Does Order Matter?
Investigating Sequential and Cotemporal Models of Collaboration

Zachari Swiecki, University of Wisconsin—Madison, swiecki@wisc.edu
Zheming Lian, University of Wisconsin—Madison, zlian4@wisc.edu
Andrew R. Ruis, University of Wisconsin—Madison, arruis@wisc.edu
David Williamson Shaffer, University of Wisconsin—Madison and Aalborg University Copenhagen, dws@education.wisc.edu

Abstract: Many researchers have argued that models of collaborative processes should account for temporality, but there exist different approaches for doing so. We compared two specific approaches to modeling collaborative processes in a CSCL context: Epistemic Network Analysis, which models events cotemporally (unordered and temporally proximate), and Sequential Pattern Mining, which models events sequentially (ordered and temporally proximate). Our results suggest that in this context cotemporal models constructed with Epistemic Network Analysis outperform sequential models constructed with Sequential Pattern Mining in terms of (a) explanatory power, (b) efficiency, and (c) interpretability.

Introduction

A central claim of computer-supported collaborative learning (CSCL) research is that collaborative processes influence group performance, and studies have shown that these processes have an important temporal dimension (Reimann, 2009). Researchers thus argue that models of collaborative processes should account for temporality (McGrath & Tschan, 2004), but there exist different approaches for doing so. *Sequential models*, which identify ordered patterns, can be used to investigate whether specific sequences of temporally proximate discourse moves, such as talk or actions, explain variation in group outcomes (Kapur, 2011). *Cotemporal models*, which identify unordered but temporally proximate patterns, can be used to investigate whether discourse moves that co-occur within some period of time explain variation in group outcomes (Siebert-Evenstone et al., 2017).

Prior work has shown that both cotemporal models (Csanadi, Eagan, Shaffer, Kollar, & Fischer, 2019) and sequential models (Kapur, 2011) have advantages over atemporal models such as coding and counting; however, no empirical comparisons have been made between cotemporal models and sequential models in collaborative contexts. Thus, it is unclear whether accounting for temporal proximity, versus both temporal proximity and sequence, is a more effective approach for modeling collaborative processes. Such a comparison will help inform researchers about the conditions under which one class of models outperforms another, potentially impacting assessments of complex thinking and performance. But more importantly, comparing techniques that model temporality in terms of local sequence versus local cotemporality can provide insight in to the nature of temporality itself in CSCL contexts.

In this paper, we present an initial attempt to address this issue by comparing two specific modeling approaches: Epistemic Network Analysis (Shaffer, Collier, & Ruis, 2016), which models events cotemporally, and Sequential Pattern Mining (Srikant & Agrawal, 1996), which models events sequentially. We use both models to analyze data collected from air defense warfare (ADW) teams participating in computer-simulated training scenarios. We compare the efficacy of these models at finding differences in group performance in terms of (a) explanatory power, (b) efficiency, and (c) interpretability.

Theory

Researchers in CSCL argue that collaboration has an important temporal dimension. For example, Kozlowski & Illgen (2006) argue that repeated interactions between individuals create behavioral, cognitive, or motivational states that influence future interactions. Similarly, Clark (1996) argues that as collaborative activities unfold in time, information is added to the *common ground*, the set of shared knowledge and experiences that exist between people when they interact, which in turn influences subsequent actions and interpretations (Dillenbourg, 1999).

Models of collaboration that do not account for temporality thus omit crucial information, limiting their validity (Kapur, 2011). One prevalent response to this critique has been to focus on *sequences* of discourse moves. Sequence is potentially important because, as Reimann (2009) argues, "human learning is inherently cumulative, [and] the sequence in which experiences are encountered affects how one learns and what one learns." Moreover, in some collaborative settings, discourse moves should be carried out in a particular order. For example, Hutchins' (1995) study of quartermasters in the U.S. Navy showed that navigation teams needed to follow a particular sequence of actions in order to accurately track the position of their ship. *Sequential models*, such as Sequential

Pattern Mining (SPM), have been used to identify sequences of discourse moves that groups make during collaborative processes in a variety of contexts (Perera et al., 2009).

There is, however, another aspect of the temporality of collaborative processes: *temporal proximity*. Events at any point in time are influenced by prior actions. However, the influence of prior activity does not always span the entire history of group interaction. Halpin and von Davier (2017) argue that the actions of one part of a group make actions by other parts more or less likely in the near future. For example, when one group member asks a question, others are likely to respond soon after. Suthers & Desiato (2012) argue that actions and interactions are interpreted with respect to the *recent temporal context*, or the immediately preceding events. Discourse moves within the recent temporal context may influence one another, directly reference one another, or build upon one another. Thus, while collaborative processes are composed of complex interactions among individuals, the most *relevant* interactions are bounded by temporal proximity.

Many SPM algorithms allow researchers to account for recent temporal context using sliding windows. Such algorithms add the constraint that the identified sequences must occur within a given window of events. However, an alternative approach is to model temporality based on co-occurrence rather than sequence, which focuses on temporal proximity irrespective of order.

In *cotemporal* models, two discourse moves are meaningfully connected if they co-occur within the same recent temporal context (Shaffer, 2017). For example, Epistemic Network Analysis (ENA) can be used to identify the connections groups make during collaboration (Shaffer et al., 2016). ENA identifies these connections by measuring how often particular discourse moves co-occur within the recent temporal context, operationalized as a sliding window that moves through each event—for example, problem step or turn of talk—in the data. ENA represents connections between discourse moves using undirected network models, meaning that connections between moves A and B in the network could mean that A followed B or that B followed A. In this way, ENA is *sensitive* to the order of events in the data—changing the order of events changes which events are present in a given window, and thus changes the results of the model—but the order in which discourse moves occur within any window is not represented in the model. ENA has been used to study CSCL and collaborative problem solving processes in many domains (e.g., Arastoopour, Shaffer, Swiecki, Ruis, & Chesler, 2016; Sullivan et al., 2018).

While it may seem counterintuitive to ignore the local order of discourse moves—after all, we perceive human actions as unfolding linearly in time—there are clearly contexts in which the specific order of these moves is less important than their cotemporality. For example, in complex and ill-formed problem solving, groups might consider issues A, B, and C at one point in the problem solving process; however, it may make little difference whether in that brief span of time they talk about A then B then C, or C then B then A—or any of the possible orderings of those issues. This approach has potential advantages over sequence models. For example, the results of SPM are lists of ordered patterns. A typical analysis identifies frequent patterns using SPM, then clusters those patterns using similarity metrics. Finally, researchers interpret these clusters in terms of the patterns within them (Jovanović, Gašević, Dawson, Pardo, & Mirriahi, 2017). In contrast, ENA produces network models for each unit of analysis, and provides an integrated visualization that helps to interpret the dimensions along which groups of networks differ. Because visualizations can reduce cognitive load in making inferences (Norman, 1993), cotemporal models with integrated visualizations may have interpretive advantages. Moreover, the number of possible permutations of discourse moves rises far more rapidly than their possible pairwise combinations as the number of significant types of moves increases. So in a case where specific ordering is not relevant, sequential models, which distribute sample variance across many variables, may have less explanatory power and may be overfit unless the analyzed dataset is large.

In this paper, we present an attempt to compare cotemporal and sequential approaches to modeling collaboration: ENA, which models events cotemporally, and SPM, which models events sequentially. We use both models to analyze data collected from air defense warfare teams (ADW) as they participated in computer-simulated training scenarios. In these scenarios, ADW teams detect, identify, assess, and take action toward nearby radar contacts. In theory, this process should follow a specific sequence for each contact; however, the demands of the task may lead to deviations from the sequence. Thus, the collaborative problem is neither so well-formed that only specific sequences are of interest nor so ill-formed that sequences are likely to be of little interest. We use this data to compare two hierarchical linear models—one using predictors derived from ENA and one using predictors derived from SPM—to assess whether ENA or SPM provides a more effective model of group performance in terms of (a) explanatory power, (b) efficiency, and (c) interpretability.

Methods

As part of the Tactical Decision Making Under Stress project, 16 teams composed of six members each participated in four simulated training scenarios to test the impact of a decision-support system on team performance (Johnston, Poirier, & Smith-Jentsch, 1998). During the scenarios, teams performed the *detect-to-*

engage sequence. A *watch-station* provided basic information about identification and behavior of ships and aircraft in the vicinity (referred to as *tracks*). Teams needed to detect and identify multiple tracks, often simultaneously, assess whether they were threats, and decide how to respond, although the full sequence of actions did not apply to every track. (For example, non-hostile tracks did not require a response.) Teams in the control condition ($n = 8$) had access to standard watch-stations. Teams in the experimental condition ($n = 8$) had access to watch-stations enhanced with information about the tactical situation. Each team participated in the same four 30-minute scenarios; scenario order was counterbalanced using a Latin square.

We analyzed two data sources for each *team-scenario* (that is, each team in each scenario): (1) a transcript of team communications and (2) a teamwork behavior score. Transcripts were segmented into 12,027 turns of talk. Teamwork behavior was assessed in a prior study using the Air Defense Warfare Team Observation Measure (ATOM) (Johnston, Smith-Jentsch, & Cannon-Bowers, 1997), which summarizes four dimensions of teamwork performance—supporting behavior, leadership, information exchange, and communication—into an overall score from 1 (worst) to 55 (best).

Coding

We analyzed the transcripts using the codes in Table 1, which were developed using a grounded analysis informed by both the existing ADW literature (e.g., Paris et al., 2000) and prior qualitative analyses conducted on similar data (e.g., Morrison et al., 1996). To code the data, we developed automated classifiers for each of the codes in Table 1 using the *ncodeR* package for the statistical programming language R (Marquart, Swiecki, Eagan, & Shaffer, 2018). We assessed concept validity by requiring that two human raters achieve acceptable values of Cohen's kappa ($\kappa > 0.65$) with statistically significant values of Shaffer's rho ($\rho < 0.05$), and we assessed reliability by requiring that both human raters independently achieve acceptable values of kappa and rho compared to the automated classifier (1). For each code, all pairwise combinations of raters (humans and automated classifier) achieved $\kappa > 0.80$ and $\rho(0.65) < 0.05$.

Table 1: Qualitative codes, definitions, and examples

Code	Definition	Example
Detection	Talk about radar detection of a track or the identification of a track, (e.g., vessel type).	IR/EW NEW BEARING, BEARING 078 APQ120 CORRELATES TRACK 7036 POSSIBLE F-4
Track Behavior	Talk about kinematic data about a track or a track's location	AIR/IDS TRACK NUMBER 7021 DROP IN ALTITUDE TO 18 THOUSAND FEET
Assessment	Talk about whether a track is friendly or hostile, the threat level of a track, or indicating tracks of interest	TRACKS OF INTEREST 7013 LEVEL 5 7037 LEVEL 5 7007 LEVEL 4 TRACK 7020 LEVEL 5 AND 7036 LEVEL 5.
Status Updates	Talk about procedural information, e.g., track responses, or talk about tactical actions taken by the team	TAO ID, STILL NO RESPONSE FROM TRACK 37, POSSIBLE PUMA HELO.
Seeking Information	Asking questions regarding track behavior, identification, or status.	TAO CO, WE'VE UPGRADED THEM TO LEVEL 7 RIGHT?
Recommendation	Recommending or requesting tactical actions	AIR/TIC RECOMMEND LEVEL THREE ON TRACK 7016 7022
Deterrent Orders	Giving orders meant to warn or deter tracks.	TIC AIR, CONDUCT LEVEL 2 WARNING ON 7037
Defensive Orders	Giving orders to prepare defenses or engage hostile tracks	TAO/CO COVER 7016 WITH BIRDS

Epistemic Network Analysis

To conduct a cotemporal analysis, we used the *rENA* package for the statistical programming language R (Marquart, Swiecki, Collier, et al., 2018). ENA uses a sliding window to construct a network model for each turn of talk in the data. Connections in the network are defined as the co-occurrence between codes in the current turn of talk and codes within the recent temporal context, which we defined as each line plus the four previous lines

based on our qualitative analysis of the data (a window size of 5 turns of talk). The resulting networks are aggregated for all turns of talk for each unit of analysis (team-scenario), such that each team-scenario is represented by a vector whose elements are the number of co-occurrences between each pair of codes for that team-scenario. ENA normalizes the matrix of co-occurrence vectors to account for variation in the amount of talk between teams and performs a dimensional reduction on the matrix via singular value decomposition.

Networks were visualized using two coordinated representations: (1) an *ENA score,* which represents the location of a team-scenario's network in the space (or *ENA space*) created by the dimensional reduction, and (2) a weighted network graph in which the nodes correspond to codes, and the edges are proportional to the relative frequency of connection between two codes. The positions of the network graph nodes are fixed across networks, and those positions are determined by an optimization algorithm that minimizes the difference between the ENA scores and their corresponding network centroids. Thus, ENA scores toward the extremes of a dimension have network graphs with strong connections between nodes located on the extremes. As a result, dimensions in this ENA space distinguish team-scenarios in terms of cotemporality between codes whose nodes are located at the extremes.

Sequential Pattern Mining

To conduct a sequential analysis, we used the *TraMineR* package for the statistical programming language R (Gabadinho, Ritschard, Müller, & Studer, 2011). SPM identifies frequent patterns using a *support threshold*, where the support for a given pattern is the percentage of sequences that contain at least one instance of the pattern. We defined a sequence as the ordered list of codes across all turns of talk for a given team-scenario. In our data, codes can co-occur within a single turn of talk, so the SPM algorithm treats these codes as occurring simultaneously by defining *event sequences* which may contain patterns of un-ordered as well as ordered events (Ritschard, Bürgin, & Studer, 2013). We used a support threshold of 0.75 to limit the total number of patterns returned by the algorithm, which eases interpretation, and a window size of five turns of talk to match our ENA model (2). To interpret the SPM results, we counted the frequency of each high-support pattern for each team-scenario and applied Principal Components Analysis (PCA) to this data, resulting in a *PCA score* for each team-scenario—that is, the position of each team-scenario on the PCA dimensions (3). To interpret the PCA results, we used the dimension loadings, which show how much each variable—in this case, each frequent pattern—contributes to each dimension.

Model comparison

We compared cotemporal and sequential models by constructing two Hierarchical Linear Models (HLMs). HLM is a regression technique for data with a nested-structure: in this case, team-scenarios (level-one) were nested into teams (level-two). In both HLMs, teams were random effects, the team-scenario ATOM score was the outcome variable at level-one, and Scenario was a control variable at level-two. For the cotemporal HLM (CT-HLM), ENA scores were explanatory variables at level-one. For the sequential HLM (S-HLM), PCA scores were explanatory variables at level-one.

We assessed model fit using two estimates of the total variance explained for HLMs: one from Snijders and Bosker (2012, TVE1) and the other from LaHuis and colleagues (2014, TVE2) (4). We assessed the efficiency of the models (model fit adjusted for number of parameters) using the Akaike information criterion corrected for small sample sizes. Following Burnham and Anderson (2004), we used a minimum AIC_c difference of 4 to indicate that the models were distinguishable.

Results

Epistemic Network Analysis

The first six ENA dimensions accounted for more variance in the data than any original variable. To reduce the chance of overfitting, we only used ENA scores on the first two dimensions in the CT-HLM. These two dimensions accounted for the highest proportion of the total variance: 51%. Figure 1 shows the average network across all team-scenarios and the ENA scores for each team-scenario in this space.

On the left side of the space are connections to Defensive Orders, Deterrent Orders, Status Updates, and Recommendations, all of which relate to actions taken by teams toward tracks. On the right side of the space are connections to Seeking Information. This suggests that the first dimension distinguishes team-scenarios in terms of whether they focused on *Tactical Actions* versus *Seeking Information*. Toward the top of the space are connections to Detection and Track Behavior; toward the bottom are connections to the remaining codes. Detection and Track Behavior relate to passing information about tracks, while the remaining codes relate to using

information about tracks. This suggests that the second dimension distinguishes team-scenarios in terms of whether they focused on *Track Information* versus *Track Processing*.
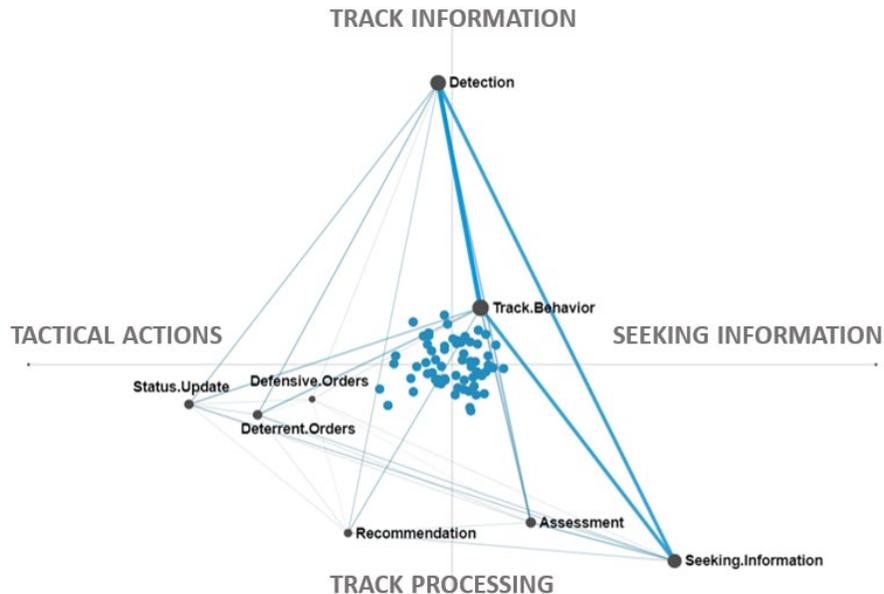

Figure 1. Average ENA network across all team-scenarios and ENA scores for each team-scenario.

Sequential Pattern Mining

Our SPM analysis returned 165 patterns. PCA on counts of these patterns returned 23 dimensions that accounted for more variance than any original variable. To maintain consistency with the ENA analysis, we only used PCA scores on the first three dimensions, which accounted for 53% of the variance. As there are 165 original variables, interpretation requires considering 165 loadings for each dimension. However, it is common to use only the loadings with high magnitudes for interpretation. We interpreted each dimension by considering commonalities between the ten patterns that loaded at either extreme. To conserve space, we show only the three patterns at each extreme for each dimension in Table 2.

Table 2: Patterns with extreme loadings on first three principal components

PC1	PC2	PC3
Patterns with Extreme Negative Loadings		
(Track Behavior)-(Detection)-(Track Behavior)	(Assessment)-(Seeking Information)	(Track Behavior, Assessment)-(Seeking Information)
(Track Behavior)-(Detection, Track Behavior)	(Seeking Information)	(Deterrent Orders)
(Track Behavior)-(Track Behavior)-(Detection)	(Seeking Information)-(Seeking Information)-(Track Behavior)	(Status Update)-(Assessment)
Patterns with Extreme Positive Loadings		
(Defensive Orders)	(Detection, Track Behavior)-(Deterrent Orders)	(Seeking Information)-(Detection)-(Detection, Track Behavior)
(Status Update)-(Recommendation)	(Detection)-(Deterrent Orders)	(Seeking Information)-(Detection)-(Detection)
(Status Update)-(Deterrent Orders)	(Status Update)-(Detection)	(Detection)-(Detection)

The extreme negative side of the first dimension includes patterns involving Track Behavior and Detection. The extreme positive side includes patterns involving Defensive Orders, Deterrent Orders, Status Updates, and Recommendations. This suggests that the first dimension distinguishes team-scenarios in terms of whether they focused on *Track Information* versus *Tactical Actions*. The extreme negative side of the second dimension includes many patterns involving Seeking Information, while the extreme positive side includes patterns involving Status Updates or ending with Deterrent Orders. This suggests that the second dimension distinguishes team-scenarios in terms of whether they focused on *Seeking Information* versus *Deterring Tracks.* Finally, the extreme negative side of the third dimension includes many patterns involving Assessment. The extreme positive side includes patterns involving Seeking Information, Detection, and Track Behavior. This suggests that the third dimension distinguishes team-scenarios in terms of whether they focused on *Assessing Tracks* versus *Exchanging Information*.

Model comparison

We compared the CT-HLM (ENA scores as explanatory variables) to the S-HLM (PCA scores as explanatory variables). Coefficients, standard errors (with corresponding *p*-values), as well as measures of model fit and efficiency for both models are shown in Table 3.

Table 3: Cotemporal HLM and sequential HLM, including parameter estimates and model fit

Model	Intercept	ENA		PCA			Scenarios			Model Fit		
		1	2	1	2	3	B	C	D	TVE1	TVE2	AICc
CT-HLM	**36.55*** **(1.06)**	**–10.1*** **(2.77)**	–0.42 (3.49)				2.78 (1.58)	–2.34 (1.57)	2.84 (1.49)	0.259	0.253	**377.19**†
S-HLM	**36.32*** **(1.11)**			0.12 (0.09)	**0.38*** **(0.18)**	–0.18 (0.14)	2.67 (1.61)	–1.61 (1.66)	3.1 (1.55)	0.238	0.237	381.72

( ) indicates standard error; * indicates $p < 0.05$; † indicates significantly lower AICc score ($\Delta > 4$)

For the CT-HLM (cotemporal), the variance estimates for the random effects were 0.67 at level two and 16.70 at level one. Of the explanatory variables, only the coefficient for scores on the first ENA dimension (ENA1) was significant. The negative coefficient (–10.1) indicates that teams with higher teamwork behavior ratings focused more on *tactical actions* toward tracks than *seeking information about tracks*.

For the S-HLM (sequential), the variance estimates for the random effects were 0.06 at level two and 17.79 at level one. Of the explanatory variables, only the coefficient for scores on the second PCA dimension (PC2) was significant. The positive coefficient (0.38) indicates that teams with higher teamwork behavior ratings focused more on *deterring tracks* than *seeking information about tracks*.

For both estimates of total variance explained, the CT-HLM performed better. Moreover, the CT-HLM had an AICc score 4.53 points lower than the S-HLM. Thus, the two models are distinguishable: the cotemporal model performed better than the sequence model in explaining differences in team behavior scores efficiently.

Discussion

Our results suggest that cotemporal models of collaborative processes constructed with ENA outperform sequential models constructed with SPM on several dimensions. The cotemporal models explained more of the difference between team-scenarios, and while differences in total variance explained were small, model selection for efficiency via AICc showed that the cotemporal-HLM was distinguishable from, and preferable to, the sequential-HLM. A possible explanation is that our ENA model used 28 un-ordered pairs of codes to derive predictors for the regression analysis, while our SPM model used 165 frequent patterns. The dimensional reduction on the ENA variables needed only *two* dimensions to account for a large proportion of the total variance while SPM needed *three* dimensions to account for approximately the same proportion of the total variance. Thus, the HLM model with ENA predictors was more parsimonious and less likely to overfit the data.

Our results also suggest that cotemporal models can have interpretive advantages compared to sequence models. The ENA algorithm combined connection identification, dimensional reduction, and visualization into one technique; once we generated the model it was ready to interpret. SPM, on the other hand, required several steps of post processing. More importantly, however, interpretation using ENA is done via integrated visualizations—network graphs projected into a low-dimensional space. SPM does not include such

visualizations. Because visualizations can reduce cognitive load by replacing cognitive calculations with perceptual inferences, ENA has an interpretive advantage compared to SPM.

Together, these results suggest that in CSCL and collaborative problem solving more generally, the specific *local order* of discourse moves may be less important than their *local cotemporality*. In the context we examined, teams were expected to follow a specific sequence of steps regarding each track; however, teams had to manage multiple tracks at once and the full sequence of actions did not apply to every track: it mattered more that specific discourse moves occurred together than that they occurred in a specific order.

Our results have several important limitations. First, we examined only one particular context of collaboration. However, the general features of this collaborative task, which contained both well-formed and ill-formed components, suggests that our findings may generalize to similar conditions in CSCL. Our future work will continue to investigate cotemporal and sequence models in similar contexts, as well as contexts that are more well-formed and more ill-formed. Second, it is possible that dimensional reduction on the SPM results via other techniques, such as clustering or factor analysis, could yield sequence models that perform better and are easier to interpret. Our results suggest that a two-cluster solution would be required to compete with ENA in terms of model efficiency, but each cluster would then need to be interpreted using more than 80 patterns. Allowing more clusters could ease interpretation but sacrifice efficiency. Similarly, we could apply a factor analysis to the SPM results using a two-factor solution, but it is unclear whether the performance and interpretability of the resulting sequence model would improve. Thus, we hypothesize results similar to those reported here. Our future work will compare cotemporal models developed using ENA to sequential models developed using clustering or factor analysis on SPM results to test this hypothesis.

Despite these limitations, our comparisons suggest that in CSCL contexts, cotemporal models can outperform sequential models. More importantly however, our results suggest that in some CSCL contexts, local order appears to be less important than local cotemporality. These results have implications for research and assessment in CSCL—in contexts that share both well-formed and ill-formed features, cotemporal models are a potentially more effective means of assessing complex thinking and performance. In turn, these models may better inform pedagogy and learning in such contexts.

Endnotes

(1) Shaffer's rho is a Monte Carlo rejective statistic that quantifies Type I error for generalizing from a sample of data coded by two raters to the true rate of agreement.
(2) There are no established guidelines for choosing a support threshold. We tested multiple support thresholds above and below 0.75, and the quantitative results of the HLM comparisons were similar in all cases.
(3) Performing a cluster analysis on the high-support patterns is more commonly used to interpret the results of SPM; however, methods are not reliably available for event sequences. One implementation exists (Ritschard et al., 2018), but it has not been validated. PCA has a standard method of interpretation and is agnostic to the type of pattern.
(4) Because these models are non-nested, we were unable to test for significant differences in model fit.

References

Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, *32*(3B), 1492–1501.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Csanadi, A., Eagan, B., Shaffer, D. W., Kollar, I., & Fischer, F. (2019). 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*.

Dillenbourg, P. (1999). *Collaborative learning: Cognitive and computational approaches. advances in learning and instruction series.* ERIC.

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, *40*(4), 1–37.

Halpin, P. F., & von Davier, A. A. (2017). Modeling Collaboration Using Point Process. In *Innovative Assessment of Collaboration* (pp. 233–247). Springer.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Johnston, J. H., Poirier, J., & Smith-Jentsch, K. A. (1998). Decision making under stress: Creating a research methodology. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 39–59). Washington, D.C.: American Psychological Association.

Johnston, J. H., Smith-Jentsch, K. A., & Cannon-Bowers, J. A. (1997). Performance measurement tools for enhancing team decision making. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, method, and application* (pp. 331–327). Hillsdale, NJ: Erlbaum.

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning Analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*, 74–85.

Kapur, M. (2011). Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International Journal of Computer-Supported Collaborative Learning*, *6*(1), 39–56.

Kozlowski, S. W., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, *7*(3), 77–124.

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, *17*, 433–451.

Marquart, C. L., Swiecki, Z., Collier, W., Eagan, B., Woodward, R., & Shaffer, D. W. (2018). rENA: Epistemic Network Analysis (Version 0.1.3).

Marquart, C. L., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). ncodeR (Version 0.1.2).

McGrath, J. E., & Tschan, F. (2004). *Temporal matters in social psychology: Examining the role of time in the lives of groups and individuals*. Washington: American Psychological Association.

Morrison, J. G., Kelly, R. T., Moore, R. A., & Hutchins, S. G. (1996). Tactical decision making under stress (TADMUS) decision support system, 13.

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. New York: Basic Books.

Paris, C., Johnston, J. H., & Reeves, D. (2000). A Schema-Based Approach to Measuring Team Decision Making in a Navy Combat Information Center. In *The Human in Command* (pp. 263–278). Springer, Boston, MA.

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 759–772.

Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 239–257.

Ritschard, G., Bürgin, R., & Studer, M. (2013). Exploratory mining of life event histories. In J. J. McArdle & G. Ritschard (Eds.), *Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences* (pp. 221–253). New York: Routledge.

Ritschard, G., Studer, M., Buergin, R., Gabadinho, A., Fonta, P.-A., Muller, N., & Rousset, P. (2018). TraMineRextras: TraMineR extension (Version 0.43).

Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9–45.

Siebert-Evenstone, A., Arastoopour Irgens, G., Collier, W., Swiecki, Z., Ruis, A. R., & Williamson Shaffer, D. (2017). In Search of Conversational Grain Size: Modelling Semantic Structure Using Moving Stanza Windows. *Journal of Learning Analytics*, *4*(3), 123–139.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. *Advances in Database Technology—EDBT*, 1–17.

Sullivan, S. A., Warner-Hillard, C., Eagan, B. R., Thompson, R., Ruis, A. R., Haines, K., … Jung, H. S. (2017). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery*, in press.

Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. In *System Science (HICSS), 2012 45th Hawaii International Conference on* (pp. 3368–3377). IEEE.

Acknowledgments