x} \mathscr{t}_{\tau_i}(\rho_{\boldsymbol{m}_i\boldsymbol{\lambda}_x} \Delta t_{ix}) \tag{7}$$

#### 2.7.3 Horizon Functions

Horizon functions are typically determined by the structure of the learning setting. That is, a researcher specifies what information can be seen by a learner under specific conditions.

#### 2.7.4 TMA Functions as Hypotheses

Regardless of their derivation, TI, LI, and horizon functions (collectively, TMA functions) represent *claims* or *hypotheses* about (1) how different modalities impact the learning process, (2) the ways in which students might systematically differ as learners, and (3) the structure of the learning environment itself. Wise and Shaffer (2015) argue that developing models based on such hypotheses, grounded in prior theoretical work and qualitative analyses of the data, leads to interpretable and transparent models, and that such models are critical in accounting for underrepresented groups in the data and systematically checking researcher bias. TMA can help researchers avoid inaccurate conclusions about learners—and particularly minoritized groups of learners—by providing models that account for differences in how students use learning resources and engage in learning interactions.

We thus hypothesize that TMA models can provide a more nuanced, more accurate, and more equitable view of learning processes for diverse learners and thus expand our understanding of effective multimodal learning processes and allow researchers to account for diversity and address questions of equity in multimodal learning.

## 3. Worked Example of TMA

In this section of the paper, we summarize a worked example of TMA analysis using a TMA-enhanced version of *epistemic network analysis* (ENA) (Shaffer et al., 2016). ENA is a widely used tool for analyzing learning process data in the LA community (Porter et al., 2021). The details of the worked example are provided in Appendix A.

[2]Specifically, over some subset of the data, let $R_\tau$ be a set such that for every pair of events in $R_\tau$ with event type $\boldsymbol{e}_i = \tau$, and $t_i < t_x$, there exists an ordered pair $(\Delta t_{ix}, f(\boldsymbol{e}_i, \boldsymbol{e}_x)) \in R'_\tau$. For any time $t \geq 0$, let $R'^t_\tau = \{(\Delta t_{ix}, f(\boldsymbol{e}_i, \boldsymbol{e}_x)) | \Delta t_{ix} \leq t\} \subseteq R'_\tau$. The cumulative distribution function is estimated as $\hat{F}_\tau(t) = (\sum_{(\Delta t_{ix}, f(\boldsymbol{e}_i, \boldsymbol{e}_x)) \in R'^t_\tau} f(\boldsymbol{e}_i, \boldsymbol{e}_x)) / |R_\tau|$. Fit analytic function $F_\tau(t)$ as a smoothed cumulative distribution function. The TI function is $\mathscr{t}_\tau(\Delta t_{ix}) \equiv 1 - F_\tau(\Delta t_{ix})$. In this sense, TMA functions resemble propagators in physics, which describe the probability of events occurring at some future time and place.

## 3.1 Methods

### 3.1.1 Context

We used a dataset with two-mode data from a virtual internship, *RescuShell*, in which undergraduate engineering students work as interns at a fictional engineering company (Chesler et al., 2015). This virtual internship simulates the engineering design process using an online work portal with research reports and a built-in chat interface. The internship had two parts:

1. In the *discovery phase* of the simulation, students review and summarize research reports, create and evaluate device prototypes, and discuss design choices with teammates and a mentor.
2. In a subsequent *design phase*, students create a final proposal that balances the requirements of prospective customers given the constraints of the possible device components.

We used three conditions:

C1. ENA applied to the student chat messages,

C2. ENA applied to a low-level fusion of student chat messages and logfile data of resource usage, and

C3. TMA-enhanced ENA (T/ENA) applied to student chat messages and logfile data of resource usage.

We compared these conditions to see which model would

(a) account for more variance,

(b) be more efficient, and

(c) have higher fidelity to the data.

### 3.1.2 Qualitative Analysis

To establish a baseline for interpreting the quantitative models, we conducted a secondary qualitative analysis (Anderson & Paulus, 2021). Prior work looked at the multimodal relationship between chat messages and resource use by students (Sung et al., 2019) and showed qualitatively that there was a relationship between students' use of resources and their chat discussions. But this work did not model their interaction quantitatively.

### 3.1.3 Discourse Hypotheses

To construct a TMA model, we created a *discourse model*—that is, a set of hypotheses about the *structure of transmodal interactions* in the *RescuShell* learning environment, including

1. that the impact of chats from mentors would diminish at a different rate than those from students;
2. that the impact of reading a resource would rise asymptotically while a student was reading, and once a document was closed, its influence would diminish;
3. that a student's chat would influence every other student in the same discussion, but that opening a resource would influence only the student who opened it.

We used these discourse hypotheses to construct TI and LI functions $\ell_{student}(t_{chat}, teacherChat, \Delta t)$, $\ell_{student}(t_{chat}, studentChat, \Delta t)$, and $t_{reading}$, using an empirical IDP. We used an a priori IDP to specify horizon functions $\hbar_{discussion,reading}$ and $\hbar_{discussion,chat}$. The derivation of these functions is critical to any TMA model, and details are provided in Appendix A.

### 3.1.4 ENA

In ENA, $\boldsymbol{c}_{ij}$ describes whether event $\boldsymbol{e}_i$ has some code $j$ associated with it. The *grounding function* for an event $\boldsymbol{e}_x$ is the sum of the code vectors $(\boldsymbol{c}_x + \boldsymbol{c}_{x-1} + \cdots + \boldsymbol{c}_{x-w+1})$ in a window of the $w$ lines preceding it in the student's discussion group $m_{xd}$:

$$G_{ENA}(\boldsymbol{e}_x) = \sum_{i|\boldsymbol{m}_{i,group}=\boldsymbol{m}_{x,group},(x-w)<i\leq x} \boldsymbol{c}_i \tag{8}$$

In T/ENA, the grounding function is not computed using a preset window but based on the functions above that describe the temporal influence of events in this setting:

$$G^T(\boldsymbol{e}_x) = \sum_{i|t_i \le t_x} \hbar_{\boldsymbol{\omega}_x}(\dot{\boldsymbol{m}}_i)\ell_{\boldsymbol{\lambda}_x}(t_{\tau_i}, \dot{\boldsymbol{m}}_i, \Delta t_{ix})\boldsymbol{c}_i \tag{9}$$

For both the ENA and T/ENA models, the estimation function for each unit of analysis was a cumulative adjacency matrix that represents the *connections* made by that student between codes the student uses and codes in the common ground. These matrices are normalized and then subjected to a dimensional reduction with *singular value decomposition* (SVD) that represents each student with two *ENA scores* that can be represented as a single point in a two-dimensional space $(\hat{\Phi}_1, \hat{\Phi}_2)$ .

In this study, we constructed a *subtracted network graph* for each model by (a) constructing two networks for each student, one in the discovery phase and the other in the design phase; (b) computing the mean of each of these networks across all students; and (c) subtracting the two networks. We used the subtracted network graphs to compare model results to the qualitative analysis. We then tested how well the ENA points predicted which phase of the game the student was in. Details are provided in Appendix A.

#### 3.1.5 Model Comparison

Across the ENA and T/ENA models (conditions 1–3), we compared ENA scores between students in the discovery and design phases of the project using a logistic regression of the form $logit\,(discovery) = \beta_0 + \beta_1\hat{\Phi}_1 + \beta_2\hat{\Phi}_2$. ENA and T/ENA models were compared using the Akaike information criterion (AIC) and McFadden's pseudo $R^2$.

#### 3.1.6 Qualitative Comparison

To establish fidelity to the data, we evaluated the ENA and T/ENA models not merely on predictive accuracy or variance explained but also based on how well each model represented a qualitative analysis of the data.

### 3.2 Results

#### 3.2.1 Qualitative Analysis

Appendix A shows a qualitative analysis of chat messages and students' resource use in each phase of the simulation. One of the key outcomes of this analysis was the importance of the linkages between DEVICE COMPONENTS, PERFORMANCE, DESIGN, and DATA during the discovery phase. In contrast, in the design phase, there were significant connections between PERFORMANCE, DATA, and DESIGN.

However, as shown in Figure 5, these connections were only apparent when considering both the chat messages and the logs that indicated what resources students were consulting as they went through these different phases of the design process.

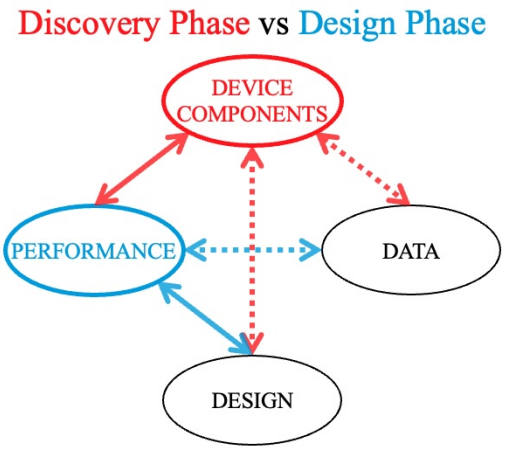


**Figure 5.** In the discovery phase of the simulation, students in *RescuShell* made connections from DEVICE COMPONENTS to PERFORMANCE in their ***chat*** conversations, but linkages from DEVICE COMPONENTS to DATA and DESIGN (shown as dashed lines) were only evident when we included ***reading*** data. In contrast, in the design phase students made connections from PERFORMANCE to DESIGN in their ***chat*** conversations, but linkages between PERFORMANCE and DATA (shown as dashed lines) were only evident when we included ***reading*** data.

#### 3.2.2 Quantitative Models

Details of the regression models are provided in Appendix A and summarized in Table 3.

**Table 3.** A comparison of model performance shows that the T/ENA model (condition 3) outperforms both ENA on chats (condition 1) and multimodal ENA on chats and resources (condition 2). [** = $p < 0.01$, * = $p < 0.05$.]

| Condition | Predictors | | | Model Evaluation | |
|---|---|---|---|---|---|
| | Intercept | SVD 1 | SVD 2 | McFadden's $R^2$ | AIC |
| (1) ENA with only chats | 1.21** | 10.64** | 2.69 | 0.18 | 61.45 |
| (2) ENA with chats and resources | 1.89* | 14.39* | 8.59 | 0.13 | 66.58 |
| (3) TMA with chats and resources | 0.96 | 13.90* | 10.46** | **0.32** | **51.46** |

### 3.2.3 Model Comparison

As shown in Table 3, model C1 explains more variance ($R^2 = 0.18$) than model C2 ($R^2 = 0.13$) and is more efficient ($AIC = 61.45$) than model C2 ($AIC = 66.58$)

Table 4 shows that both models C1 and C2 differ on relative connection strength from the qualitative analysis 33% of the time (two of six connections), while model C3 shows the same relative strengths as the qualitative analysis across all connections in the analysis.

**Table 4.** Connections that are identified as stronger in one phase than another by models in the three conditions relative to connections identified in the qualitative model.

| Connection | Qualitative | ENA Chats | ENA Chat+Resources | T/ENA Chat+Resources |
|---|---|---|---|---|
| DEVICE COMPONENTS & DATA | discovery | ✓ | ✓ | ✓ |
| DEVICE COMPONENTS & DESIGN | discovery | ✓ | ✗ | ✓ |
| DEVICE COMPONENTS & PERFORMANCE | discovery | ✓ | ✓ | ✓ |
| PERFORMANCE & DATA | design | ✗ | ✗ | ✓ |
| PERFORMANCE & DESIGN | design | ✓ | ✓ | ✓ |
| DATA & DESIGN | neither | ✗ | ✓ | ✓ |

We thus conclude that model C1 outperforms model C2 in terms of variance explained and efficiency, but neither model provides a more accurate representation of the qualitative analysis than the other.

Model C3 explains more variance ($R^2 = 0.32$) and is more efficient ($AIC = 51.46$) than C1 and C2. Moreover, Table 4 shows that C3 has the same relative strength as the qualitative analysis on all six connections.

## 4. Discussion

The worked example we summarize above (and describe in detail in Appendix A) shows a widely used SML (ENA) applied to multimodal data from a learning setting.

The setting was an engineering simulation divided into two phases. In the discovery phase, students worked together to understand the design properties and key performance parameters of a mechanical exoskeleton design project. In the design phase, students used this information to create and test prototype exoskeleton designs. Throughout the simulation, students were working in teams using an online collaboration portal. In the portal, students communicated with each other through chat messages and were able to access written resources, such as technical reports and schematics, that were related to the design task.

We constructed three models:

C1. ENA applied to the student chat messages,

C2. ENA applied to a low-level fusion of student chat messages and logfile data of resource usage, and

C3. TMA-enhanced ENA (T/ENA) applied to student chat messages and logfile data of resource usage.

When we compared the performance of these models, the results showed that model C1 outperformed model C2 and model C3 outperformed both of the other models.

In saying this, we do not argue that this worked example provides dispositive evidence that TMA outperforms ENA or SMLs more generally. We anticipate that (a) there are multimodal datasets and settings where an SML will outperform its corresponding T/SML *and* (b) it is possible to construct bespoke data fusion methods that would enable some SML models to outperform T/ENA for other specific datasets.

Our point in presenting this example is to show how TMA works with one specific SML, and that in some cases, a TMA-enhanced SML can outperform an SML that has not been adapted for multimodal data. Moreover, this example illustrates some of the core principles by which TMA accounts for significant features of multimodal data.

## 4.1 Core Principles of TMA

These results present something of a conundrum: an underlying hypothesis of MMLA is that the information from additional data modalities should provide a more complete picture of student learning. But in this case, modelling both chat messages and resources resulted in *poorer* model performance than modelling chat messages alone. On the other hand, the TMA model, which also used both chat messages and resources, resulted in *better* performance than either of the SML models.

This suggests the importance of several features of TMA.

### 4.1.1 Multimodal Models Require Appropriate Discourse Models

These results show that multimodal models *can* account for more information about the learning environment than unimodal models. However, the results here also suggest that multimodal models can do so *only if* the interactions between the different modalities are modelled appropriately.

Like any discourse setting, multimodal interactions have structure: ways in which events systematically influence (or do not influence, or influence more or less) future events. And one critical feature of multimodal data is that different modalities have different interactional properties. Indeed, that is one of the reasons we consider multiple modalities in the first place.

Because of the structure of the SML we considered here—and more generally by the nature of all extant SMLs—the SML-only model applied to two modalities treated two *different data modalities* as if they had the *same interactional properties*: the same impact on future events, the same impact on individual learners, and the same visibility to different participants in the setting. But this is clearly a set of assumptions that are not warranted except in very particular circumstances.

By accounting for the structural features of the multimodal discourse being modelled (or attempting to account for them), the TMA-enhanced model was able to use the information contained in multiple data modalities more effectively.

It may be possible to construct bespoke data fusion methods that would enable an SML model to outperform its TMA-enhanced version for some datasets. However, we argue that such an approach both (a) is less efficient than using TMA and (b) potentially leads to more biased models. We argue this based on a second important principle underlying TMA, namely *data transfusion*.

### 4.1.2 Data Transfusion

Any data transformation is based on some underlying hypothesis or hypotheses about the nature of the phenomenon that produced the data. In the case of MMLA, these claims include *discourse hypotheses*, or assertions about how the different modalities of data act and interact in the learning process. And we argue that

> *it is better to encode discourse hypotheses as functions that are applied to the data in the modelling process rather than by reformatting the data in advance of model construction.*

We refer to this process of constructing functions to describe how modalities act and interact as data *transfusion* rather than data *fusion*. The term *fusion* (in the context of multimodal data) connotes combining multimodal data streams to construct a fixed multimodal data object. In contrast, the term *transfusion* suggests an ongoing process by which multimodal data is combined and recombined to express discourse hypotheses.

The process of constructing TMA functions for such data transfusion is non-trivial, as our worked example suggests. However, the process of formatting data to account for the different properties and complex interactions between data modalities is non-trivial as well (Ochoa et al., 2018; Buckingham Shum et al., 2019)—and perhaps more difficult, in the end, than constructing functions. But regardless of the comparative level of difficulty of these two approaches, there are at least three advantages to using functions rather than formatting to integrate multiple data modalities:

1. **Iterativity.** It is easier to change discourse hypotheses if they are modelled in functions rather than encoded in data formatting. Changing a parameter in a function is far more efficient than the complex refactoring of data that might be required to change (for example) which data is visible to each student. This means that it is easier to test multiple hypotheses about the nature of the learning environment—and thus about student learning—in a given setting. Moreover, because functions and parameters are very flexible compared to the options in formatting data in a spreadsheet (or even a relational database), researchers are less likely to make assumptions (consciously or unconsciously) because they are easy to implement.

2. **Clarity.** Representing discourse hypotheses as functions and function parameters rather than encoding them in data formatting forces researchers to be explicit about the hypotheses they are making. Once data has been formatted, it is easy to forget the assumptions that went into the construction of a dataset, and these assumptions are not always clearly articulated in data analyses—particularly when secondary analyses are conducted on a formatted dataset. Moreover, this

explicit representation of discourse hypotheses lays out a framework that prompts researchers to declare (and test) the assumptions they are making.

3. **Flexibility.** Simply put, the range of discourse hypotheses that can be expressed with functions is larger than what can easily be encoded in data formatting. TMA makes it possible to simultaneously account for (a) the varying temporal influence of different modalities, (b) the systematic impacts data modalities have on different learners, and (c) how the structure of the learning environment shapes these effects across different modalities.

In sum, using TMA functions to transfuse multimodal data streams makes it possible to model multimodal data more efficiently, more effectively, and more transparently.

Formatting that fuses rather than transfuses multiple data streams does have a place—an important place—in MMLA. Humans have a difficult time making sense of disparate data streams, potentially across multiple files, that are integrated using complex functions. We are much better at interpreting data when it has been formatted to show the relevant relationships between modalities, learners, actions, and events in a learning environment. And we completely agree that data fusion representations that integrate multiple data modalities are essential to making grounded interpretations of multimodal data.

However, we argue that it is better to construct these visualizations dynamically *based on an underlying set of functions* that describe the relationships in the multimodal data. This is a set of techniques we are working on actively, as are other researchers who are constructing interpretable representations of learning analytic results (Martinez-Maldonado et al., 2020; Buckingham Shum et al., 2019).

We thus suggest that, compared to existing methods of data fusion, TMA models that use data transfusion can provide a more nuanced and more accurate view of learning processes—as well as the capability to systematically account for unique characteristics of diverse learners and groups of learners. This, in turn, will allow researchers using MMLA to expand our understanding of effective learning processes and more appropriately address questions of equity in multimodal learning.

To be clear, existing SMLs provide *some* of the features needed to create such models, but no technique provides all the functionality that a TMA approach offers. TMA lets researchers apply existing analysis techniques to multimodal data without having to develop bespoke models—that is, without having to create custom data formats and analysis tools for every dataset.

### 4.2 Limitations

Of course, we recognize that this initial description and exploration of TMA has a number of significant limitations.

First, and most obviously, we have not in any way tested whether the specific functions described here are the ideal mathematical representation for relationships across different forms of multimodal data. We anticipate that as researchers use TMA, they will suggest other families of functions, other parameters, and other properties of multimodal data that need to be represented in order to construct appropriate multimodal grounding functions. Indeed, it would be quite surprising if this were not the case.

Second, and almost equally obviously, the worked example we have provided here would be woefully inadequate if we were using it to definitively demonstrate the utility of TMA as a methodological approach. We looked at only one dataset, using only one learning analytic approach. We did little to systematically compare TMA to a range of data formatting solutions. But, as we hope has been clear, our example is intended to be illustrative rather than definitive. More work is clearly needed to systematically test the relative efficacy of TMA compared to other approaches to data fusion in MMLA—to identify which approaches work best and under what conditions.

Third, we recognize that implementing TMA is currently difficult. However, we anticipate that over time, tools will be developed to (a) integrate TMA into multiple families of SML functions, (b) construct human-readable representations of TMA functions through multimodal data formatting, (c) simplify the process of parameter estimation, and (d) provide clear guidelines as to whether and when TMA provides advantages over other modelling techniques.

Finally, and perhaps most importantly, just as there are surely learning contexts in which not all data modalities add value to models of student learning, we anticipate that there will be some—and perhaps many—situations in which existing data fusion methods are sufficient to answer research questions, and even where existing data fusion methods may outperform TMA. The fact that we can account for complex interactions in students' learning processes does not always mean that we should.

However, we argue that as the field of LA becomes (appropriately) more concerned with questions of equity, it is good practice to consider (and ideally test) differential impacts at the group or individual level before constructing a simpler model. Thus, the properties of iterativity, clarity, and flexibility that TMA provides in representing hypotheses about the structure of discourse may provide advantages in modelling unimodal data as well as multimodal data.

## 5. Conclusion

We are clearly at the very early stages in research and development of TMA as an analytic technique—or, more properly, as a set of analytic techniques that use data transfusion to augment existing models of learning processes. There are clear

limitations to this first explication of the method and the worked example we have provided. However, we believe that TMA has the potential to significantly expand researchers' ability to analyze multimodal learning processes. If the promise of TMA is realized, it could improve the speed, efficiency, transparency, and fairness of multimodal analyses of learning processes.

## Declaration of Conflicting Interest

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

## Funding

This work was funded in part by the National Science Foundation (DRL-2100320, DRL-2201723, DRL-2225240, DRL-2405238), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

## References

Ahmad, S., Fatima, R., Mazhar, S. S., Bajpai, S., Yadav, R. R., & Kanaujia, D. S. (2023). Assessing the linkage between vocational education and economic growth using autoregression analysis: Evidence from India. *Journal of Namibian Studies: History Politics Culture*, *35*, 434–449.

Airey, J., & Linder, C. (2009). A disciplinary discourse perspective on university science learning: Achieving fluency in a critical constellation of modes. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(1), 27–49. https://doi.org/10.1002/tea.20265

Akçapınar, G., & Hasnine, M. N. (2022). Discovering the effects of learning analytics dashboard on students' behavioral patterns using differential sequence mining. *Procedia Computer Science*, *207*, 3818–3825. https://doi.org/10.1016/j.procs.2022.09.443

Al-Adwan, A. S., Albelbisi, N. A., Hujran, O., Al-Rahmi, W. M., & Alkhalifah, A. (2021). Developing a holistic success model for sustainable e-learning: A structural equation modeling approach. *Sustainability*, *13*(16), 9453. https://doi.org/10.3390/su13169453

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247–286. https://doi.org/10.1080/10508406.2011.611446

Alkabbany, I., Ali, A. M., Foreman, C., Tretter, T., Hindy, N., & Farag, A. (2023). An experimental platform for real-time students engagement measurements from video in STEM classrooms. *Sensors*, *23*(3), 1614. https://doi.org/10.3390/s23031614

Almond, R., Kim, Y. J., Shute, V. J., & Ventura, M. (2013). Debugging the evidence chain. In *Proceedings of the 2013 UAI Application Workshops: Big Data Meet Complex Models and Models for Spatial, Temporal and Network Data*, 15 July 2013, Bellevue, Washington, USA. Association for Uncertainty in Artificial Intelligence. http://purl.flvc.org/fsu/fd/FSU_libsubv1_scholarship_submission_1472579448

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, *141*, 103612. https://doi.org/10.1016/j.compedu.2019.103612

Alwahaby, H., & Cukurova, M. (2022). The ethical implications of using multimodal learning analytics: Towards an ethical research and practice framework. *EdArXiv preprint*. https://doi.org/10.35542/osf.io/4znby

Anderson, L. A., & Paulus, T. M. (2021). Secondary qualitative analysis in the family sciences. *Family and Consumer Sciences Research Journal*, *49*(4), 362–375. https://doi.org/10.1111/fcsr.12403

Ba, S., Hu, X., Stein, D., & Liu, Q. (2023). Assessing cognitive presence in online inquiry-based discussion through text classification and epistemic network analysis. *British Journal of Educational Technology*, *54*(1), 247–266. https://doi.org/10.1111/bjet.13285

Balogh, Z., & Kuchárik, M. (2019). Predicting student grades based on their usage of LMS Moodle using Petri nets. *Applied Sciences*, *9*(20), 4211. https://doi.org/10.3390/app9204211

Bowman, D., Swiecki, Z., Cai, Z., Wang, Y., Eagan, B., Linderoth, J., & Shaffer, D. W. (2021). The mathematical foundations of epistemic network analysis. In *Advances in quantitative ethnography* (pp. 91–105). Springer International Publishing. https://doi.org/10.1007/978-3-030-67788-6_7

Buckingham Shum, S., Echeverria, V., & Martinez-Maldonado, R. (2019). The multimodal matrix as a quantitative ethnography methodology. In A. Ruis & S. Lee (Eds.), *Advances in quantitative ethnography. ICQE 2021. Communications in computer and information science* (pp. 26–40, Vol. 1312). Springer International. https://doi.org/10.1007/978-3-030-33232-7_3

Buigut, S. K., & Valev, N. T. (2005). Is the proposed East African Monetary Union an optimal currency area? A structural vector autoregression analysis. *World Development*, *33*(12), 2119–2133. https://doi.org/10.1016/j.worlddev.2005.06.006

Cagiltay, B., Ho, H.-R., Michaelis, J. E., & Mutlu, B. (2020). Investigating family perceptions and design preferences for an in-home robot. In *Proceedings of the Interaction Design and Children Conference* (IDC 2020), 21–24 June 2020, London, UK (pp. 229–242). ACM. https://doi.org/10.1145/3392063.3394411

Chan, K. I., Tse, R., & Lei, P. I. (2022). Tracing students' learning performance on multiple skills using Bayesian methods. In *Proceedings of the Sixth International Conference on Education and Multimedia Technology* (ICEMT 2022), 13–15 July 2022, Guangzhou, China (pp. 84–89). ACM. https://doi.org/10.1145/3551708.3556202

Chan, M. C. E., Ochoa, X., & Clarke, D. (2019). Multimodal learning analytics in a laboratory classroom. In M. Virvou, E. Alepis, G. Tsihrintzis, & L. Jain (Eds.), *Machine learning paradigms* (pp. 131–156, Vol. 158). Springer International. https://doi.org/10.1007/978-3-030-13743-4_8

Chango, W., Cerezo, R., & Romero, C. (2021). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Computers & Electrical Engineering*, *89*, 106908. https://doi.org/10.1016/j.compeleceng.2020.106908

Chatfield, S. L. (2020). Recommendations for secondary analysis of qualitative data. *The Qualitative Report*, *25*(3), 833–842. https://doi.org/10.46743/2160-3715/2020.4092

Chen, F., & Lu, C. (2022). Learning outcome modeling in computer-based assessments for learning. In F. Ouyang, P. Jiao, B. M. McLaren, & A. H. Alavi (Eds.), *Artificial intelligence in STEM education* (pp. 175–194). CRC Press. https://doi.org/10.1201/9781003181187-15

Chen, K.-Z., & Li, S.-C. (2021). Sequential, typological, and academic dynamics of self-regulated learners: Learning analytics of an undergraduate chemistry online course. *Computers and Education: Artificial Intelligence*, *2*, 100024. https://doi.org/10.1016/j.caeai.2021.100024

Chen, L., Feng, G., Leong, C. W., Joe, J., Kitchen, C., & Lee, C. M. (2016). Designing an automated assessment of public speaking skills using multimodal cues. *Journal of Learning Analytics*, *3*(2), 261–281. https://doi.org/10.18608/jla.2016.32.13

Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Williamson Shaffer, D. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of Biomechanical Engineering*, *137*(2), 024701. https://doi.org/10.1115/1.4029235

Chiu, M. M. (2018). Statistically modelling effects of dynamic processes on outcomes: An example of discourse sequences and group solutions. *Journal of Learning Analytics*, *5*(1), 75–91. https://doi.org/10.18608/jla.2018.51.6

Choi, Y., & Cho, Y. I. (2020). Learning analytics using social network analysis and Bayesian network analysis in sustainable computer-based formative assessment system. *Sustainability*, *12*(19), 7950. https://doi.org/10.3390/su12197950

Chouldechova, A., & Roth, A. (2018). The frontiers of fairness in machine learning. *arXiv preprint arXiv:1810.08810*. https://doi.org/10.48550/arXiv.1810.08810

Cloude, E. B., Azevedo, R., Winne, P. H., Biswas, G., & Jang, E. E. (2022). System design for using multimodal trace data in modeling self-regulated learning. *Frontiers in Education*, *7*, 928632. https://doi.org/10.3389/feduc.2022.928632

Conijn, R., Van Waes, L., & van Zaanen, M. (2020). Human-centered design of a dashboard on students' revisions during writing. In C. Alario-Hoyos, M. Rodríguez-Triana, M. Scheffel, I. Arnedillo-Sánchez, & S. Dennerlein (Eds.), *Addressing global challenges and quality education. EC-TEL 2020. Lecture notes in computer science* (pp. 30–44, Vol. 12315). Springer International. https://doi.org/10.1007/978-3-030-57717-9_3

Csanadi, A., Eagan, B., Kollar, I., Shaffer, D. W., & Fischer, F. (2018). When coding-and-counting is not enough: using epistemic network analysis (ENA) to analyze verbal data in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, *13*(4), 419–438. https://doi.org/10.1007/s11412-018-9292-z

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology*, *51*(5), 1441–1449. https://doi.org/10.1111/bjet.13015

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, *8*(1). https://doi.org/10.1186/1744-9081-8-33

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 338–349. https://doi.org/10.1111/jcal.12288

Dominguez, F., Ochoa, X., Zambrano, D., Camacho, K., & Castells, J. (2021). Scaling and adopting a multimodal learning analytics application in an institution-wide setting. *IEEE Transactions on Learning Technologies*, *14*(3), 400–414. https://doi.org/10.1109/tlt.2021.3100778

Dorodchi, M., Mahzoon, M. J., Maher, M., & Benedict, A. (2020). A learning analytics approach to assessing student risk in active learning. In J. Keith-Le & M. Morgan (Eds.), *Faculty experiences in active learning* (pp. 86–100). UNC Press. https://omp.charlotte.edu/library/catalog/download/9/13/243?inline=1

Dowell, N. M. M., Nixon, T. M., & Graesser, A. C. (2018). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, *51*(3), 1007–1041. https://doi.org/10.3758/s13428-018-1102-z

Eagan, B., Rogers, B., Pozen, R., Marquart, C., & Shaffer, D. W. (2019). rhoR: Rho for inter rater reliability. *R package version*, *1*(0.0). https://cran.r-project.org/web/packages/rhoR/rhoR.pdf

Eagan, B., Swiecki, Z., Farrell, C., & Shaffer, D. W. (2019). The binary replicate test: Determining the sensitivity of CSCL models to coding error. In K. Lund, G. P. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *A wide lens: Combining embodied, enactive, extended, and embedded learning in collaborative settings, 13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019* (pp. 328–335, Vol. 1). International Society of the Learning Sciences (ISLS). https://repository.isls.org//handle/1/4421

Emerson, A., Henderson, N., Rowe, J., Min, W., Lee, S., Minogue, J., & Lester, J. (2020). Investigating visitor engagement in interactive science museum exhibits with multimodal Bayesian hierarchical models. In I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial intelligence in education. AIED 2020. Lecture notes in computer science* (pp. 165–176, Vol. 12163). Springer International. https://doi.org/10.1007/978-3-030-52237-7_14

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37. https://doi.org/10.1609/aimag.v17i3.1230

Fincham, E., Whitelock-Wainwright, A., Kovanović, V., Joksimović, S., van Staalduinen, J.-P., & Gašević, D. (2019). Counting clicks is not enough: Validating a theorized model of engagement in learning analytics. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 501–510). ACM. https://doi.org/10.1145/3303772.3303775

Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 225–234). ACM. https://doi.org/10.1145/3303772.3303791

Gee, J. P. (2015). Discourse, small d, big D. In K. Tracy (Ed.), *The international encyclopedia of language and social interaction*. Wiley. https://doi.org/10.1002/9781118611463.wbielsi016

González-Brambila, S., Sánchez-Guerrero, L., González-Beltrán, B., & Figueroa-González, J. (2021). Using the trajectories analysis for determining computer engineering students' risks at an superior educational institution. In L. Gomez Chova, A. Lopez, & I. Candel Torres (Eds.), *Proceedings of the 14th Annual International Conference of Education, Research and Innovation* (ICERI 2021), 8–9 November 2021, online (pp. 4441–4445). IATED. https://doi.org/10.21125/iceri.2021.1023

Gupta, A., Garg, D., & Kumar, P. (2022). Mining sequential learning trajectories with hidden Markov models for early prediction of at-risk students in e-learning environments. *IEEE Transactions on Learning Technologies*, *15*(6), 783–797. https://doi.org/10.1109/tlt.2022.3197486

Huang, L., Zheng, J., Lajoie, S. P., Chen, Y., Hmelo-Silver, C. E., & Wang, M. (2023). Examining university teachers' self-regulation in using a learning analytics dashboard for online collaboration. *Education and Information Technologies*, *29*(7), 8523–8547. https://doi.org/10.1007/s10639-023-12131-7

Hung, Y.-W., & Higgins, S. (2015). Learners' use of communication strategies in text-based and video-based synchronous computer-mediated communication environments: Opportunities for language learning. *Computer Assisted Language Learning*, *29*(5), 901–924. https://doi.org/10.1080/09588221.2015.1074589

Hutchins, E. (1995). *Cognition in the wild*. MIT press. https://doi.org/10.7551/mitpress/1881.001.0001

Jeng, H.-L., & Chung-Nien, C. (2022). Where can we find the differences between experts and novices with lag sequential analysis of spatial behavioral patterns in digital pentomino games. *Journal of Research in Education Sciences*, *67*(4), 105–142. https://doi.org/10.6209/JORIES.202212_67(4).0004

Jewitt, C., Kress, G., Ogborn, J., & Tsatsarelis, C. (2001). Exploring learning through visual, actional and linguistic communication: The multimodal environment of a science classroom. *Educational Review*, *53*(1), 5–18. https://doi.org/10.1080/00131910123753

Jiang, S., Huang, X., Sung, S. H., & Xie, C. (2023). Learning analytics for assessing hands-on laboratory skills in science classrooms using Bayesian network analysis. *Research in Science Education*, *53*(2), 425–444. https://doi.org/10.1007/s11165-022-10061-x

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. https://doi.org/10.18608/jla.2017.43.2

Kokoç, M., Akçapınar, G., & Hasnine, M. N. (2021). Unfolding students' online assignment submission behavioral patterns using temporal learning analytics. *Educational Technology & Society*, *24*(1), 223–235. https://www.jstor.org/stable/26977869

Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal data fusion: An overview of methods, challenges, and prospects. *Proceedings of the IEEE*, *103*(9), 1449–1477. https://doi.org/10.1109/jproc.2015.2460697

Lämsä, J., Hämäläinen, R., Koskinen, P., Viiri, J., & Mannonen, J. (2020). The potential of temporal analysis: Combining log data and lag sequential analysis to investigate temporal differences between scaffolded and non-scaffolded group inquiry-based learning processes. *Computers & Education*, *143*, 103674. https://doi.org/10.1016/j.compedu.2019.103674

Larmuseau, C., Cornelis, J., Lancieri, L., Desmet, P., & Depaepe, F. (2020). Multimodal learning analytics to investigate cognitive load during online problem solving. *British Journal of Educational Technology*, *51*(5), 1548–1562. https://doi.org/10.1111/bjet.12958

Lee, M. P., Croteau, E., Gurung, A., Botelho, A. F., & Heffernan, N. T. (2023). Knowledge tracing over time: A longitudinal analysis. In M. Feng, T. Käser, & P. Talukdar (Eds.), *Proceedings of the 16th International Conference on Educational Data Mining* (EDM 2023), 11–14 July 2023, Bengaluru, India. International Educational Data Mining Society. https://doi.org/10.5281/zenodo.8115788

Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin & R. Veel (Eds.), *Reading science: Critical and functional perspectives on discourses of science* (pp. 87–113). Routledge.

Lewandowski, L. J., Codding, R. S., Kleinmann, A. E., & Tucker, K. L. (2003). Assessment of reading rate in postsecondary students. *Journal of Psychoeducational Assessment*, *21*(2), 134–144. https://doi.org/10.1177/073428290302100202

Liu, F., Fan, Z., Huang, F., Li, Y., He, Y., & Hu, W. (2022). Modeling learner behavior analysis based on educational big data and dynamic Bayesian network. In *2022 IEEE Eighth International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS)*, 6–8 May 2022, Jinan, China (pp. 48–53). IEEE. https://doi.org/10.1109/bigdatasecurityhpscids54978.2022.00019

Liu, R., Stamper, J., Davenport, J., Crossley, S., McNamara, D., Nzinga, K., & Sherin, B. (2019). Learning linkages: Integrating data streams of multiple modalities and timescales. *Journal of Computer Assisted Learning*, *35*(1), 99–109. https://doi.org/10.1111/jcal.12315

Liu, X. (2022). Primary science curriculum student acceptance of blended learning: Structural equation modeling and visual analytics. *Journal of Computers in Education*, *9*(3), 351–377. https://doi.org/10.1007/s40692-021-00206-8

Ma, Y., Celepkolu, M., & Boyer, K. E. (2022). Detecting impasse during collaborative problem solving with multimodal learning analytics. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 45–55). ACM. https://doi.org/10.1145/3506860.3506865

Marquart, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2019). *ncodeR: Techniques for automated classifiers*. University of Wisconsin–Madison. https://cran.r-project.org/web/packages/ncodeR/index.html

Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., & Buckingham Shum, S. (2020). From data to insights: A layered storytelling approach for multimodal learning analytics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI 2020), 25–30 April 2020, Honolulu, Hawaii, USA (pp. 1–15). ACM. https://doi.org/10.1145/3313831.3376148

Mayfield, E., Madaio, M., Prabhumoye, S., Gerritsen, D., McLaughlin, B., Dixon-Román, E., & Black, A. W. (2019). Equity beyond bias in language technologies for education. In H. Yannakoudakis, E. Kochmar, C. Leacock, N. Madnani, I. Pilán, & T. Zesch (Eds.), *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications*, 2 August 2019, Florence, Italy (pp. 444–460). Association for Computational Linguistics. https://doi.org/10.18653/v1/w19-4446

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Michaelis, J. E., & Mutlu, B. (2018). Reading socially: Transforming the in-home reading experience with a learning-companion robot. *Science Robotics*, *3*(21). https://doi.org/10.1126/scirobotics.aat5999

Michaelis, J. E., & Mutlu, B. (2019). Supporting interest in science learning with a social robot. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (IDC 2019), 12–15 June 2019, Boise, Idaho, USA (pp. 71–82). ACM. https://doi.org/10.1145/3311927.3323154

Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012). Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In V. Farrell, G. Farrell, C. Chua, W. Huang, R. Vasa, & C. Woodward (Eds.), *Proceedings of the 24th Australian Computer-Human Interaction Conference* (OzCHI 2012), 26–30 November 2012, Melbourne, Australia (pp. 420–423). ACM. https://doi.org/10.1145/2414536.2414602

Ochoa, X. (2022). Multimodal learning analytics—Rationale, process, examples, and direction. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), *The handbook of learning analytics* (pp. 54–65). SoLAR. https://doi.org/10.18608/hla22.006

Ochoa, X., Chiluiza, K., Granda, R., Falcones, G., Castells, J., & Guamán, B. (2018). Multimodal transcript of face-to-face group-work activity around interactive tabletops. In *Companion Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (CrossMMLA@ LAK 2018), 7–9 March 2018, Sydney, New South Wales, Australia. ACM. https://ceur-ws.org/Vol-2163/paper4.pdf

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (ICMI 2013), 9–13 December 2013, Sydney, Australia (pp. 583–590). ACM. https://doi.org/10.1145/2522848.2533789

Ochoa, X., & Worsley, M. (2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219. https://doi.org/10.18608/jla.2016.32.10

Poitras, E. G., Behnagh, R. F., & Bouchet, F. (2020). A dimensionality reduction method for time series analysis of student behavior to predict dropout in massive open online courses. In D. Ifenthaler & D. Gibson (Eds.), *Adoption of data analytics in higher education learning and teaching* (pp. 391–406). Springer. https://doi.org/10.1007/978-3-030-47392-1_20

Porter, C., Donegan, S., Eagan, B., Geröly, A., Jeney, A., Jiao, S., Peters, G.-J., & Zörgő, S. (2021). A systematic review of quantitative ethnography methods. In B. Wasson & S. Zörgő (Eds.), *Second International Conference on Quantitative Ethnography: Conference Proceedings Supplement*, 6–11 November 2021, online (pp. 35–38). International Society for Quantitative Ethnography. https://doi.org/10.31235/osf.io/erbt5

Priestley, M. (1980). State-dependent models: A general approach to non-linear time series analysis. *Journal of Time Series Analysis*, *1*(1), 47–71. https://doi.org/10.1111/j.1467-9892.1980.tb00300.x

Raca, M., & Dillenbourg, P. (2013). System for assessing classroom attention. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK 2013), 8–13 April 2013, Leuven, Belgium (pp. 265–269). ACM. https://doi.org/10.1145/2460296.2460351

Rahmani, D., & Fay, D. (2022). A state-dependent linear recurrent formula with application to time series with structural breaks. *Journal of Forecasting*, *41*(1), 43–63. https://doi.org/10.1002/for.2778

Reilly, J. M., & Dede, C. (2019). Differences in student trajectories via filtered time series analysis in an immersive virtual world. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 130–134). ACM. https://doi.org/10.1145/3303772.3303832

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*, 239–257. https://doi.org/10.1007/s11412-009-9070-z

Reimann, P., Markauskaite, L., & Bannert, M. (2014). E-research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528–540. https://doi.org/10.1111/bjet.12146

Reimann, P., Yacef, K., & Kay, J. (2011). Analyzing collaborative interactions with data mining methods for the benefit of learning. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues. Computer-supported collaborative learning series* (pp. 161–185, Vol. 12). Springer. https://doi.org/10.1007/978-1-4419-7710-6_8

Rivers, K., Harpstead, E., & Koedinger, K. R. (2016). Learning curve analysis for programming: Which concepts do students struggle with? In *Proceedings of the 2016 ACM Conference on International Computing Education Research* (ICER 2016), 8–12 September 2016, Melbourne, Australia (pp. 143–151, Vol. 16). ACM. https://doi.org/10.1145/2960310.2960333

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*(4), 734. https://doi.org/10.1037//0033-295x.103.4.734

Ruis, A., Siebert-Evenstone, A., Pozen, R., Eagan, B., & Shaffer, D. W. (2019). Finding common ground: A method for measuring recent temporal context in analyses of complex, collaborative thinking. In K. Lund, G. P. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning* (CSCL 2019), 17–21 June 2019, Lyon, France (Vol. 1). International Society of the Learning Sciences. https://repository.isls.org//handle/1/4395

Runyan, M. K. (1991). The effect of extra time on reading comprehension scores for university students with and without learning disabilities. *Journal of Learning Disabilities*, *24*(2), 104–108. https://doi.org/10.1177/002221949102400207

Schneider, B., & Blikstein, P. (2015). Unraveling students' interaction around a tangible interface using multimodal learning analytics. *Journal of Educational Data Mining*, *7*(3), 89–116. https://doi.org/10.5281/zenodo.3554729

Shaffer, D. W. (2017). *Quantitative ethnography*. Lulu.com. https://books.google.com/books/about/Quantitative_Ethnography.html?id=H-iMDwAAQBAJ

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. https://doi.org/10.18608/jla.2016.33.3

Shaffer, D. W., & Ruis, A. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics*. SoLAR. https://doi.org/10.18608/hla17.015

Shaffer, D. W., & Ruis, A. R. (2021). How we code. In A. Ruis & S. Lee (Eds.), *Advances in quantitative ethnography. ICQE 2021. Communications in computer and information science* (pp. 62–77, Vol. 1312). Springer. https://doi.org/10.1007/978-3-030-67788-6_5

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, *51*(5), 1450–1484. https://doi.org/10.1111/bjet.12993

Sharma, K., Papamitsiou, Z., & Giannakos, M. N. (2019). Modelling learners' behaviour: A novel approach using GARCH with multimodal data. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Transforming learning with meaningful technologies. EC-TEL 2019. Lecture notes in computer science* (pp. 450–465, Vol. 11722). Springer. https://doi.org/10.1007/978-3-030-29736-7_34

Siebert-Evenstone, A. L., Irgens, G. A., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2017). In search of conversational grain size: Modeling semantic structure using moving stanza windows. *Journal of Learning Analytics*, *4*(3), 123–139. https://doi.org/10.18608/jla.2017.43.7

Spätgens, T., & Schoonen, R. (2019). Individual differences in reading comprehension in monolingual and bilingual children: The influence of semantic priming during sentence reading. *Learning and Individual Differences*, *76*, 101777. https://doi.org/10.1016/j.lindif.2019.101777

Sümer, Ö., Goldberg, P., D'Mello, S., Gerjets, P., Trautwein, U., & Kasneci, E. (2021). Multimodal engagement analysis from facial videos in the classroom. *IEEE Transactions on Affective Computing*, *14*(2), 1012–1027. https://doi.org/10.1109/taffc.2021.3127692

Sung, H., Cao, S., Ruis, A. R., & Shaffer, D. W. (2019). Reading for breadth, reading for depth: Understanding the relationship between reading and complex thinking using epistemic network analysis. In K. Lund, G. P. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning* (CSCL 2019), 17–21 June 2019, Lyon, France. International Society of the Learning Sciences. https://repository.isls.org//handle/1/4428

Sung, H., Swart, M. I., & Nathan, M. J. (2022). Methods for analyzing temporally entangled multimodal data. In A. Weinberger, W. Chen, D. Hernández-Leo, & B. Chen (Eds.), *Proceedings of the 15th International Conference on Computer-Supported Collaborative Learning* (CSCL 2022), 6–10 June 2022, Hiroshima, Japan, and online (pp. 242–249). International Society of the Learning Sciences. https://repository.isls.org//handle/1/8282

Swiecki, Z., Lian, Z., Ruis, A., & Shaffer, D. W. (2019). Does order matter? Investigating sequential and cotemporal models of collaboration. In K. Lund, G. P. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning* (CSCL 2019), 17–21 June 2019, Lyon, France (Vol. 1). International Society of the Learning Sciences. https://repository.isls.org//handle/1/1556

Tang, H., Dai, M., Yang, S., Du, X., Hung, J.-L., & Li, H. (2022). Using multimodal analytics to systemically investigate online collaborative problem-solving. *Distance Education*, *43*(2), 290–317. https://doi.org/10.1080/01587919.2022.2064824

Tang, K.-s., Delgado, C., & Moje, E. B. (2014). An integrative framework for the analysis of multiple and multimodal representations for meaning-making in science education. *Science Education*, *98*(2), 305–326. https://doi.org/10.1002/sce.21099

Teasley, S. D., Popov, V., Bae, J.-S., & Elkins, S. (2023). Self-regulated learning theory and epistemic network analysis: Understanding university students' use of a learning analytics dashboard. In T. Urdan & E. N. Gonida (Eds.), *Remembering the life, work, and influence of Stuart A. Karabenick: A legacy of research on self-regulation, help seeking, teacher motivation, and more* (pp. 215–240). Emerald Publishing Limited. https://doi.org/10.1108/s0749-742320230000022015

Thiyagarajan, G., & Prasanna, S. (2022). Process mining-based behavioral modeling of learners in self-paced learning environment. In K. Ray, A. Dixit, D. Adhikari, & R. Mathew (Eds.), *Proceedings of the 2nd international conference on signal and data processing. CSDP 2022. Lecture notes in electrical engineering* (pp. 121–132, Vol. 1026). Springer. https://doi.org/10.1007/978-981-99-1410-4_11

Thorne, S. (2013). Secondary qualitative data analysis. In C. Tatano Beck (Ed.), *Routledge international handbook of qualitative nursing research* (pp. 393–404). Routledge. https://doi.org/10.4324/9780203409527

Wang, Y., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2021). Simplification of epistemic networks using parsimonious removal with interpretive alignment. In A. Ruis & S. Lee (Eds.), *Advances in quantitative ethnography. ICQE 2021. Communications in computer and information science* (pp. 137–151, Vol. 1312). Springer. https://doi.org/10.1007/978-3-030-67788-6_10

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, *2*(2), 5–13. https://doi.org/10.18608/jla.2015.22.2

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, *25*, 731–739. https://doi.org/10.3758/bf03211316

Wong, T.-L., Zou, D., Cheng, G., Tang, J. K. T., Cai, Y., & Wang, F. L. (2021). Enhancing skill prediction through generalising Bayesian knowledge tracing. *International Journal of Mobile Learning and Organisation*, *15*(4), 358–373. https://doi.org/10.1504/ijmlo.2021.10040632

Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. In *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (MLA 2014), 12 November 2014, Istanbul, Türkiye (pp. 1–4). ACM. https://doi.org/10.1145/2666633.2666634

Worsley, M. (2022). Framing the future of multimodal learning analytics. In M. Giannakos, D. Spikol, D. Di Mitri, K. Sharma, X. Ochoa, & R. Hammad (Eds.), *The multimodal learning analytics handbook* (pp. 359–369). Springer. https://doi.org/10.1007/978-3-031-08076-0_14

Worsley, M., & Blikstein, P. (2011). What's an expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the Fourth International Conference on Educational Data Mining* (EDM 2011), 6–8 July 2011, Eindhoven, Netherlands (pp. 235–240). International Educational Data Mining Society. https://educationaldatamining.org/EDM2011/wp-content/uploads/proc/edm2011_paper18_short_Worsley.pdf

Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK 2013), 8–13 April 2013, Leuven, Belgium (pp. 94–101). ACM. https://doi.org/10.1145/2460296.2460315

Worsley, M., & Blikstein, P. (2014). Deciphering the practices and affordances of different reasoning strategies through multimodal learning analytics. In *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (MLA 2014), 12 November 2014, Istanbul, Türkiye (pp. 21–27). ACM. https://doi.org/10.1145/2666633.2666637

Yan, L., Zhao, L., Gasevic, D., & Martinez-Maldonado, R. (2022). Scalability, sustainability, and ethicality of multimodal learning analytics. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 13–23). ACM. https://doi.org/10.1145/3506860.3506862

Yusuf, A., Noor, N. M., & Bello, S. (2023). Using multimodal learning analytics to model students' learning behavior in animated programming classroom. *Education and Information Technologies*, 1–44. https://doi.org/10.1007/s10639-023-12079-8

Zhang, S., Gao, Q., Sun, M., Cai, Z., Li, H., Tang, Y., & Liu, Q. (2022). Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers & Education*, *183*, 104485. https://doi.org/10.1016/j.compedu.2022.104485

Zhou, G., Moulder, R. G., Sun, C., & D'Mello, S. K. (2022). Investigating temporal dynamics underlying successful collaborative problem solving behaviors with multilevel vector autoregression. In *Proceedings of the 15th International Conference on Educational Data Mining* (EDM 2022), 24–27 July 2022, Durham, UK. International Educational Data Mining Society. https://educationaldatamining.org/edm2022/proceedings/2022.EDM-long-papers.25/index.html

Zhou, Y., & Kang, J. (2023). Enriching multimodal data: A temporal approach to contextualize joint attention in collaborative problem-solving. *Journal of Learning Analytics*, *10*(3), 87–101. https://doi.org/10.18608/jla.2023.7989

# Appendices

## Appendix A: Worked Example of TMA

In this appendix, we present a worked example of TMA analysis using a TMA-enhanced version of *epistemic network analysis* (ENA) (Shaffer & Ruis, 2017; Shaffer et al., 2016; Bowman et al., 2021). ENA is a widely used tool for analyzing learning process data in the LA community (Porter et al., 2021).

We show the derivation of appropriate TMA functions—temporal influence (TI), learner impact (LI), and horizon functions—and then compare three different models from the existing data:

1. ENA applied to unimodal data,
2. ENA applied to multimodal data, and
3. T/ENA applied to multimodal data.

The comparison shows that T/ENA on multimodal data outperforms either of the other two models.

### A.1 Methods

#### A.1.1 Context

To provide an example of how TMA functions are derived and used, we use a dataset with two-mode data from a virtual internship, *RescuShell*, in which undergraduate engineering students work as interns at a fictional engineering company to design the robotic legs of a mechanical exoskeleton for search and rescue workers (Chesler et al., 2015).

This virtual internship simulates the engineering design process using an online work portal with text resources, simulated design tools, and a built-in chat interface for students to collaborate with their project teams. Information about the design problem comes from a set of technical reports and research briefs. These documents provide the technical knowledge needed to design, test, and evaluate the performance of a mechanical exoskeleton. Students integrate information from these documents and then collaborate with their team to make design decisions.

The internship was composed of two parts:

1. In the *discovery phase* of the simulation, students review and summarize research reports, create and evaluate device prototypes, and discuss design choices with teammates and a mentor.

   In other words, in this first phase of the virtual internship, students focus on understanding the different DEVICE COMPONENTS that can be used to manufacture an exoskeleton in terms of the DATA about device PERFORMANCE that they need to consider in their DESIGN process[3].

2. In a subsequent *design phase*, students create a final proposal that balances the requirements of prospective customers given the constraints of the possible device components.

   In other words, in this second phase of the virtual internship, students focus on understanding how to improve the PERFORMANCE of the device by considering both DESIGN reasoning and DATA from device prototypes.

These hypothesized differences in student activity between the two phases of the simulation are shown schematically in Figure 6.

#### A.1.2 Test Conditions and Hypotheses

For this worked example, we compare three conditions:

C1. ENA applied to the student chat messages,

C2. ENA applied to a low-level fusion of student chat messages and logfile data of resource usage, and

C3. TMA-enhanced ENA (T/ENA) applied to student chat messages and logfile data of resource usage.

We hypothesize that

H1. there will be differences in student activity between the discovery and design phases of the simulation,

H2. the ENA model with chat only (C1) will outperform the ENA model with chat and resources (C2), and

[3] In what follows, we use text in SMALL CAPS to denote *codes*—that is, critical aspects of the activity being modelled.

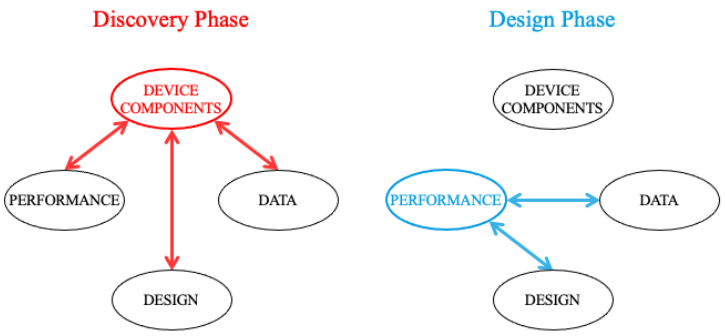


**Figure 6.** Two phases of the virtual internship *RescuShell* showing the key conceptual connections students were expected to make in each phase of the project.

H3. T/ENA (C3) will outperform both other conditions in modelling these differences.

In other words, we hypothesize that, *for this specific dataset*, (a) ENA on unimodal data will outperform ENA on multimodal data (low-level fusion) and (b) the TMA-enhanced version of ENA will outperform ENA, whether applied to a single data stream or to multimodal data—where by "outperform" we mean that these models will

(a) account for more variance,

(b) be more efficient, and

(c) have higher fidelity to the data.

#### A.1.3 Data

The dataset consists of (a) chat messages between students working in teams and (b) logfile data indicating when students have opened company memos that contain information about the simulation. The dataset contains a total of 3,613 chat messages and 3,029 reading events when students opened technical reports.

#### A.1.4 Coding

Our analysis of this data was based on a previously developed and validated coding scheme (Chesler et al., 2015; Eagan, Swiecki, et al., 2019)[4]. Chat data was segmented by utterance, defined as when a student sent a single message in the chat program. Technical reports included descriptions, graphs, images, and tables to provide background information about the design task. Because we could not track what part of a document students were reading at any one time—and because technical information was distributed throughout each document—we coded each technical report as a whole (Sung et al., 2019).

The codes used are shown in Table 5.

Automated classifiers for the chat logs were developed using the `ncodeR` R package (Marquart et al., 2019). `ncodeR` uses two inter-rater reliability statistics to establish code validity and reliability: Cohen's kappa and Shaffer's rho (Eagan, Rogers, et al., 2019; Shaffer, 2017)[5]. Concept validity of each code was assessed by requiring that two human raters achieve acceptable values of Cohen's kappa ($\kappa > 0.65$) with statistically significant values of rho ($\rho < 0.05$). Reliability was assessed by requiring that both human raters achieve acceptable values of kappa and rho compared to the automated classifier. For each code, all pairwise combinations of raters (humans and automated classifier) achieved $\kappa > 0.80$ and $\rho(0.65) < 0.05$, meaning that all codes were significantly above $\kappa = 0.65$. Resource documents were coded by two human raters using social moderation, meaning that both raters agreed on all codes for all resources.

[4]We changed the name of three of the codes in the coding scheme originally developed by Chesler and colleagues (2015). We changed PERFORMANCE PARAMETERS to PERFORMANCE, TECHNICAL CONSTRAINTS to DEVICE COMPONENTS, and DESIGN REASONING to DESIGN. We made these changes because we think that these code names more accurately reflect the definitions of the codes and thus are less confusing when describing the data and our analysis of it. As Shaffer and Ruis (2021) argue, the *definition* is the critical part of a code, and as we did not change the definitions of these codes, the original coding and inter-rater reliability statistics remain valid despite the change in names.

[5]The kappa statistic measures the agreement between two raters while accounting for agreement due to chance. Rho measures whether the level of agreement found for a sample coded by two raters generalizes to the rest of the dataset.

**Table 5.** Codebook for the *RescuShell* dataset.

| Code | Definition | Example |
|---|---|---|
| DESIGN | Referring to design development, prioritization, tradeoffs, and design decisions | "Aluminum and Composite are good options. Steel can carry a big load, but it is heavy and weighs down on the recharge interval, and it is a costly option." |
| PERFORMANCE | Referring to functional attributes of the device, such as payload, recharge interval, agility, safety, or cost | "My device has a pretty good safety, payload, agility, and recharge interval; the cost is a little high though." |
| DEVICE COMPONENTS | Referring to specific design choices associated with a prototype, such as actuators, ROM, materials, power sources, or sensors | "Our two best were both made with Aluminum, NiCd Batteries, Piezoelectric sensors, and Pneumatic actuators." |
| DATA | Referring to or justifying decisions based on numerical values, results tables, graphs, research papers, or relative quantities | "I thought that safety near the maximum was not very good (close to 225—one had 218 RPN), but other than that I was fine with the safety as long as it was around 200 or lower." |

#### A.1.5 Qualitative Analysis

To establish a baseline for interpreting the quantitative models, we conducted a secondary qualitative analysis (Chatfield, 2020; Anderson & Paulus, 2021), meaning that we reinterpreted the existing data. As Thorne (2013) and others argue, this means carefully considering the original context in which the data was collected as well as the methods of collection. In our case, we relied heavily on the existing analyses of the data, particularly Ruis and colleagues (2019)[6]. In particular, prior work looked at the multimodal relationship between chat messages and resource use by students (Sung et al., 2019) and showed qualitatively that there was a relationship between students' use of resources and their chat discussions. But this work did not model their interaction quantitatively.

#### A.1.6 Constructing TMA Functions

**Discourse Hypotheses** To construct a TMA model, we created a *discourse model*—that is, a set of hypotheses about the *structure of transmodal interactions* in the *RescuShell* learning environment. Specifically, we hypothesized that

- dH1: the impact of chats would diminish over time, but we did not know the rate;
- dH2: the impact of chats from mentors would diminish at a different rate than those from students, but we did not know by how much;
- dH3: following Runyan (1991), the impact of reading a resource would rise asymptotically while a student was reading, with the shape of the curve determined by the length of the document;
- dH4: once a document was closed, its influence would diminish, but we did not know the rate;
- dH5: a student's chat would influence every other student in the same discussion; and
- dH6: opening a resource would influence only the student who opened it.

These discourse hypotheses are summarized in Figure 7, which shows the functions $\ell_{student}(t_{chat}, teacherChat, \Delta t)$, $\ell_{student}(t_{chat}, studentChat, \Delta t)$, $t_{reading}$, $\hbar_{discussion,reading}$, and $\hbar_{discussion,chat}$.

[6] We note that several of the authors of this paper were involved in the design of the original learning environment, collection of data, and analyses on which our secondary qualitative analysis was based.

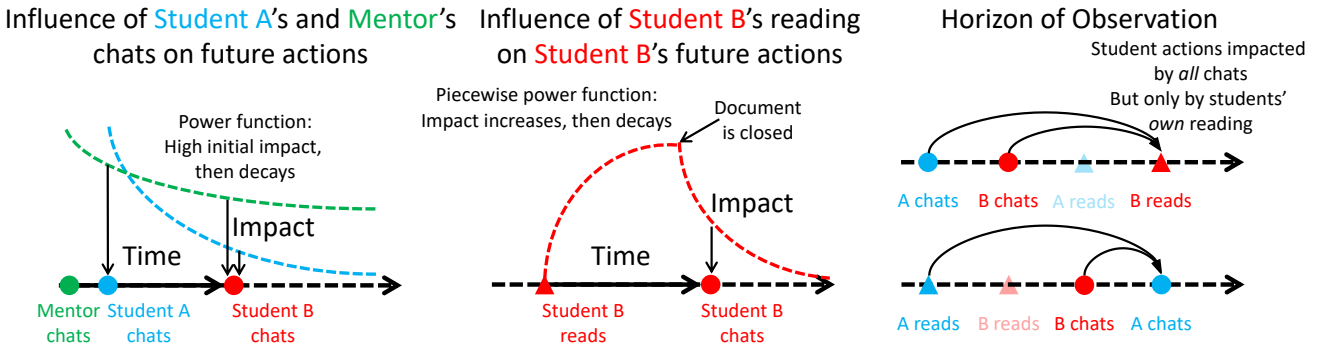


**Figure 7.** Modelling students' multimodal interactions in an engineering design simulation using TMA functions. Left are the LI and TI functions $\ell_{student}(t_{chat}, teacherChat, \Delta t)$, $\ell_{student}(t_{chat}, studentChat, \Delta t)$. Center is the TI function $t_{read}$. Right are the horizon functions $\hbar_{discussion,reading}$ and $\hbar_{discussion,chat}$. Functions are not illustrated on the same time scale.

### Determining Values for TMA Functions

**TI functions** Research on *forgetting* (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1997) suggests that it can be best modelled with a power function. Thus, the general form for TI functions can be written as a function of (a) the time difference $\Delta t_{ix}$ between any two events $e_i$ and $e_x$ and (b) a parameter $\theta$:

$$d_\theta(\Delta t_{ix}) = \frac{1}{\theta * \Delta t_{ix} + 1} \tag{10}$$

Based on this, we define a TI function for chats with a parameter $\theta_{chat}$:

$$t_{chat}(\Delta t_{ix}) = d_{\theta_{chat}}(\Delta t_{ix}) \tag{11}$$

To model the impact of a reading event $\boldsymbol{e}_i$ on future events $\boldsymbol{e}_x$ for each document, we estimated the average time a student would take to read the document, $\delta_{document}$, based on the number of words in the document and the average reading speed for post-secondary students of 200 words per minute (Lewandowski et al., 2003). Thus, a document of 500 words would have $\delta_{document} = \frac{500}{200/60} = 150$ seconds.

For each reading event $\boldsymbol{e}_i$, we identified the corresponding event $\boldsymbol{e}_c^i$ that indicated the *closing* of the document. Thus, the time a document remained open if it was opened at $t_i$ is $\Delta t_{ic}$.

We then constructed two TI functions for reading. We used $t_{whileReading}$ to model the impact of reading when $t_x < t_c$—that is, on events that take place while a student is still reading a document. We used $t_{afterReading}$ when $t_c < t_x$—that is, for events that take place after a student has closed the document.

Following Runyan (1991), $t_{whileReading}$ is a function of time difference, $\Delta t_{ix}$, with a parameter $\theta_{whileReading} = \frac{\mu}{\delta_{document}}$, where $\mu$ was determined empirically:

$$t_{whileReading}(\Delta t_{ix}) = 1 - d_{\theta_{whileReading}}(\Delta t_{ix}) \tag{12}$$

As a result, $t_{whileReading}$ represents a pattern of temporal influence for reading comprehension before closing the document: longer reading duration leads to more impact but levels off as the time spent reading approaches the average time it takes for a student to read the document.

For events after the document was closed ($t_c < t_x$), we used a parameter $\theta_{afterReading} = \frac{\sigma}{\delta_{document}}$, where $\sigma$ was determined empirically:

$$t_{afterReading}(\Delta t_{ix}) = \frac{d_{\theta_{afterReading}}(\Delta t_{ix}) * t_{whileReading}(\Delta t_{ic})}{d_{\theta_{afterReading}}(\Delta t_{ic})} \tag{13}$$

That is, $t_{afterReading}$ represents a pattern of temporal influence for decaying impact after closing the document: the highest impact occurs right after the document is closed, and then the impact diminishes as the current event gets further from the time the document was closed.

Using a random sample of 61 chat events and 78 reading events, we qualitatively determined, for each chat or reading event, the most distant event that had influenced it. We used an empirical IDP and fitted the parameters as follows[7]:

$$\begin{aligned} \theta_{chat} &= 0.0310 \\ \mu &= 8.8452 \\ \sigma &= 0.0263 \end{aligned} \tag{14}$$

The resulting $t_{chat}$ and $t_{reading}$ are plotted in Figure 8 using $t_{reading}$ for a document with 390 words that was read for 5 minutes. We had not expected a priori that the impact of reading would extend as long as the empirical IDP suggested, but students were often reading a document right before class or the day before class and then referring to it in the discussion, hence the long "tail" on the $t_{reading}$ function.

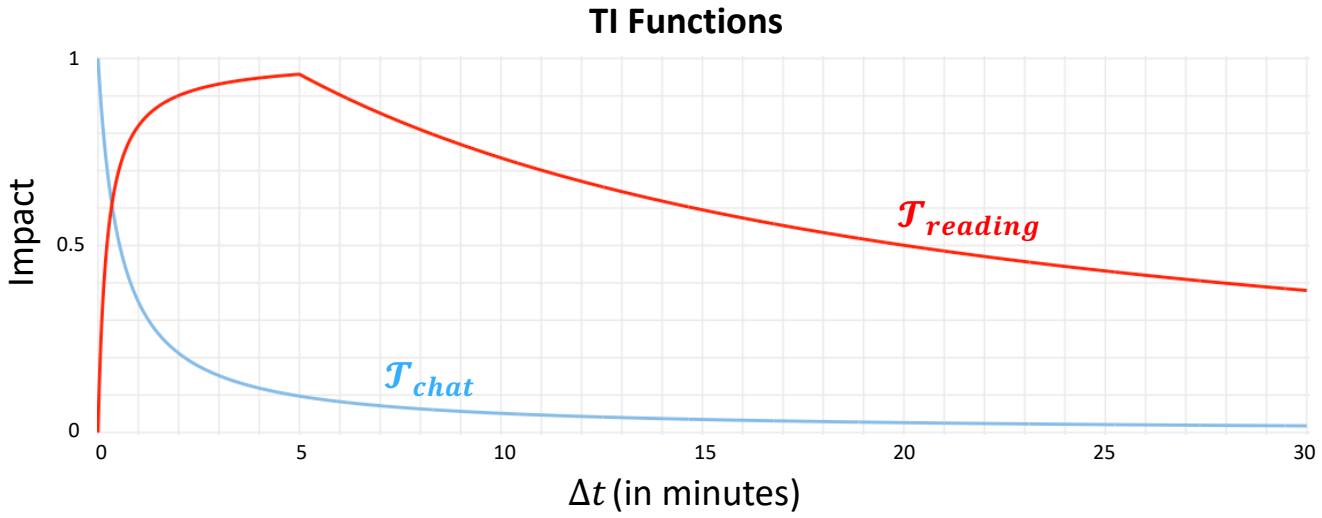


**Figure 8.** TI functions for chat messages and reading.

**LI functions** We tested design hypothesis dH2 (*the impact of chats from mentors would diminish at a different rate than chats from students*) by conducting a separate empirical IDP for mentor chats using the same method as for students. We found that while there was a small difference in the influence of mentor chats compared to student chats, the difference was not statistically significant and did not change the T/ENA models. We thus set the learner impact coefficients $v = \rho = 1$, which results in

$$\ell_{\boldsymbol{\lambda}_{teacher}}(t_{chat}, \dot{\boldsymbol{m}}_i, \Delta t_{ix}) = \ell_{\boldsymbol{\lambda}_{student}}(t_{chat}, \dot{\boldsymbol{m}}_i, \Delta t_{ix}) = t_{chat}(\Delta t_{ix}) \tag{15}$$

**Horizon functions** The horizon function for the influence of event $e_i$ on event $e_x$ was

$$\hbar_{\boldsymbol{\omega}_x}(\boldsymbol{m}_i) = \begin{cases} 0, & ((m_{i,eventType} = reading) \wedge (m_{i,student} \neq \omega_{x,student})) \\ 0, & (m_{i,group} \neq \omega_{x,group}) \\ 1, & \text{otherwise} \end{cases} \tag{16}$$

That is, event $\boldsymbol{e}_i$ will not impact event $\boldsymbol{e}_x$ if (a) $\boldsymbol{e}_i$ is a reading event from a student other than the student in $\boldsymbol{e}_x$ or (b) the students in $\boldsymbol{e}_i$ and $\boldsymbol{e}_x$ are not in the same group. Thus, events are only impacted by chats in the same group or reading by the same student.

[7]Thus, the set of TI functions, $\mathcal{T}$, can be written as

$$t_{chat}(\Delta t_{ix}) = \frac{1}{0.0310 * \Delta t_{ix} + 1} \text{ and } t_{reading}(\Delta t_{ix}) = \begin{cases} 1 - \frac{1}{\frac{8.8542}{\delta_{document}} * \Delta t_{ix} + 1}, & \Delta t_{ix} \leq \Delta t_{ic} \\ \frac{(\frac{0.0263}{\delta_{document}} * \Delta t_{ic}) * (\frac{8.8452}{\delta_{document}} * \Delta t_{ic} + 1)}{(\frac{0.0263}{\delta_{document}} * \Delta t_{ic} + 1) * (\frac{8.8452}{\delta_{document}} * \Delta t_{ix} + 1)}, & \Delta t_{ic} < \Delta t_{ix} \end{cases}$$

While this looks complex, these functions are, of course, implemented in R code, and, as shown in Figure 8, the results can be shown in interpretable form.

**ENA** In ENA, $\boldsymbol{c}_{ij}$ describes whether event $\boldsymbol{e}_i$ has some code $j$ associated with it. The *grounding function* for an event $\boldsymbol{e}_x$ is the sum of the code vectors $(\boldsymbol{c}_x + \boldsymbol{c}_{x-1} + \cdots + \boldsymbol{c}_{x-w+1})$ in a window of the $w$ lines preceding it in the student's discussion group $m_{xd}$:

$$G_{ENA}(\boldsymbol{e}_x) = \sum_{i|\boldsymbol{m}_{i,group}=\boldsymbol{m}_{x,group},(x-w)<i\leq x} \boldsymbol{c}_i \tag{17}$$

The estimation function for each unit of analysis is a cumulative adjacency matrix that represents the *connections* made by that student between codes the student uses and codes in the common ground.

More specifically, we have the following:

(a) The estimation function for each student $s$ in ENA produces a network $\Phi(s)$:

$$\Phi^s = \Phi(s) = \sum_{x|\boldsymbol{m}_{is}=s} G_{ENA}(\boldsymbol{e}_x)\,\boldsymbol{c}_x^T \tag{18}$$

(b) These matrices are normalized and then subjected to a dimensional reduction (in this example we used SVD) that represents each student with two *ENA scores*, which can be represented as a single point in a two-dimensional space as $(\hat{\Phi}^s_1, \hat{\Phi}^s_2)$.

(c) These points are then *co-registered* with the network graph representations for each student, such that the two-dimensional space can be interpreted in terms of the connections that are more or less present in student networks (Shaffer & Ruis, 2017; Shaffer, 2017)[8].

Each network, $\Phi^s$, is thus represented in two ways: first as a point $\hat{\Phi}^s$ in a projected metric space, and second as a weighted network graph. Critically, these two representations are placed in the same metric space. These coordinated representations make it possible to (a) interpret the meaning of each network's location in the projected space and (b) verify that the meaning of network positions is aligned with qualitative interpretations of the original data.

In this study, we constructed a *subtracted network graph* for each model by (a) constructing two networks for each student, one in the discovery phase ($\Phi^s_{discovery}$) and the other in the design phase ($\Phi^s_{design}$); (b) computing the mean of each of these networks across all students ($\bar{\Phi}_{discovery}$ and $\bar{\Phi}_{design}$); and (c) subtracting the two networks ($\Phi_\Delta = \bar{\Phi}_{discovery} - \bar{\Phi}_{design}$). We used the subtracted network graphs $\Phi_\Delta$ to compare model results to the qualitative analysis to determine how well the model from each condition represents the qualitative analysis. We then tested how well the points $\hat{\Phi}^s$ predicted which phase of the game the student was in.

**T/ENA** In T/ENA, the grounding function is not computed using a preset window but based on the functions above that describe the temporal influence of events in this setting:

$$G^T(\boldsymbol{e}_x) = \sum_{i|t_i\leq t_x} \hbar_{\boldsymbol{\omega}_x}(\ddot{\boldsymbol{m}}_i)\ell_{\boldsymbol{\lambda}_x}(t_{\tau_i}, \dot{\boldsymbol{m}}_i, \Delta t_{ix})\boldsymbol{c}_i \tag{19}$$

Because in our example $\ell_{\boldsymbol{\lambda}_x}(t_{\tau i}, \boldsymbol{m}, \Delta t_{ix}) = t_{\tau i}(\Delta t_{ix})$, the ground function reduces to

$$G^T_{ENA}(\boldsymbol{e}_x) = \sum_{i|t_i\leq t_x} \hbar_{\boldsymbol{\omega}_x}(\boldsymbol{m}_i)\, t_{\tau_i}(\Delta t_{ix})\, \boldsymbol{c}_i \tag{20}$$

### A.1.7 Model Construction

**ENA Model on Chat Messages** We modelled the collaborative discourse of 29 students by computing an ENA model for student chat messages using the standard ENA grounding function and a window length of seven, which was determined by previous research on the same data (Ruis et al., 2019).

[8] More specifically, we have the following:

(a) The elements $\Phi^s_{ij}$ of the matrix $\Phi(s)$ are normalized using the L2 norm (normalized to unit length) for each student to form a normalized matrix ${}^{\eta}\Phi^s_{ij}$.

(b) The normalized matrices are centralized so that $\forall i, \sum_s \frac{{}^{C}\Phi^s_i}{|S|} = 0$. That is, every ${}^{C}\Phi^s_i$ has zero mean over all students.

(c) The centralized matrices ${}^{C}\Phi^s$ are embedded in a high-dimensional space such that ${}^{H}\Phi^s_{i,(k^2-k)+j} = {}^{C}\Phi^s_{kj}$, and a dimensional reduction is performed using SVD, resulting in a two-dimensional representation $\hat{\Phi}^s$ for each student.

(d) The codes (the nodes of the networks) are placed in the space using an optimization algorithm, described in more detail in Bowman and colleagues (2021), to create two coordinated representations.

**ENA Models on Chat Messages and Logfile of Resources Accessed** To analyze the interdependence between reading and chatting in this task, we constructed a low-level fusion of chat messages and logfile data on resource access and sorted the fused dataset based on the timestamps of chat messages and readings opened. We again constructed an ENA model using a standard ENA grounding function with a window of seven events, regardless of the event type.

**T/ENA Model** We constructed a T/ENA model using the same estimation function and process as for the two ENA models. We substituted the TMA grounding function described above for the standard ENA grounding function used in the two ENA models.

#### A.1.8 Model Comparison

**Quantitative Comparison** Across the ENA and T/ENA models (conditions 1–3), we compared ENA scores (the location of points on the first two dimensions of the dimensional reduction) between students in the discovery and design phases of the project using a logistic regression of the form $logit\,(discovery) = \beta_0 + \beta_1\hat{\Phi}_1 + \beta_2\hat{\Phi}_2$.

ENA and T/ENA models were compared using the AIC and McFadden's pseudo $R^2$.

**Qualitative Comparison** To establish fidelity to the data, we evaluated the ENA and T/ENA models using *interpretive alignment* (Shaffer, 2017; Wang et al., 2021). In interpretive alignment, models are compared not merely on predictive accuracy or variance explained, but by three criteria:

1. Do the models show different conclusions about the conditions being tested?
2. Are the interpretations of the models different?
3. Which models best represent a qualitative analysis of the data?

### A.2 Worked Example of TMA: Results

#### A.2.1 Qualitative Analysis

As described above, we analyzed chat messages and students' resource use in each phase of the simulation. One of the key outcomes of this analysis was the importance of the linkages between DEVICE COMPONENTS, PERFORMANCE, DESIGN, and DATA during the discovery phase.

Table 6 shows a small portion of the data from one group during the discovery phase, showing how these codes are connected to one another in the discourse.

The ***chat*** components of the excerpt start when Omar responds (line 1) "sweet," closing out a previous conversation, after which there is a 2-minute pause before Mateo suggests (line 2) that they discuss a new stakeholder, "paulo henriquez [sic]." Mateo says Paulo is concerned with "Payload and safety." That is, they suggest discussing what this stakeholder wants in terms of two PERFORMANCE criteria.

Omar argues (line 3) that they should move to a different stakeholder ("Next is Benjamin"), again talking about the stakeholder in terms of two PERFORMANCE criteria. This is followed by some confusion, where Omar (lines 4–5) and Nia (line 6) are not sure as to whether "Benjamin" or "Laura" is the stakeholder in the simulation who is interested in "payload and recharge interval."

At this point (line 7), Mateo says definitively: "laura is cost and long battery life." Here, they are referring to the stakeholder's needs ("laura is . . . ") in terms of a specific DEVICE COMPONENT (the battery) and two PERFORMANCE criteria ("cost" and "long . . . life").

Thus, if we only consider the ***chat***, we can see Mateo talking about DEVICE COMPONENTS in the context of a long discussion of PERFORMANCE criteria.

However, if we look further back in the data, we can see that 6.5 minutes earlier, Mateo opened the reference material *The Effect of Exoskeleton Range of Motion on Pneumatic Artificial Muscle Actuated System Level Performance Metrics*, which they kept open for a total of 286 seconds, or 50% longer than the estimated time to read this resource.

Importantly, the document provides specific DATA about the PERFORMANCE of the exoskeleton to explain different DESIGN tradeoffs that the students should consider. In other words, Mateo's comment about DEVICE COMPONENTS was almost certainly informed by the DATA they had just been reading about and the DESIGN choices the DATA suggests.

As shown conceptually in Figure 9, in the context of the ***chat***, Mateo is connecting DEVICE COMPONENTS to the PERFORMANCE of the device. If we consider ***reading*** as well, then this turn of talk is also linking DEVICE COMPONENTS to DATA and DESIGN. Moreover, this relationship was only apparent when we considered the interaction between students' chat messages and resource use.

When we analyzed chat messages and students' resource use in the design phase, there were significant connections between PERFORMANCE, DATA, and DESIGN. For example, Table 7 shows one project team during the design phase thinking about the DESIGN tradeoffs they will make to produce a device that meets the PERFORMANCE objectives.

The discussion begins with the mentor reminding students (line 1) about how to submit their "batch" of prototype designs for testing. Almost immediately, Omar asks (line 2) if everyone agrees with the DESIGN choices for a prototype they have just

**Table 6.** Chat discussion among students in *RescuShell* during the discovery phase about which DEVICE COMPONENTS will produce the best PERFORMANCE.

| Line | Time | Student | Chat | Click | Codes |
|---|---|---|---|---|---|
| 0 | 18:21:45 | Mateo | | The Effect of Exoskeleton Range of Motion on Pneumatic Artificial Muscle Actuated System Level Performance Metrics<br>*Reading duration: 286 s*<br>*Expected time to complete: 182 s* | PERFORMANCE, DEVICE COMPONENTS, **DATA**, **DESIGN** |
| | ... | | | | |
| 1 | 18:25:17 | Omar | sweet | | |
| 2 | 18:27:15 | Mateo | are we on to paulo henriquez? Payload and safety | | PERFORMANCE |
| 3 | 18:27:23 | Omar | Next is Benjamin: Energy efficiency and Cost | | PERFORMANCE |
| 4 | 18:27:41 | Omar | hold on | | |
| 5 | 18:28:02 | Omar | isn't Laura Payload and Recharge Interval? | | PERFORMANCE |
| 6 | 18:28:22 | Nia | I have Benjamin as payload and recharge interval | | PERFORMANCE |
| **7** | **18:29:16** | **Mateo** | **laura is cost and long battery life** | | **PERFORMANCE, DEVICE COMPONENTS** |

been discussing.

The other students agree (lines 3–6), after which Omar (line 7) continues the discussion by starting to talk about the PERFORMANCE of a new prototype designed to maximize the agility of the exoskeleton they are designing: "for agility heres what I think."

Thus, if we only consider the ***chat***, we can see Omar talking about maximizing PERFORMANCE in the context of a longer discussion of DESIGN tradeoffs in the prototype designs they will test.

However, if we look further back in the data, we can see that 4 minutes earlier, Omar had opened the reference material *Actuator Descriptions and Technical Specifications*, which they had open for a total of 151 seconds, or a little less than the estimated time to read that resource.

Importantly, this document provides specific DATA about the PERFORMANCE of the actuators that move the exoskeleton. In other words, Omar's comment about device PERFORMANCE is almost certainly also informed by the DATA they had just been reading about.

That is, in the context of the ***chat***, Omar is connecting device PERFORMANCE to DESIGN tradeoffs about which prototypes to test. If we consider ***reading*** as well, then this turn of talk is also linking PERFORMANCE to DATA (see Figure 9). Moreover, this relationship was only apparent when we considered the interaction between students' chat messages and resource use.

In sum, our qualitative analysis suggested that, as intended in the curriculum design (see Figure 6 above), students in the discovery phase made connections between DEVICE COMPONENTS and DATA, DESIGN tradeoffs, and PERFORMANCE criteria. Students in the design phase made connections between device PERFORMANCE and DATA and DESIGN tradeoffs.

However, these connections were only apparent when considering both the chat messages *and* the logs that indicated what resources students were consulting as they went through these different phases of the design process.

#### A.2.2 Quantitative Models

**Condition 1: ENA on Chat Messages** Using the ENA model with chat messages only, a logistic regression with an outcome variable of the discovery versus design phases obtained an AIC of 61.45 and a McFadden's $R^2$ of 0.18. The ENA score on the first dimension was a significant predictor of the two phases ($\beta = 10.64$, $e^{\beta} = 41772.77$, $p = 0.0078$), but the second

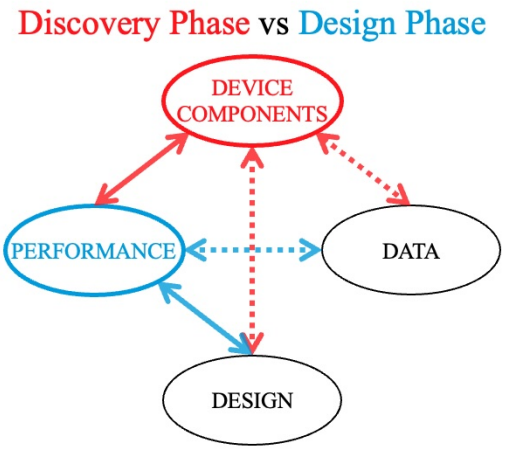


**Figure 9.** In the discovery phase of the simulation, students in *RescuShell* made connections from DEVICE COMPONENTS to PERFORMANCE in their ***chat*** conversations, but linkages from DEVICE COMPONENTS to DATA and DESIGN (shown as dashed lines) were only evident when we included ***reading*** data. In contrast, in the design phase students made connections from PERFORMANCE to DESIGN in their ***chat*** conversations, but linkages between PERFORMANCE and DATA (shown as dashed lines) were only evident when we included ***reading*** data.

dimension was not significant ($\beta = 2.69$, $e^{\beta} = 14.73$, $p = 0.2005$).

When we examine the subtracted network for the discovery and design phases (see Figure 10) compared to the design phase, students in the discovery phase made stronger connections between (a) DEVICE COMPONENTS and DATA, (b) DEVICE COMPONENTS and PERFORMANCE, and (c) DEVICE COMPONENTS and DESIGN. Compared to the discovery phase, students in the design phase made stronger connections between (d) PERFORMANCE and DESIGN, and (e) DESIGN and DATA, but *not* between (f) PERFORMANCE and DATA.

Referring to Figure 9, which illustrates the qualitative results, we can see that (a), (b), (c), and (d) are consistent with the qualitative analysis; however, (e) and (f) are not. Relative to our qualitative analysis, the ENA model on the chat messages underemphasizes connections between PERFORMANCE and DATA and overemphasizes connections between DESIGN and DATA in the design phase.

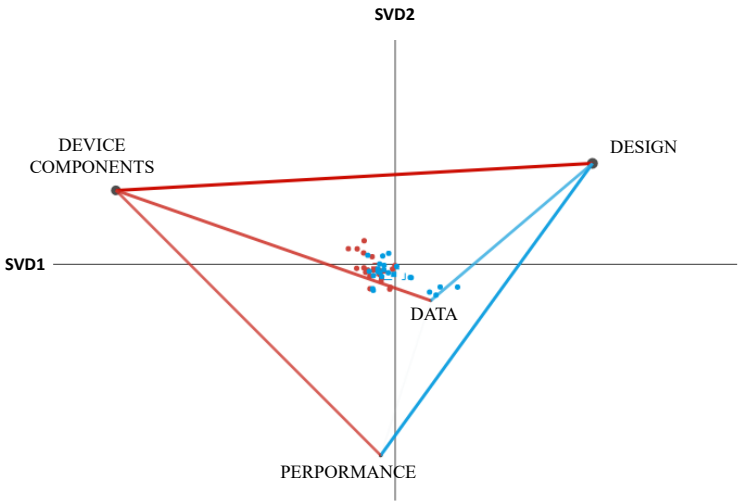


**Figure 10.** ENA model of the chat data in the *RescuShell* simulation. The subtracted network shows the differences between the means of the ENA networks for the discovery and design phases.

**Condition 2: ENA on Chat Messages and Resources** Using the ENA model with chat messages and resource clicks, a logistic regression with an outcome variable of the discovery versus design phases obtained an AIC of 66.58 and a McFadden's $R^2$ of 0.13. The ENA score on the first dimension was a significant predictor of two phases ($\beta = 14.39$, $e^{\beta} = 1776223$,

**Table 7.** Chat discussion among students in *RescuShell* during the design phase engaging in a discussion of DESIGN involved in achieving good PERFORMANCE for a prototype design they will test.

| Line | Time | Student | Chat | Click | Codes |
|---|---|---|---|---|---|
| 0 | 18:39:38 | Omar | | Actuator Descriptions and Technical Specifications<br>*Time read: 151 s*<br>*Expected time to complete: 179 s* | PERFORMANCE, DEVICE COMPONENTS, DESIGN, **DATA** |
| | ... | | | | |
| 1 | 18:42:20 | Mentor | Don't forget to include a link to your batch in your notebook entry. | | DESIGN |
| 2 | 18:42:23 | Omar | so everyone agree with the above design? | | DESIGN |
| 3 | 18:42:36 | Tariq | yea | | |
| 4 | 18:42:41 | Esmeralda | yes | | |
| 5 | 18:42:48 | Juan | yes? | | |
| 6 | 18:42:58 | Yolanda | yes | | |
| **7** | **18:43:53** | **Omar** | **so for agility heres what I think:** | | **PERFORMANCE** |

$p = 0.0167$), but the second dimension was not significant ($\beta = 8.59$, $e^{\beta} = 5377.61$, $p = 0.0887$).

When we examine the subtracted network for the discovery and design phases (see Figure 11), students in the discovery phase made stronger connections between (a) DEVICE COMPONENTS and DATA, and (b) DEVICE COMPONENTS and PERFORMANCE, but *not* between (c) DEVICE COMPONENTS and DESIGN. Students in the design phase made stronger connections between (d) PERFORMANCE and DESIGN, but there was also little difference between (e) phases on the connection between PERFORMANCE and DATA.

Once again referring to Figure 9, connections (a), (b), and (d) are consistent with the qualitative analysis; however, connections (c) and (e) are not. Relative to our qualitative analysis, the ENA model on chat messages and resources underemphasizes connections between DEVICE COMPONENTS and DESIGN in the discovery phase and underemphasizes connections between PERFORMANCE and DATA in the design phase.

**Condition 3: T/ENA Model on Chat Messages and Resources** Using the T/ENA model with chat messages and resource clicks, a logistic regression with an outcome variable of the discovery versus design phases obtained an AIC of 51.45 and a McFadden's $R^2$ of 0.32. ENA scores on both the first dimension ($\beta = 13.90$, $e^{\beta} = 1083384$, $p = 0.0109$) and the second dimension ($\beta = 10.46$, $e^{\beta} = 34891.55$, $p = 0.0072$) were significant predictors of the two phases.

When we examine the subtracted network for the discovery and design phases (see Figure 12), students in the discovery phase made stronger connections between (a) DEVICE COMPONENTS and DATA, (b) DEVICE COMPONENTS and PERFORMANCE, and (c) DEVICE COMPONENTS and DESIGN. Students in the design phase made stronger connections between (d) PERFORMANCE and DESIGN and (e) PERFORMANCE and DATA.

Referring to Figure 9, which illustrates the qualitative results, connections (a), (b), (c), (d), and (e)—that is, all of the key connections identified in the qualitative analysis—are consistent with the T/ENA model in condition 3.

### A.2.3 Model Comparison

We proposed three hypotheses about the data and constructed three models to evaluate each hypothesis.

**Hypothesis 1: We would see differences in student activity between phases** Our first hypothesis was that there were differences in the way students addressed the design problem they were solving between the discovery and design phases of the simulation.

First, our qualitative analysis suggested that there were differences between the two phases in the patterns of connection

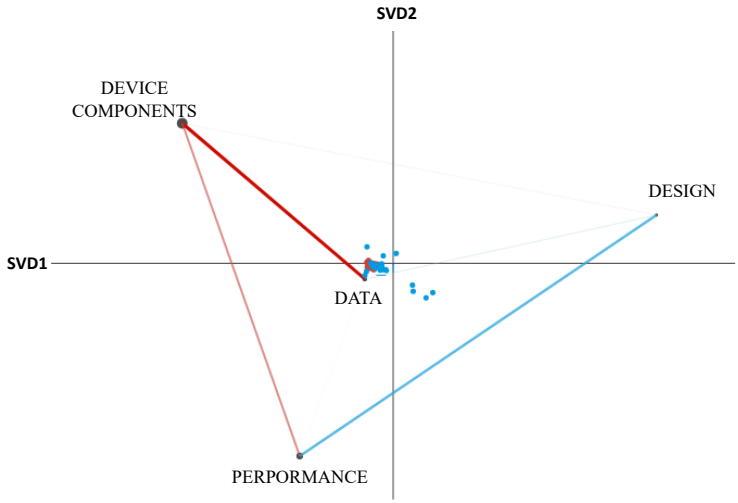


**Figure 11.** Multimodal ENA model of chat data and resource data in the *RescuShell* simulation. The subtracted network shows the differences between the means of the ENA networks for the discovery and design phases.

among the key design practices. These differences are summarized in Table 8.

Next, as shown in Figures 10, 11, and 12, in each of the analytic conditions, there were differences in the pattern of connections between the two phases of the simulation.

Finally, as shown in the result above (summarized in Table 9), in all three conditions, these differences were statistically significant.

We thus conclude that there were differences in student activity between the discovery and design phases of the simulation.

**Hypothesis 2: ENA on chat would outperform ENA on chat + resources** Our second hypothesis was that the ENA model on chats (C1) would outperform the multimodal ENA model on chats + resources (C2).

As shown in Table 9,

1. model C1 explains more variance ($R^2 = 0.18$) than model C2 ($R^2 = 0.13$), and
2. model C1 is more efficient ($AIC = 61.45$) than model C2 ($AIC = 66.58$).[9]

Table 8 shows that both models C1 and C2 differ on relative connection strength from the qualitative analysis 33% of the time (two of six connections).

We thus conclude that model C1 outperforms model C3 in terms of variance explained and efficiency, but neither model provides a more accurate representation of the qualitative analysis than the other.

**T/ENA would outperform both other conditions** Our third hypothesis was that T/ENA (C3) would outperform both of the other conditions.

As shown in Table 9, C3 explains more variance ($R^2 = 0.32$) and is more efficient ($AIC = 51.46$) than C1 and C2.

Moreover, Table 8 shows that while both models C1 and C2 differ on relative connection strength from the qualitative analysis on two of six connections, C3 shows the same relative strength as the qualitative analysis on all six connections.

We thus conclude that C3 outperforms the other models in the sense that it (a) accounts for more variance, (b) is more efficient, and (c) provides the best representation of the qualitative analysis.

### A.3 Summary

In the preceding sections, we presented a worked example of a widely used SML (ENA) applied to multimodal data from a learning setting.

The setting was an engineering simulation divided into two phases. In the discovery phase, students worked together to understand the design properties and key performance parameters of a mechanical exoskeleton design project. In the design phase, students used this information to create and test prototype exoskeleton designs. Throughout the simulation, students

[9] AIC measures model efficiency, with *lower* values being more efficient. The usual interpretation is that values that differ by more than four are significantly different.

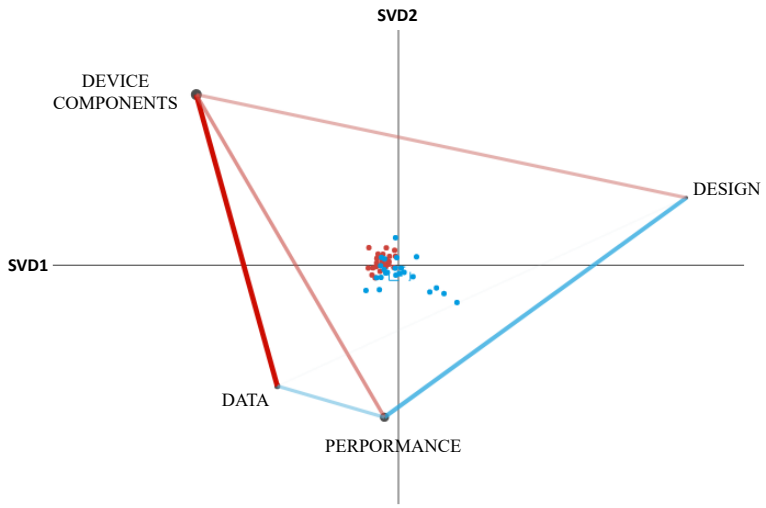


**Figure 12.** T/ENA model of chat data and resource data in the *RescuShell* simulation. The subtracted network shows the differences between the means of the ENA networks for the discovery and design phases.

**Table 8.** Connections that are identified as stronger in one phase than another by models in the three conditions relative to connections identified in the qualitative model.

| Connection | Qualitative | ENA Chats | ENA Chat+Resources | T/ENA Chat+Resources |
|---|---|---|---|---|
| DEVICE COMPONENTS & DATA | discovery | ✓ | ✓ | ✓ |
| DEVICE COMPONENTS & DESIGN | discovery | ✓ | ✗ | ✓ |
| DEVICE COMPONENTS & PERFORMANCE | discovery | ✓ | ✓ | ✓ |
| PERFORMANCE & DATA | design | ✗ | ✗ | ✓ |
| PERFORMANCE & DESIGN | design | ✓ | ✓ | ✓ |
| DATA & DESIGN | neither | ✗ | ✓ | ✓ |

were working in teams using an online collaboration portal. In the portal, students communicated with each other through chat messages and were able to access written resources, such as technical reports and schematics, that were related to the design task.

We constructed three models:

C1. ENA applied to the student chat messages,

C2. ENA applied to a low-level fusion of student chat messages and logfile data of resource usage, and

C3. TMA-enhanced ENA (T/ENA) applied to student chat messages and logfile data of resource usage.

When we compared the performance of these models, the results showed that model C1 outperformed model C2 and model C3 outperformed both of the other models.

In saying this, we want to be clear that we do not consider this worked example dispositive evidence that TMA outperforms ENA or SMLs more generally. We anticipate that (a) there are multimodal datasets and settings where an SML will outperform its corresponding T/SML *and* (b) it is possible to construct bespoke data fusion methods that would enable some SML models to outperform T/ENA for other specific datasets.

Our point in presenting this example is to show how TMA works with one specific SML and that in some cases a TMA-enhanced SML can outperform an SML that has not been adapted for multimodal data.

ISSN 1929-7750 (online). 

**Table 9.** A comparison of model performance shows that the T/ENA model (condition 3) outperforms both ENA on chats (condition 1) and multimodal ENA on chats and resources (condition 2). [** = $p < 0.01$, * = $p < 0.05$.]

| Condition | Predictors | | | Model Evaluation | |
|---|---|---|---|---|---|
| | Intercept | SVD 1 | SVD 2 | McFadden's $R^2$ | AIC |
| (1) ENA with only chats | 1.21** | 10.64** | 2.69 | 0.18 | 61.45 |
| (2) ENA with chats and resources | 1.89* | 14.39* | 8.59 | 0.13 | 66.58 |
| (3) TMA with chats and resources | 0.96 | 13.90* | 10.46** | **0.32** | **51.46** |

These results present something of a conundrum: an underlying hypothesis of MMLA is that the information from additional data modalities should provide a more complete picture of student learning. But in this case, modelling both chat messages and resources resulted in *poorer* model performance than modelling chat messages alone. On the other hand, the TMA model, which also used both chat messages and resources, resulted in *better* performance than either of the SML models.

These results show that multimodal models *can* account for more information about the learning environment than unimodal models (and, of course, this is not in dispute). However, the results here also suggest that multimodal models can do so *only if* the interactions between the different modalities are modelled appropriately.

Like any discourse setting, multimodal interactions have structure: ways in which events systematically influence (or do not influence, or influence more or less) future events. And one critical feature of multimodal data is that different modalities have different interactional properties. Indeed, that is one of the reasons we consider multiple modalities in the first place.

Because of the structure of the SML we considered here—and more generally by the nature of all extant SMLs—the SML-only model applied to two modalities treated two *different data modalities* as if they had the *same interactional properties*: the same impact on future events, the same impact on individual learners, and the same visibility to different participants in the setting. But this is clearly a set of assumptions that are not warranted except in very particular circumstances.

By accounting for the structural features of the multimodal discourse being modelled (or attempting to account for them), the TMA-enhanced model was able to use the information contained in multiple data modalities more effectively.

#### A.3.1 Limitations

The worked example we have provided here would be woefully inadequate if we were using it to definitively demonstrate the utility of TMA as a methodological approach. We looked at only one dataset, using only one learning analytic approach. We did little to systematically compare TMA to a range of data formatting solutions. But, as we hope has been clear, our example is intended to be illustrative rather than definitive. More work is clearly needed to systematically test the relative efficacy of TMA compared to other approaches to data fusion in MMLA—to identify which approaches work best and under what conditions.

## Appendix B: Additional References

This appendix provides additional background and references for those who wish to see a more comprehensive view of the literature on multimodal learning as it relates to TMA.

### B.1 Introduction

Learning is a multimodal process (K.-s. Tang et al., 2014; Jewitt et al., 2001). *State-dependent and state-space models of learning (SMLs)* include techniques such as Markov chains (Sharma et al., 2019; Kokoç et al., 2021; Gupta et al., 2022), process mining (Balogh & Kuchárik, 2019; Thiyagarajan & Prasanna, 2022; Huang et al., 2023), lag sequential models (K.-Z. Chen & Li, 2021; Lämsä et al., 2020; Jeng & Chung-Nien, 2022), Bayesian knowledge tracing (Wong et al., 2021; Lee et al., 2023; K. I. Chan et al., 2022), and ENA (Zhang et al., 2022; Teasley et al., 2023; Ba et al., 2023). SMLs model learning at some point in time as a function of the events that came before it (Priestley, 1980; Rahmani & Fay, 2022).

### B.2 Background

STEM learning investigates how students develop scientific, mathematical, and technical understanding by simultaneously using different modalities within and across multiple representations (Jewitt et al., 2001; Airey & Linder, 2009; Lemke, 1998). Analyses of learning that integrate multiple modes of data are potentially more accurate and more equitable—and therefore better able to influence both current practice and future research (Worsley, 2022; Alwahaby & Cukurova, 2022).

The field of MMLA uses multimodal data to model how students develop understanding, such as how data on talk, motion, location, gaze, emotion, and self-reports can be used to model how students learn from a lecture (Raca & Dillenbourg, 2013; Sümer et al., 2021; Alkabbany et al., 2023) or how data on talk, writing, drawing, movement, and galvanic skin response can be used to model how students learn from problem-solving exercises (Ochoa et al., 2013; Worsley & Blikstein, 2013; Larmuseau et al., 2020; H. Tang et al., 2022).

Ochoa and Worsley (2016) argue that different modes of data in a multimodal context are produced by different physical, psychological, and social processes and therefore influence learning through different mechanisms. For example, Hung and Higgins (2015) compared communications between learners using synchronous video conferencing and synchronous chat. Data collected from different modalities uses different *extraction processes*, which often have different characteristics (Ma et al., 2022; Cloude et al., 2022; Ochoa, 2022; Yusuf et al., 2023).

There are challenges in *processing* multimodal data, such as those identified by Di Mitri and colleagues (2018), Cukurova and colleagues (2020), and Ochoa (2022), including representing, segmenting, and integrating such different data streams in a way that preserves the important properties of each type of data (Worsley, 2014; Sharma & Giannakos, 2020; M. C. E. Chan et al., 2019; Conijn et al., 2020). Moreover, models that researchers develop and validate on such data may be biased toward majority groups and thus ultimately unfair to subgroups (Mehrabi et al., 2021; Chouldechova & Roth, 2018), with the potential to reify and even augment existing inequities (Gardner et al., 2019; Mayfield et al., 2019). But despite broad attention to issues of equity in education, model bias in MMLA has received little systematic attention (Yan et al., 2022).

Integration of complex, multimodal data can take place at different levels: including *naive fusion* (Schneider & Blikstein, 2015; L. Chen et al., 2016; Emerson et al., 2020; Worsley & Blikstein, 2011), *low-level fusion*, (Worsley & Blikstein, 2014; Dominguez et al., 2021; Ma et al., 2022), and *high-level fusion* (Alibali & Nathan, 2012; R. Liu et al., 2019; Chango et al., 2021; Sung et al., 2022).

Reimann and colleagues (2014) and Alonso-Fernández and colleagues (2019) provide useful overviews of analytic methods that model fused data of this kind, including Petri nets for process mining (Balogh & Kuchárik, 2019; Thiyagarajan & Prasanna, 2022; Huang et al., 2023), Markov chains (Sharma et al., 2019; Kokoç et al., 2021; Gupta et al., 2022), sequential analysis (Swiecki et al., 2019; K.-Z. Chen & Li, 2021; Akçapınar & Hasnine, 2022), ENA (Zhang et al., 2022; Teasley et al., 2023; Ba et al., 2023), group communication analysis (Dowell et al., 2018), Bayesian networks (Choi & Cho, 2020; Jiang et al., 2023; F. Liu et al., 2022), Bayesian knowledge tracing (Wong et al., 2021; Lee et al., 2023; K. I. Chan et al., 2022), autoregression (González-Brambila et al., 2021; Ahmad et al., 2023), vector autoregression (Buigut & Valev, 2005; G. Zhou et al., 2022; Y. Zhou & Kang, 2023), additive factors models (Rivers et al., 2016; F. Chen & Lu, 2022), time series analysis (Reilly & Dede, 2019; Poitras et al., 2020; Dorodchi et al., 2020), and structural equation models (Fincham et al., 2019; Al-Adwan et al., 2021; X. Liu, 2022).

However, as we argue in this paper, a multimodal model must account for the way that different events in different circumstances influence events in the future. Specifically, a multimodal SML needs to account for

1. *types of events*, which ground future events in different ways;
2. *characteristics of student populations*, allowing types of events to ground future events differently depending on characteristics of the learner involved; and
3. *horizons of observation*, allowing events to be included in the ground only for some students or groups of students.

In other words, discourse has *structure* (Gee, 2015; Shaffer, 2017): there are particular ways in which events unfold in a specific context, and the structure of discourse determines how one event influences another.

As Table 10 shows, with few exceptions, SMLs are not designed to address these features of multimodal data. Existing approaches that account for multiple modes of data (a) make assumptions that do not always hold in the context of LA and (b) do not account for all relevant features of multimodal heterogeneity in learning process data.

**Table 10.** Multimodal features of state-dependent models.

| | Multiple Data Sources | Multiple Event Types | Student Population & Event Type Interactions | Horizon of Observation |
|---|---|---|---|---|
| Additive Factors Model | ✗ | ✗ | ✗ | ✗ |
| Autoregression | ✗ | ✗ | ✗ | ✗ |
| Bayesian Knowledge Tracing | ✗ | ✗ | ✗ | ✗ |
| Group Conversation Analysis | ✗ | ✗ | ✗ | ✗ |
| Lag Sequential Models | ✗ | ✗ | ✗ | ✗ |
| Markov Chains | ✗ | ✗ | ✗ | ✗ |
| Time Series Analysis | ✗ | ✗ | ✗ | ✗ |
| Bayesian Networks | ✗ | ✗ | ✓ | ✗ |
| Petri Nets | ✗ | ✗ | ✓ | ✗ |
| ENA | ✗ | ✗ | ✓ | ✓ |
| Structural Models | ✓ | ✓ | ✗ | ✗ |
| Vector Autoregression | ✓ | ✓ | ✗ | ✗ |