Springer

Dear Author

Here are the proofs of your article.

- You can submit your corrections **online,** via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For **fax** submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
- Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL:

`http://dx.doi.org/10.1007/s11412-018-9288-8`

If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information, go to: http://www.link.springer.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us, if you would like to have these documents returned.

AUTHOR'S PROOF

Metadata of the article that will be visualized in OnlineFirst

Please note: Images will appear in color online but will be printed in black and white.

Article Title	**When coding-and-counting is not enough: using epistemic network analysis (ENA) to analyze verbal data in CSCL research**	
Article Sub-Title		
Article Copyright Year	**International Society of the Learning Sciences, Inc. 2018 (This will be the copyright line in the final PDF)**	
Journal Name	**International Journal of Computer-Supported Collaborative Learning**	

Corresponding Author	Family Name	**Csanadi**
	Particle	
	Given Name	**Andras**
	Suffix	
	Organization	Bundeswehr University of Munich
	Address	Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
	e-mail	andras.csanadi@unibw.de
Author	Family Name	**Eagan**
	Particle	
	Given Name	**Brendan**
	Suffix	
	Organization	University of Wisconsin-Madison
	Address	Madison, WI, USA
	e-mail	beagan@wisc.edu
Author	Family Name	**Kollar**
	Particle	
	Given Name	**Ingo**
	Suffix	
	Organization	University of Augsburg
	Address	Augsburg, Germany
	e-mail	ingo.kollar@phil.uni-augsburg.de
Author	Family Name	**Shaffer**
	Particle	
	Given Name	**David Williamson**
	Suffix	
	Organization	University of Wisconsin-Madison
	Address	Madison, WI, USA
	e-mail	dws@education.wisc.edu

Author	Family Name	**Fischer**
	Particle	
	Given Name	**Frank**
	Suffix	
	Organization	Ludwig Maximilian University of Munich
	Address	Munich, Germany
	e-mail	frank.fischer@psy.lmu.de

Schedule	Received	13 November 2017
	Revised	
	Accepted	5 November 2018

Abstract

Research on *computer-supported collaborative learning* (CSCL) is often concerned with the question of how scaffolds or other characteristics of learning may affect learners' social and cognitive engagement. Such engagement in socio-cognitive activities frequently materializes in discourse. In quantitative analyses of discourse, utterances are typically coded, and differences in the frequency of codes are compared between conditions. However, such traditional *coding-and-counting-based strategies* neglect the *temporal* nature of verbal data, and therefore provide limited and potentially misleading information about CSCL activities. Instead, we argue that analyses of the temporal proximity, specifically temporal co-occurrences of codes, provide a more appropriate way to characterize socio-cognitive activities of learning in CSCL settings. We investigate this claim by comparing and contrasting a traditional coding-and-counting analysis with *epistemic network analysis* (ENA), a discourse analysis technique that models temporal co-occurrences of codes in discourse. We apply both methods to data from a study that compared the effects of individual vs. collaborative problem solving. The results suggest that compared to a traditional coding-and-counting approach, ENA provides more insight into the socio-cognitive learning activities of students.

Keywords (separated by '-'): Discourse analysis - Coding-and-counting - Epistemic network analysis - Problem solving

Foot note information

Introduction

A major goal of research in *computer supported collaborative learning* (CSCL) is to understand how to use technology to improve collaborative learning. For example, Bause et al. (2018) investigated whether a particular design of a multitouch table that separates a private from a joint screen area is more effective for groups working on a problem-solving task than a design that does not include a joint working space. Likewise, many empirical studies look at whether CSCL scripts evoke different socio-cognitive actions than unscripted CSCL (e.g., Schwaighofer et al. 2017).

Central to such studies is the analysis of how differently designed learning environments impact how students interact during learning. For that purpose, researchers often rely on verbal data that are captured during learning, such as transcripts of within-group talk. These data are then analyzed to model how different learning conditions impact learners' actions, such as developing explanations or evaluating evidence (Teasley 1995).

Such analyses are typically based on *coding-and-counting* (e.g., Vogel and Weinberger 2018). In this approach, a researcher (1) develops a coding scheme to identify different actions that occurred during learning; (2) applies that coding scheme to the data corpus; and (3) typically counts the frequencies by which learners in different experimental conditions engaged in these actions. Frequency-based methods of this coding-and-counting-strategy thus provide a means for comparing the effects that different conditions have on the learners' socio-cognitive actions.

Despite its wide adoption in the CSCL community, however, coding-and-counting-based analyses as the one just described have been repeatedly criticized in CSCL research (Kapur 2011; Reimann 2009). In particular, critics of such an approach argue that (1) it ignores *temporality* in verbal data, and (2) it does not afford analyzing *patterns* of learning activities. That is, such traditional coding-and-counting-based approaches model the frequency of each kind of learner action (each code), but do not provide information about whether and how these actions might be related to one another.

For example, during collaboration, learners often develop questions and expectations that guide their interaction with each other and with the learning material. Counting how often each learner formulates questions and also counting independently how often each learner refers to the learning material tells us nothing about whether the learners have made *connections* between their questions and the learning material over time. We thus argue that using traditional coding-and-counting-based techniques as described above is often a suboptimal strategy to model learning in verbal data. In many cases, a more appropriate and informative approach is to use methods that model *temporal relationships* between coded socio-cognitive actions in verbal data.

In this article, we compare a traditional coding-and-counting-based analysis of a data corpus to epistemic network analysis (ENA; Shaffer et al. 2009; Shaffer 2017), an analysis method that models temporality in verbal data. We apply both a typical coding-and-counting approach and an ENA analysis on the same data set, and then examine the inferences that can be drawn from the two analyses.

To further investigate the impact of failing to account for temporality in the analysis of verbal data from a CSCL environment, we also compare the results of ENA on the original data set with the results of ENA on a *randomized version* of the original dataset. Randomizing the order of coded learning actions within each transcript preserves the *frequency* of occurrence of learner' actions in a verbal protocol, but eliminates temporal information from the original transcripts. Therefore, comparing the original data set to a randomized data provides an opportunity to understand more deeply the impact of temporality on the learning activities being modelled.

Engaging in socio-cognitive activities during CSCL: An example

The data we use to address these questions comes from an experiment in which pre-service teachers were asked to reason about a pedagogical problem (Csanadi et al. 2016). In one condition, students were asked to discuss the problem in pairs; in the other condition, students reflected on the problem individually using a *think aloud protocol* (e.g., Ericsson and Simon 1980; Fox et al. 2011). Using transcripts of discourse, Csanadi et al. (2016) investigated whether and how participants' engagement in actions of scientific reasoning such as hypothesizing and evaluating evidence, differed between the two conditions.

Tables 1 and 2 show two excerpts from this study. In what follows, we will refer to these two examples to describe how both traditional coding-and-counting approaches and ENA model this data.

The transcripts from dyadic discussions and individual think-aloud protocols were segmented into propositional units, and each proposition was coded (Csanadi et al. 2016) using a coding scheme developed by Csanadi et al. (2015) based on a heuristic framework of scientific reasoning (Fischer et al. 2014). The coding scheme identifies one of eight kinds of epistemic actions for each propositional unit:

(1) *Problem Identification* (PI): an initial attempt to build an understanding of the problem
(2) *Questioning* (Q): statements or questions triggering further inquiry
(3) *Hypothesis Generation* (HG): developing explanations of the problem
(4) *Generating Solutions* (GS): developing interventions or solution plans
(5) *Evidence Generation* (EG): reference to information or lack of information that could support a claim
(6) *Evidence Evaluation* (EE): evaluating a claim
(7) *Communicating and Scrutinizing* (CS): planned discussions with others (e.g., in order to find out further information)
(8) *Drawing Conclusions* (DC): concluding outcomes of reasoning

More specific details of segmentation and coding are discussed in the methods section below.

Measuring socio-cognitive activities by a traditional coding-and-counting approach

Both traditional coding-and-counting-based approaches and an ENA analysis begin with a *coding phase*. In the coding phase, researchers identify socio-cognitive actions that are relevant to the research question at hand. Then, they develop a coding framework to capture those actions in the data, and apply the framework to the data. The whole procedure may, in fact, include several steps and iterations of those steps (see e.g., Chi 1997; Strijbos et al. 2006; Vogel and Weinberger 2018; Shaffer 2017). The coding scheme we used in this experiment is described briefly above, and in more detail in the methods section.

While both traditional coding-and-counting-based analyses and ENA models use coded data, they differ with respect to what subsequently is done with the coded data. In typical coding-and-counting-based studies, the coding phase is followed by a *counting phase*, in which the researcher chooses units of analysis and computes the *code frequency*—the rate at which a code appears in the data—for each code within the data from each unit of analysis. Differences between code frequencies across units of analysis in different conditions are then analyzed statistically.

In our case, we used a multivariate ANOVA to test whether the two conditions (individual versus dyadic problem solving) had differential effects on learners' problem solving. The results, which we will present in more detail later, were useful to understand *the extent to which* participants engaged in actions of scientific reasoning, and whether the extent to which participants engaged in those actions was different between the two conditions, We could also have used more complex inferential methods based on code frequencies: for example, mediation analysis, which might test whether engagement in certain activities mediates the effect of the experimental conditions on learning outcomes, but that was not the goal of this study.

Although traditional coding-and-counting-based methods are often used in CSCL research (and we ourselves have used them extensively in the past; e.g., Csanadi et al. 2016; Kollar, Fischer, & Slotta, 2007; Stegmann et al. 2012), they have been heavily criticized (Kapur 2011; Mercer 2008; Reimann 2009). There are at least two main arguments that have been put forward in this respect: First, traditional analyses based on coding-and-counting do not account for temporality in verbal data. Second, they do not afford analyzing patterns of learning activities in verbal data.

Traditional coding-and-counting-based studies do not take into account the temporal development of socio-cognitive activities (Reimann 2009; Wegerif and Mercer 1997) in the sense that every instance of an action, such as hypothesis generation, is added to an overall frequency score (see rows "Sum" and "%" on Table 3). By summing each occurrence of the action, such analyses assume that each instance of the action contributes in the same way to learning outcomes (Chiu and Khoo 2005).

However, this violates a persistent finding in educational research: there are often differences between two instances of the same action (Lämsä et al. 2018; Roschelle and Teasley 1995; Shaffer 2006). For example, in the discussion presented in Table 2, two learners discuss possible reasons (Code "HG") for the problem of an underperforming student. While Learner A generates a very similar hypothesis in both Line 2 and 7, there is a clarification phase in-between. During this clarification phase, Learner B reframes the discussion by introducing the term "learning strategies," and in Line 7 Learner A changes her words accordingly. As a result, the two instances of hypothesis generation are not the same: the second is explicitly generating a hypothesis in the context of a discussion of learning strategies.

One approach to solving this kind of problem is to use more specifically defined codes, such as "general hypothesis generation" versus "hypothesis generation in response to topics from the class". However, this potentially increases the number of codes dramatically, as well as the difficultly in coding data and achieving good inter-rater reliability.

Traditional coding-and-counting-based analyses also do not take into account that learning actions often occur in relation to each other throughout verbal protocols. As such analyses focus on the occurrence of single actions in verbal data, they do not afford an analysis of their co-occurrence throughout the data. For example, in Transcript 1 from Table 3, a traditional coding-and-counting-based analysis would identify that *hypothesis generation* (HG) occurred three times and *evidence evaluation* (EE) occurred seven times, and this difference might be relevant to a researcher. However, as the table shows, the hypothesis generation occurs in the context of evidence evaluation: for example, in Lines 3 and 4 of Transcript 1.

Measuring isolated variables as "components" of learning is already a widely recognized problem in the literature (Jeong 2005; Klahr and Dunbar 1988; Suthers 2005; Wise and Chiu 2011). Researchers (Reimann 2009; Shaffer 2017) thus, highlight the need of accounting for the connections and patterns of connections among such, in fact, *interdependent* activities to capture "higher-level" entities such as learning. Neglecting the temporal relationships between learning activities may pose severe limitations for the analysis and its generalizability with respect to learning. For example, a researcher may use code frequencies to show which actions are correlated to each other in a discourse. The summed occurrence of *hypothesis generation* (HG) and *generating solutions* (GS) in Transcript 2 are higher than in Transcript 1. At the same time, the sum of *evidence evaluation* (EE) is lower. If these counts show a systematic tendency across multiple transcripts, then a traditional coding-and-counting-based approach might indicate that *hypothesis generation* (HG) is more strongly associated with *generating solutions* (GS) and negatively correlated with *evidence evaluation* (EE). Yet, such an analysis would miss information of how those activities co-occur *within* each transcript.

In contrast to such a traditional coding-and-counting strategy, a consideration of temporal co-occurrences can reveal patterns of actions within time- or event-intervals that are not apparent at the level of raw code frequencies (Dyke et al. 2012; Shaffer 2017; Siebert-Evenstone et al. 2016). Such intervals can be, for example, seconds or minutes, or pairs or triplets of propositions. This kind of approach is shown in Table 3, where subsets of event-pairs are marked. When co-occurrences are identified as event pairs, *hypothesis generation* (HG) and *generating solutions* (GS) do not occur together even once. In contrast, a connection between *hypothesis generation* (HG) and *evidence evaluation* (EE) occurs three times. If these patterns recur systematically throughout transcripts, the researcher could conclude that *hypothesis generation* (HG) and *generating solutions* (GS) are not closely related, but *hypothesis generation* (HG) and *evidence evaluation* (EE) are those socio-cognitive actions that interact with one another. More to the point, these results could be in direct contrast with the correlational strategy of traditional coding-and-counting-based approaches, where the overall frequency of *hypothesis generation* (HG) and *generating solutions* (GS) could be correlated even though the two socio-cognitive actions are never temporally co-located (see the dilemma between "global" vs "local" correlational strategies in Collier et al. 2016; Shaffer & Serlin, 2004).

One statistical consequence of not capturing learning-related actions as they develop over time and co-occur with each other is a loss of statistical power, in the sense that the original data is highly compressed when aggregated to just a few variables. This makes any statistical analysis that is concerned with predicting learning outcomes based on learning actions potentially less sensitive (Shaffer & Serlin, 2004; Kapur 2011). This reduced power could further mean that existing relationships between variables may remain undetected by traditional quantitative techniques based on coding-and-counting (this phenomenon is demonstrated both with real data and in a simulation study in Collier et al. 2016).

Researchers are also often interested in visualizing the results of an analysis. Frequency bars are typically reported in coding-and-counting-based analyses (see Fig. 1), but because frequency data does not account for temporality, such bar graphs are not very informative in this respect. More generally, traditional coding-and-counting-based approaches do not provide the opportunity to visualize how discourse is developing over time (Dyke et al. 2012; Hmelo-Silver et al. 2013; Suthers and Medina 2011).

These limitations of traditional coding-and-counting-bases approaches show that ignoring the temporally developing relationships between socio-cognitive actions of learning may affect the analysis of verbal data. As noted above, it is true that recoding data or introducing a hierarchical coding scheme may be a solution to some limitations of traditional coding-and-counting-based analyses, but it is not a general solution in the sense that it does not directly address the core problem of representing temporality (and the socio-cognitive effects of temporality) in verbal data.

Thus, we argue that CSCL research should look for and develop further methods for analyzing verbal data to address this core problem. Such methods should (1) account for the temporal development of learning actions, (2) address the interdependence between learning actions, (3) afford more powerful quantitative analyses of learning actions, and (4) visualize the dynamics of learning in an insightful manner.

Beyond traditional coding-and-counting-based analyses: Epistemic network analysis as a method to analyze temporal connections of learning activities

CSCL research has developed several methods and tools for the analysis and visual representation of verbal data that take temporality information into account. For example, *sequential analysis* (Cress and Hesse 2013; Jeong 2005; Kapur 2011) recognizes that many actions do not occur independently from each other in verbal data. In fact, an action such as hypothesis generation may be more likely to be followed by a certain action such as evidence generation rather than by another action such as drawing conclusions. The likelihood of such transitions between action pairs is called transition probability. Such transition probabilities can define a "most likely" pattern or sequence of actions across verbal protocols. This way, sequential analysis affords the analysis of activity patterns in the form of sequences.

Verbal data in CSCL has also been analyzed using *process mining* (Reimann 2009). Process mining is not a single method, but rather an approach for developing models and mining data to (a) gain empirical models or to (b) confirm existing models. These models would represent processes of learning, such as self-regulation, including actions of, e.g., planning or progress monitoring (e.g., Bannert et al. 2014).

These methods have several limitations, however. First, the number of possible sequences of actions is extremely large. As a result, these methods require very large data sets, which are not always available (Bakeman and Gottman 1997; Reimann and Yacef 2013). Second, such models are often difficult to interpret, resulting in a set of specific sequences of actions—sometimes a quite large number of such sequences—that are statistically different between one group and another. Yet, explaining the significance of these differences is often a challenge. Moreover, although there are visualizations that are used to represent sequential data, such as transition state diagrams, it is difficult to compare such representations visually.

There are also representational tools in CSCL research that can visualize temporal patterns of verbal data. For example, CORDTRA (Hmelo-Silver et al. 2011) allows a researcher to visually investigate how different activities occur over time in relation to each other. However, such visualization tools do not provide a quantitative measure of the resulting differences, and do not afford analyzing the systematic temporal relationships of events in a larger corpus of data.

For these reasons, in our investigation we chose to use *epistemic network analysis* (ENA; see Marquart et al. 2018; Shaffer and Ruis 2017; Shaffer et al. 2016; Shaffer 2006), a modeling technique that can (1) capture, (2) visualize, (3) quantitatively compare patterns of learning activities across conditions, and (4) be used with smaller datasets. ENA allows researchers to model temporal co-occurrences between socio-cognitive actions, visualize those co-occurrences, and conduct statistical comparisons between different groups of learners with respect to those models.

The theoretical background of ENA is Epistemic Frame Theory (Shaffer 2017; Shaffer 2007). Epistemic Frame Theory assumes that learning cannot be reduced to isolated components such as specific actions in the learning process. Rather, learning is the transformation of an individual's epistemic network: a set of relationships that connects skills, knowledge, and values that a learner uses to make sense of and take action on the world. This network is expressed in discourse and changes over time during the learning process (Shaffer 2012). ENA has been used in diverse research settings, including (a) surgery trainees' operative performance during a simulated procedure (Ruis et al. 2018), (b) gaze coordination during collaborative work (Andrist et al. 2015), (c) communication among health care teams (Sullivan et al. 2018), and more generally in situations where researchers want to analyze the integration of interconnected skills in contrast to the "isolationist" methodological approach that is often used in traditional coding-and-counting-based analyses (Arastoopour et al. 2016; Collier et al. 2016; Eagan and Hamilton 2018).

Like analyses following a traditional coding-and-counting approach, ENA begins with verbal data that has been segmented and coded. However, instead of computing the mere frequencies of single codes, ENA analyzes the data segment-by-segment to identify if certain actions occur either within the same segment or in neighboring segments. (More detailed explanations of the mathematics of ENA can be found in Shaffer (2017) as well as in Shaffer and Ruis (2017); Siebert-Evenstone et al. (2017); Shaffer et al. (2016).

The researcher can therefore identify how far the actions she is interested in may fall from each other in the discourse. For example, in Table 3, action pairs were marked for actions in one-step-distance from each other. ENA can analyze connections with different scopes, however: for example, activities that fall within a *window* of any size in the data, such as two, or five, or eight steps from each other (Siebert-Evenstone et al. 2016). Table 3 shows a window size of two (each event is analyzed in the context of the event that immediately preceded it), meaning that ENA counts occurrences of pairs of adjacent events. If two events occur repeatedly over time in the discourse, the connection between these events is stronger in the ENA model.

After analyzing all neighboring segments in a verbal protocol, a researcher might find several connections that organize themselves into a pattern: that is, into an epistemic network where some pairs of events are strongly connected (they co-occurred more often in the discourse) while others are weaker (they co-occurred less often in the discourse). Epistemic networks can be aggregated into *mean networks* across individuals, such as the mean network of all individual problem solvers or the mean network of all dyadic problem solvers in our data. And networks can be compared by subtracting their connection weights in one network from the weights in the other. The resulting *subtracted network* represents the differences between two epistemic networks. So, for example, by subtracting the mean network for individual problem solvers from the mean network for dyads in our data, it is possible to visualize and quantify the differences between collaborative and individual reasoning.

ENA also provides a method for performing statistical tests on epistemic networks. First, a high-dimensional space is generated where the dimensions represent each pair of possible connections within the networks. Through a dimensional reduction method, that is, *single value decomposition* (Shaffer et al. 2016), the space of networks can be reduced to a simpler, one-, or two-dimensional *projected ENA space*. (See Shaffer et al. 2016 for a more detailed explanation of the mathematics; see Arastoopour et al. 2016 and Sullivan, 2018 for examples of this kind of analysis). Then the resulting one- or two-dimensional values representing different networks in the projected ENA space can be included in further quantitative analysis. For example, the networks of dyadic reasoners and those of individual reasoners can be compared to see if they the differences between them are statistically significant: that is, if dyadic networks are quantitatively different from individual networks.

A key feature of an ENA model is that networks are visualized using network graphs, where nodes correspond to the codes, and edges reflect the relative frequency of co-occurrence, or connection, between two codes. But critically, the positions of the network graph nodes are fixed, meaning all networks in a given ENA space have the same node locations. In addition, *those positions are determined by an optimization routine* that minimizes the difference, for any given network, between the point that represents that given network in the projected ENA space and the *centroid* or center of mass of the same network, computed from the weights of the connections in the network. In other words, the optimization minimizes the difference between the point that represents a network in the projected ENA space and the network centroid for every units in the set. Thus, if two learners' epistemic networks show similar patterns of connections, their centroid values will fall close to each other in the projected ENA space (Shaffer et al. 2016).

Optimizing the position of network nodes in this way creates a *co-registration* of network graphs and projected ENA space from the dimensional reduction. As a result, the positions of the network graph nodes—and the connections they define—can be used to interpret the dimensions of the projected space and explain the positions of plotted points in the space. This makes it possible to conduct quantitative comparisons between two sets of networks in the projected ENA space (in our case, dyadic vs individual networks), and then *interpret* the differences between networks using their corresponding network graphs.

Research questions

Following concerns raised in the CSCL literature (e.g, Kapur 2011; Reimann 2009; Shaffer 2017), we have argued that traditional coding-and-counting-based analyses are limited for in-depth quantitative analysis of verbal data, both regarding the analytical process and the visualization of verbal data. ENA may add to such purely frequentist approaches as it accounts for these limitations. In what follows, we compare these approaches empirically and set the following research questions:

> RQ1: Which technique provides the best explanation of group differences with respect to learners' engagement in different learning actions?

To investigate this question, we analyze to what extent dyads and individuals differ from each other with respect to the learning activities they engage in during their problem solving. We first conduct a traditional coding-and-counting-based analysis, followed by an ENA analysis, and then compare the outcomes of the two approaches. We hypothesize that the results of ENA will reveal information that the more traditional coding-and-counting-based approach did not capture with respect to the way learners engage in learning activities.

> RQ2: To what extent are the results from RQ1 due to systematic temporal co-occurrences between learning actions?

To investigate this question, we compare the epistemic networks resulting from the analysis of RQ1 with epistemic networks generated by randomized version of the same data. Randomizing within each verbal protocol removes temporality from the data by mixing up the order of coded segments. We hypothesize that the ENA results on RQ1 will differ from those of RQ2, showing that the findings from the ENA analysis in RQ1 cannot be explained by the frequency distributions of learning actions, but also reflect the temporality information in the original data.

Method

Participants and design

The data analyzed in this study, the coding procedure and its outcomes come from a previous work (Csanadi et al. 2016). In the original study, pre-service teachers ($N = 76$; 59 female, $M_{\text{Age}} = 21.22$, $SD_{\text{Age}} = 3.98$) solved an educational problem case from their future professional field (teaching) in one of two between-subject conditions: either as individuals ($N = 16$) or as dyads ($N = 30$ dyads).

Data

In Csanadi et al. (2016), verbal problem-solving data (think aloud data from participants in the individual condition and discourse data of participants in the dyadic condition) were audio-recorded and transcribed.
The transcriptions were segmented into propositional units (Coder 1 agreed on 85.09% of the segments of Coder 2; Coder 2 agreed on 79.73% of the segments of Coder 1; Strijbos et al. 2006). Coding was based on a coding scheme developed by Csanadi et al. (2015), following a typology of epistemic actions, that is, *epistemic processes* of scientific reasoning that was suggested by Fischer et al. (2014). Based on this taxonomy, they distinguished between eight different epistemic processes: problem identification (PI), questioning (Q), hypothesis generation (HG), generating solutions (GS), evidence generation (EG), evidence evaluation (EE), drawing conclusions (DC), and communicating and scrutinizing (CS), as well as non-epistemic propositions (NE). It is important to note that the two independent coders often found it problematic to distinguish between cases when evidence was used to support a claim (EE) or was used for another epistemic purpose (EG). As a result, the two coders had most of their misclassification between evidence evaluation (EE) and evidence generation (EG). Thus, they merged these two codes into one: evidence evaluation (EE). Using this scheme, two raters independently coded 10 % of the material was randomly chosen and both coders independently applied the coding scheme to the material ($\kappa = 0.68$). Afterwards, a single rater coded the remaining data.
To answer RQ2, in the present study we created a *randomized dataset* of the previously segmented propositions (see Table 4) within each dyad and individual participant.
Table 4 illustrates our randomized dataset. This table contains the randomized version of the first six lines from Table 1. Note, that the frequencies of codes in case of Table 1 and Table 4 are the same. Yet, in Table 1, the codes follow each other in the temporal order of their occurrence, while in Table 4, such temporality does not exist.

Analysis

To answer RQ1, we applied five frequency-based inferential statistics (MANOVA, ANOVAs, Welch-test, Chi-square test, and correlations) and ENA to compare the outcomes of the two methodological approaches.

For the ENA model, we used a window size of two. We chose a dimensional reduction that maximized the difference between the mean of units (participants or dyads) in the two conditions. The value of each network in the projected ENA space (described above) was included as dependent variable to compare dyadic and individual epistemic networks of scientific reasoning. Mean networks were computed for both the dyadic and the individual reasoning conditions, respectively, and we constructed a subtracted network by subtracting the mean connection strengths for participants in the dyadic condition from the mean connection strengths for participants in the individual conditions. The resulting subtracted network showed what connections contributed to the differences between the two conditions.

To answer RQ2, we used the randomized dataset selecting the same parameters and performing the same ENA analysis as in RQ1. We compared the outcomes of this analysis with the ENA results from RQ1.

Results

> RQ1: Which technique provides the best explanation of group differences with respect to learners' engagement in different socio-cognitive actions?

Coding-and-counting

We compared the frequency of engagement in the different socio-cognitive actions across the two conditions (collaborative vs. individual). The difference in frequency of events between the two conditions was statistically significant overall (Pillai's trace = .40, $F(5,40) = 5.26$, $p < .001$, partial $\eta^2 = .40$). Subsequent univariate comparisons showed that participation in the collaborative condition resulted in a significantly higher engagement in *hypothesis generation* ($M = .24$, $SD = .09$) in contrast to the individual condition ($M = .17$, $SD = .11$), $F(1, 44) = 6.06$, $p < .05$, partial $\eta^2 = .12$. Also, engagement for dyads was higher in *evidence evaluation* ($M = .33$, $SD = .11$) in contrast to the individual condition ($M = .26$, $SD = .13$) $F(1, 44) = 4.28$, $p < .05$, partial $\eta^2 = .09$. Similarly, the odds to engage in *drawing conclusions* were 5.43 times higher for the collaborative than for the individual condition, $\chi^2(1) = 4.51$, $p < .05$. At the same time, collaboration led to a significantly lower engagement in *generating solutions* ($M = .29$, $SD = .13$) in comparison to the individual condition ($M = .45$, $SD = .24$), Welch's $F(1, 19.79) = 6.56$, $p < .05$, partial $\eta^2 = .17$.

We also correlated the frequency of epistemic processes within each condition. In the dyadic condition, *generating solutions* was negatively correlated with *evidence evaluation* ($r = .65$, $p < .001$) and *hypothesis generation* ($r = .43$, $p < .05$), and *problem identification* was negatively correlated with *communicating and scrutinizing* ($r = .38$, $p < .05$). In the individual condition, *generating solutions* was negatively correlated with *evidence evaluation* ($r = .87$, $p < .0001$), *hypothesis generation* ($r = .70$, p

$< .01$) and *non-epistemic activities* ($r = .50$, $p < .05$). *Problem identification* and *non-epistemic activities* were positively correlated ($r = .52$, $p < .05$).

ENA

As described above, we compared dyadic and individual networks. The mean centroid value for individuals' epistemic networks ($M = .21$, $SD = .32$) was significantly different from the mean centroid value for dyads' epistemic networks ($M = .11$, $SD = .21$, $t(44) = 3.65$, $p < .01$, $d = 1.32$). So, as with the traditional coding-and-counting approach, there were differences between individuals and dyads.

However, the mean network graphs for both groups (Fig. 2) showed different relationships among epistemic practices than the traditional coding-and-counting-based analytical approach. For dyadic conversations, *evidence evaluation* was central to the problem solving process: it formed connections with *hypothesis generation*, *communicating and scrutinizing*, *generating solutions* and *non-epistemic propositions*. In case of the more traditional coding-and-counting-based strategy, in contrast, only one of these four connections was visible: between evidence evaluation and generating solutions. Furthermore, while the correlation analysis indicated a strong negative association between *evidence evaluation* and *generating solutions* overall, ENA showed that there was temporal co-occurrence between the two (Fig. 2). Finally, while correlations indicated that *communicating and scrutinizing* is negatively correlated with problem identification, ENA showed temporal co-occurrence between *communicating and scrutinizing* and *evidence evaluation*.

In the case of individuals, ENA showed that *evidence evaluation* was strongly connected to *hypothesis generation* and *generating solutions*. But in contrast with the dyadic condition, individual networks did not have a clear central node. Rather, each of the three most frequent (Fig. 2) epistemic practices were connected to each other. In the case of individual problem solvers, the more traditional coding-and-counting-based analysis identified relationships between *non-epistemic propositions* and *problem identification* as well as between *non-epistemic propositions* and *generating solutions*. However, this coding-and-counting-based strategy did not show the connection between *hypothesis generation* and *generating solutions*.

Subtracting individual from dyadic networks revealed that, in comparison to individuals, dyadic conversations were strongly characterized by the connections between *evidence evaluation* and *communicating and scrutinizing* as well as between *evidence evaluation* and *hypothesis generation*. However, these relationships were not significant for the correlational findings from the coding-and-counting-based approach. In contrast, connections from *solution generation* to both *hypothesis generation* and *evidence evaluation* were stronger for individuals than for dyads.

In general, ENA showed that *evidence evaluation* was more central to the problem solving of dyads than individuals. The correlations that we found via the traditional coding-and-counting-based approach we used, however, suggested the opposite: evidence evaluation was negatively correlated with other epistemic processes.

On the one hand, the traditional coding-and-counting-based approach and ENA found the same activities to be central for each condition: evidence evaluation for dyads and generating solutions for individuals. However, our coding-and-counting-based analyses did not show the same structure of associations between different epistemic processes, particularly in the case of dyads.

> RQ2: To what extent are the results from RQ1 due to systematic temporal co-occurrences between learning activities?

To determine if the epistemic networks resulting from the analysis of RQ1 are due to temporal connections between socio-cognitive actions and not merely their frequency distribution in the data, we compared epistemic networks resulting from the analysis of RQ1 with epistemic networks resulting from the analysis of the randomized dataset.

The quantitative outcomes with the *randomized dataset* showed that the mean centroid value for the individuals' epistemic networks ($M = .17$, $SD = .26$) still significantly different from the mean centroid value for dyads' epistemic networks ($M = .09$, $SD = .20$), $t(44) = 3.35$, $p < .01$, 95%, $d = 1.15$), which prior work suggests is not surprising, as data with differences in frequencies of codes will also show differences in connections between codes if the data is randomly ordered (Collier et al. 2016).

However, the mean epistemic networks from the two conditions in the randomized data (Fig. 3) showed that in both conditions, participants only made connections among the three most frequent socio-cognitive activities (*hypothesis generation*, *solution generation* and *evidence evaluation*: compare to Fig. 1). Thus, dyadic and individual networks showed no structural differences from each other. These results were, therefore, in clear contrast with the results on the original dataset (Fig. 2) where dyadic and individual networks showed different structures of association between epistemic processes. The randomized data set suggests that there is no central epistemic practice in either dyadic or individual conditions. Finally, the subtracted network model in Fig. 3 consists of only blue lines (representing connections for dyads), indicating that ENA identified more connections among the highly frequent codes for dyads than for individuals. The results thus show that the epistemic networks captured from the analysis of RQ1 cannot be reduced only to the frequency distributions of epistemic practices. They reflect temporality information in the original data.

Discussion

CSCL research often compares different groups with respect to their learning, including quantitative analysis of verbal process data. CSCL researchers often conduct such analyses using one or more coding-and-counting strategies, such as summarizing frequencies of occurrences, and conducting ANOVAs or correlational analyses (e.g., Vogel and Weinberger 2018). However, using such traditional coding-and-counting-based techniques is often a suboptimal choice because it does not account for temporality in verbal data (Reimann 2009).

The present work aimed to (a) summarize the main limitations of traditional coding-and-counting-based approaches, (b) survey methodological solutions that account for temporality, and (c) empirically test ENA as a methodological addition to traditional coding-and-counting-based analyses to identify temporal structure of relationships between codes in learning activities.

Our analyses show that ENA revealed relationships in the data that were not found by analyses that were based on a traditional coding-and-counting approach. Through ENA, we were able to identify temporal patterns between socio-cognitive events in verbal problem-solving protocols. ENA helped us to (a) model patterns in the temporal co-occurrence between socio-cognitive events over time, (b) visualize the structure of those temporal co-occurrences in the form of epistemic networks, (c) quantify those patterns, (d) statistically compare our two conditions (individual and dyadic), and (e) use co-registered network visualizations to interpret how the patterns we identified differ from one another.

Compared to more traditional analyses based on coding-and-counting, ENA showed that *evaluating evidence* was a central epistemic practice for dyads but not for individual problem solvers. More specifically, *evaluating evidence* was associated with all of the other epistemic practices. This suggests that collaborating partners argued in a more evidence-focused manner than individuals did. Dyads referred to hypotheses and evidence more frequently in temporal proximity, and made more temporal connections between communicating and scrutinizing in the problem-solving process.

The outcomes on RQ2 further show that these results can be attributed to temporality in the data. The pairwise frequency comparisons based on traditional coding-and-counting (RQ1) showed that *evidence evaluation* was the most frequent learning action in case of dyads compared to individuals. To test whether the frequency of *evidence evaluation* alone made *evidence evaluation* a central epistemic practice for dyads, we compared our results in RQ1 to an ENA model of data where temporality information was removed through randomization of the data. The model with randomized data did not show the same pattern of connections as the original model, demonstrating that the frequency of *evidence evaluation* alone does not explain the connections between learning activities for dyads. To put it simply: Temporality mattered.

A second significant finding is that correlation-based analyses of coding-and-counting showed different relationships between learning activities than ENA captured. A correlation-based analysis did not show a relationship between *hypothesis generation* and *evidence evaluation* for dyads, which was the strongest temporal connection identified by the ENA model. This is notable because earlier qualitative data analysis on the same dataset showed that the relationship between hypothesis generation and evidence evaluation is a particularly important feature of dyadic conversations in this setting (Csanadi et al. 2016). Thus, the ENA findings quantify a salient feature of the qualitative data that is not accounted for by a traditional coding-and-counting correlation analysis. We argue, therefore, that these results further demonstrate the power of an analytical approach that accounts for temporality. The outcomes of the present article thus suggest that for collaborative reasoning, temporal models provide a better account of problem solving than code frequency models.

Finally, what can we conclude from the results about scientific reasoning in groups in comparison to individuals? The learning activities we investigated in this study were developed (Csanadi et al. 2015) based on a theoretical framework on scientific reasoning (Fischer et al. 2014). This theoretical framework proposed eight activities of scientific reasoning. However, the framework did not propose a theoretical model with respect to what patterns of engagement in these activities were more effective than others. Thus, the work here also represents an initial step toward exploring patterns with respect to learners' engagement in such epistemic practices of scientific reasoning in the context of collaborative and individual problem solving. The epistemic networks resulting from our study are quantitative empirical models that describe temporal enactment epistemic practices of scientific reasoning during problem solving.

In sum, through ENA we were able to identify complex temporal relationships in verbal data. Using ENA to model temporal co-occurrences between socio-cognitive events allowed us to (a) build quantitative models of our data, (b) visualize those models in an interpretable manner, and (c) quantitatively compare models between different groups of learners. This comparison (d) provided more detail than the traditional coding-and-counting-based approach in modeling the structure of connections between socio-cognitive events, and (e) was aligned with the findings from earlier qualitative analyses of the same data (Csanadi et al. 2016). We thus suggest that ENA in particular—and models that incorporate the temporal structure of discourse more generally—should be an important part of the toolkit of CSCL researchers, and provide a powerful addition to the widely-used approach of purely frequency-oriented coding-and-counting.

Limitations and conclusions

There are, of course, limitations to this study, as there are in any study. Four of them are mentioned in the following:

First, this work focused on the analysis of process data, but we did not include learning outcomes in the analysis. Often in CSCL research, process data is used to predict learning outcomes, that is to see how learning processes moderate the effects of learning conditions and learning outcomes. Therefore, our findings are limited in that respect that we cannot say much about how different epistemic network models may or may not have led to potentially better learning outcomes. Thus, future studies should more directly address this question.

Second, we could have used more advanced measures that would be based on a traditional coding-and-counting strategy, for example, hierarchical modelling. However, our aim was to focus on analyses that are typically applied in coding-and-counting-based research (specifically ANOVAs and correlations). To triangulate our analyses, we simulated the results of an a-temporal analysis by modeling randomized data with ENA. The results led to similar outcomes as the more traditional coding-and-counting approach: in ENA, the most frequently-occurring events formed connections in the case of the randomized dataset. Thus, despite any limitations of our coding-and-counting models, we demonstrated that temporality captures significant connections between socio-cognitive events beyond those found in a frequency-based analysis.

Third, our ENA-analysis modeled events occurring in direct succession to one another. However, it might be that other window sizes would capture different problem-solving processes. Other studies have looked at how to identify the optimal scope of analysis when analyzing verbal data with ENA (Csanadi et al. 2017; Ruis et al. 2018).

Fourth, our findings were limited by the fact that we were comparing two different data collection methods (recordings of discussion versus recordings of a think-aloud protocol) as well as different problem-solving conditions (dyadic versus individual). While prior research has shown that such comparison may not compromise the results in theory (Csanadi et al. 2016; Mullins et al. 2011; Teasley 1995), the differences between ENA and a more traditional coding-and-counting approach were more pronounced for dyads than for individual problem solvers. There are examples of analyzing temporality in the context of individual learning (e.g., Bannert et al. 2014), but many studies that use temporal patterns focus on collaborative learning contexts (Jeong 2005; Kapur 2011). Thus, more studies are needed to see to what extent temporality is less of a factor in individual problem solving as recorded in think-aloud protocols in comparison to dyadic problem solving as captured through discussion.

Despite these limitations, our results suggest that traditional coding-and-counting-based approaches are limited in their ability to model temporality in verbal process data. We do not suggest that such coding-and-counting approaches should be abandoned, because counting frequencies of occurrences may reveal important information to a researcher. However, we do argue that any analysis that aims to understand how learners engage in activities and how such engagement contributes to learning needs to also use analytical approaches that account for temporal characteristics of data. Temporal analyses are not a luxury that CSCL researchers might choose to enjoy or not; rather, they are an analytical necessity for researchers interested in generating meaningful analyses of collaborative learning. Based on our research, we can say that ENA is a powerful means to perform such analyses.

JrnlID 11412_ArtID 9288_Proof# 1 - 07/11/2018

International Journal of Computer-Supported Collaborative Learning
https://doi.org/10.1007/s11412-018-9288-8

When coding-and-counting is not enough: using epistemic network analysis (ENA) to analyze verbal data in CSCL research

Andras Csanadi [1] **· Brendan Eagan** [2] **· Ingo Kollar** [3] **· David Williamson Shaffer** [2] **· Frank Fischer** [4]

Received: 13 November 2017 / Accepted: 5 November 2018

Abstract
Research on *computer-supported collaborative learning* (CSCL) is often concerned with the question of how scaffolds or other characteristics of learning may affect learners' social and cognitive engagement. Such engagement in socio-cognitive activities frequently materializes in discourse. In quantitative analyses of discourse, utterances are typically coded, and differences in the frequency of codes are compared between conditions. However, such traditional *coding-and-counting-based strategies* neglect the *temporal* nature of verbal data, and therefore provide limited and potentially misleading information about CSCL activities. Instead, we argue that analyses of the temporal proximity, specifically temporal co-occurrences of codes, provide a more appropriate way to characterize socio-cognitive activities of learning in CSCL settings. We investigate this claim by comparing and contrasting a traditional coding-and-counting analysis with *epistemic network analysis* (ENA), a discourse analysis technique that models temporal co-occurrences of codes in discourse. We apply both methods to data from a study that compared the effects of individual vs. collaborative problem solving. The results suggest that compared to a traditional coding-and-counting approach, ENA provides more insight into the socio-cognitive learning activities of students.

Keywords Discourse analysis · Coding-and-counting · Epistemic network analysis · Problem solving

✉ Andras Csanadi
andras.csanadi@unibw.de

Brendan Eagan
beagan@wisc.edu

Ingo Kollar
ingo.kollar@phil.uni-augsburg.de

David Williamson Shaffer
dws@education.wisc.edu

Frank Fischer
frank.fischer@psy.lmu.de

Introduction

A major goal of research in *computer supported collaborative learning* (CSCL) is to understand how to use technology to improve collaborative learning. For example, Bause et al. (2018) investigated whether a particular design of a multitouch table that separates a private from a joint screen area is more effective for groups working on a problem-solving task than a design that does not include a joint working space. Likewise, many empirical studies look at whether CSCL scripts evoke different socio-cognitive actions than unscripted CSCL (e.g., Schwaighofer et al. 2017).

Central to such studies is the analysis of how differently designed learning environments impact how students interact during learning. For that purpose, researchers often rely on verbal data that are captured during learning, such as transcripts of within-group talk. These data are then analyzed to model how different learning conditions impact learners' actions, such as developing explanations or evaluating evidence (Teasley 1995).

Such analyses are typically based on *coding-and-counting* (e.g., Vogel and Weinberger 2018). In this approach, a researcher (1) develops a coding scheme to identify different actions that occurred during learning; (2) applies that coding scheme to the data corpus; and (3) typically counts the frequencies by which learners in different experimental conditions engaged in these actions. Frequency-based methods of this coding-and-counting-strategy thus provide a means for comparing the effects that different conditions have on the learners' socio-cognitive actions.

Despite its wide adoption in the CSCL community, however, coding-and-counting-based analyses as the one just described have been repeatedly criticized in CSCL research (Kapur 2011; Reimann 2009). In particular, critics of such an approach argue that (1) it ignores *temporality* in verbal data, and (2) it does not afford analyzing *patterns* of learning activities. That is, such traditional coding-and-counting-based approaches model the frequency of each kind of learner action (each code), but do not provide information about whether and how these actions might be related to one another.

For example, during collaboration, learners often develop questions and expectations that guide their interaction with each other and with the learning material. Counting how often each learner formulates questions and also counting independently how often each learner refers to the learning material tells us nothing about whether the learners have made *connections* between their questions and the learning material over time. We thus argue that using traditional coding-and-counting-based techniques as described above is often a suboptimal strategy to model learning in verbal data. In many cases, a more appropriate and informative approach is to use methods that model *temporal relationships* between coded socio-cognitive actions in verbal data.

In this article, we compare a traditional coding-and-counting-based analysis of a data corpus to epistemic network analysis (ENA; Shaffer et al. 2009; Shaffer 2017), an analysis method that models temporality in verbal data. We apply both a typical coding-and-counting approach and an ENA analysis on the same data set, and then examine the inferences that can be drawn from the two analyses.

To further investigate the impact of failing to account for temporality in the analysis of verbal data from a CSCL environment, we also compare the results of ENA on the original data set with the results of ENA on a *randomized version* of the original dataset. Randomizing the order of coded learning actions within each transcript preserves the *frequency* of occurrence of learner' actions in a verbal protocol, but eliminates temporal information from the original transcripts. Therefore, comparing the original data set to a randomized data provides an opportunity to understand more deeply the impact of temporality on the learning activities being modelled.

Engaging in socio-cognitive activities during CSCL: An example

The data we use to address these questions comes from an experiment in which pre-service teachers were asked to reason about a pedagogical problem (Csanadi et al. 2016). In one condition, students were asked to discuss the problem in pairs; in the other condition, students reflected on the problem individually using a *think aloud protocol* (e.g., Ericsson and Simon 1980; Fox et al. 2011). Using transcripts of discourse, Csanadi et al. (2016) investigated whether and how participants' engagement in actions of scientific reasoning such as hypothesizing and evaluating evidence, differed between the two conditions.

Tables 1 and 2 show two excerpts from this study. In what follows, we will refer to these two examples to describe how both traditional coding-and-counting approaches and ENA model this data.

The transcripts from dyadic discussions and individual think-aloud protocols were segmented into propositional units, and each proposition was coded (Csanadi et al. 2016) using a coding scheme developed by Csanadi et al. (2015) based on a heuristic framework of scientific reasoning (Fischer et al. 2014). The coding scheme identifies one of eight kinds of epistemic actions for each propositional unit:

(1) *Problem Identification* (PI): an initial attempt to build an understanding of the problem
(2) *Questioning* (Q): statements or questions triggering further inquiry
(3) *Hypothesis Generation* (HG): developing explanations of the problem
(4) *Generating Solutions* (GS): developing interventions or solution plans
(5) *Evidence Generation* (EG): reference to information or lack of information that could support a claim
(6) *Evidence Evaluation* (EE): evaluating a claim

Table 1 Excerpt from a think aloud protocol (individual condition)

Line	Excerpt from segmented transcript	Code
1	Well, I would first inform myself, //	EE
2	what can be the reason, //	Q
3	that she is not so good at the exams. //	EE
4	If it can be her learning method, //	HG
5	or perhaps she learns well //	HG
6	but then she always has exam-anxiety. //	HG
7	There can be many reasons for it, //	NE
8	and one should tell it in the context, //	EE
9	if there is not anything special. //	EE
10	The parents say she learns diligently at home… //	EE
11	I would then look up some books //	EE
12	and I would write out a couple of things. //	EE
13	For example, I would recommend her //	GS
14	that she should have a learning plan for the homework //	GS

Table 2 Excerpt from a dyadic discourse (collaborative condition)

Line	Excerpt from segmented transcript	Code
1	A: I think it may rather be that although she learns a lot, //	EE
2	yet, she learns it in the wrong way. //	HG
3	B: That she has the wrong learning strategies. //	HG
4	A: Exactly, she studies in a wrong way. I mean… //	HG
5	B: That she does not elaborate, //	HG
6	rather learns by heart. //	HG
7	A: Exactly, she learns the whole stuff superficially. //	HG
8	I mean, of course, I can recite something to myself for hours, //	EE
9	but when I don't understand it, //	EE
10	it won't stay long in memory. //	EE
11	B: In that case you could try some counselling with him, //	GS
12	to find the right learning strategies, //	GS
13	A: Right.	
14	B: how she learns best. //	GS

(7) *Communicating and Scrutinizing* (CS): planned discussions with others (e.g., in order to find out further information)

(8) *Drawing Conclusions* (DC): concluding outcomes of reasoning

More specific details of segmentation and coding are discussed in the methods section below.

Measuring socio-cognitive activities by a traditional coding-and-counting approach

Both traditional coding-and-counting-based approaches and an ENA analysis begin with a *coding phase*. In the coding phase, researchers identify socio-cognitive actions that are relevant to the research question at hand. Then, they develop a coding framework to capture those actions in the data, and apply the framework to the data. The whole procedure may, in fact, include several steps and iterations of those steps (see e.g., Chi 1997; Strijbos et al. 2006; Vogel and Weinberger 2018; Shaffer 2017). The coding scheme we used in this experiment is described briefly above, and in more detail in the methods section.

While both traditional coding-and-counting-based analyses and ENA models use coded data, they differ with respect to what subsequently is done with the coded data. In typical coding-and-counting-based studies, the coding phase is followed by a *counting phase*, in which the researcher chooses units of analysis and computes the *code frequency*—the rate at which a code appears in the data—for each code within the data from each unit of analysis. Differences between code frequencies across units of analysis in different conditions are then analyzed statistically.

In our case, we used a multivariate ANOVA to test whether the two conditions (individual versus dyadic problem solving) had differential effects on learners' problem solving. The results, which we will present in more detail later, were useful to understand *the extent to which* participants engaged in actions of scientific reasoning, and whether the extent to which participants engaged in those actions was different between the two conditions, We could also have used more complex inferential methods based on code frequencies: for example, mediation analysis, which might test whether engagement in certain activities mediates the effect of the experimental conditions on learning outcomes, but that was not the goal of this study.

Although traditional coding-and-counting-based methods are often used in CSCL research (and we ourselves have used them extensively in the past; e.g., Csanadi et al. 2016; Kollar, Fischer, & Slotta, 2007; Stegmann et al. 2012), they have been heavily criticized (Kapur 2011; Mercer 2008; Reimann 2009). There are at least two main arguments that have been put forward in this respect: First, traditional analyses based on coding-and-counting do not account for temporality in verbal data. Second, they do not afford analyzing patterns of learning activities in verbal data.

Traditional coding-and-counting-based studies do not take into account the temporal development of socio-cognitive activities (Reimann 2009; Wegerif and Mercer 1997) in the sense that every instance of an action, such as hypothesis generation, is added to an overall frequency score (see rows "Sum" and "%" on Table 3). By summing each occurrence of the action, such analyses assume that each instance of the action contributes in the same way to learning outcomes (Chiu and Khoo 2005).

However, this violates a persistent finding in educational research: there are often differences between two instances of the same action (Lämsä et al. 2018; Roschelle and Teasley 1995; Shaffer 2006). For example, in the discussion presented in Table 2, two learners discuss possible reasons (Code "HG") for the problem of an underperforming student. While Learner A generates a very similar hypothesis in both Line 2 and 7, there is a clarification phase in-between. During this clarification phase, Learner B reframes the discussion by introducing the term "learning strategies," and in Line 7 Learner A changes her words accordingly. As a result, the two instances of hypothesis generation are not the same: the second is explicitly generating a hypothesis in the context of a discussion of learning strategies.

One approach to solving this kind of problem is to use more specifically defined codes, such as "general hypothesis generation" versus "hypothesis generation in response to topics from the class". However, this potentially increases the number of codes dramatically, as well as the difficultly in coding data and achieving good inter-rater reliability.

Traditional coding-and-counting-based analyses also do not take into account that learning actions often occur in relation to each other throughout verbal protocols. As such analyses focus on the occurrence of single actions in verbal data, they do not afford an analysis of their co-occurrence throughout the data. For example, in Transcript 1 from Table 3, a traditional coding-and-counting-based analysis would identify that *hypothesis generation* (HG) occurred three times and *evidence evaluation* (EE) occurred seven times, and this difference might be relevant to a researcher. However, as the table shows, the hypothesis generation occurs in the context of evidence evaluation: for example, in Lines 3 and 4 of Transcript 1.

Measuring isolated variables as "components" of learning is already a widely recognized problem in the literature (Jeong 2005; Klahr and Dunbar 1988; Suthers 2005; Wise and Chiu 2011). Researchers (Reimann 2009; Shaffer 2017) thus, highlight the need of accounting for the connections and patterns of connections among such, in fact, *interdependent* activities to capture "higher-level" entities such as learning. Neglecting the temporal relationships between learning activities may pose severe limitations for the analysis and its generalizability with respect to learning. For example, a researcher may use code frequencies to show which actions are correlated to each other in a discourse. The summed occurrence of *hypothesis generation* (HG) and *generating solutions* (GS) in Transcript 2 are higher than in Transcript 1. At the same time, the sum of *evidence evaluation* (EE) is lower. If these counts show a systematic tendency across multiple transcripts, then a traditional coding-and-counting-based approach might t3.1 **Table 3** A typical coding-and-counting strategy

	Transcript 1			Transcript 2		
	HG	GS	EE	HG	GS	EE
L1	0	0	1	0	0	1
L2	0	0	0	1	0	0
L3	0	0	1	1	0	0
L4	1	0	0	1	0	0
L5	1	0	0	1	0	0
L6	1	0	0	1	0	0
L7	0	0	0	1	0	0
L8	0	0	1	0	0	1
L9	0	0	1	0	0	1
L10	0	0	1	0	0	1
L11	0	0	1	0	1	0
L12	0	0	1	0	1	0
L13	0	1	0	0	0	0
L14	0	1	0	0	1	0
Sum	3	2	7	6	3	4
%	21	14	50	43	21	29

Note: HG, EE, and *GS* are mutually exclusive codes assigned to each line (*L1, L2…L5*) for each transcripts. 1″ indicates occurrence, "0" indicates absence of a certain code. Row "*Sum*" represents the total while row "*%*" represents the proportional frequency of occurrence of each code within the transcript, following a traditional coding-and-counting-based strategy

indicate that *hypothesis generation* (HG) is more strongly associated with *generating solutions* (GS) and negatively correlated with *evidence evaluation* (EE). Yet, such an analysis would miss information of how those activities co-occur *within* each transcript.

In contrast to such a traditional coding-and-counting strategy, a consideration of temporal co-occurrences can reveal patterns of actions within time- or event-intervals that are not apparent at the level of raw code frequencies (Dyke et al. 2012; Shaffer 2017; Siebert-Evenstone et al. 2016). Such intervals can be, for example, seconds or minutes, or pairs or triplets of propositions. This kind of approach is shown in Table 3, where subsets of event-pairs are marked. When co-occurrences are identified as event pairs, *hypothesis generation* (HG) and *generating solutions* (GS) do not occur together even once. In contrast, a connection between *hypothesis generation* (HG) and *evidence evaluation* (EE) occurs three times. If these patterns recur systematically throughout transcripts, the researcher could conclude that *hypothesis generation* (HG) and *generating solutions* (GS) are not closely related, but *hypothesis generation* (HG) and *evidence evaluation* (EE) are those socio-cognitive actions that interact with one another. More to the point, these results could be in direct contrast with the correlational strategy of traditional coding-and-counting-based approaches, where the overall frequency of *hypothesis generation* (HG) and *generating solutions* (GS) could be correlated even though the two socio-cognitive actions are never temporally co-located (see the dilemma between "global" vs "local" correlational strategies in Collier et al. 2016; Shaffer & Serlin, 2004). Q3

One statistical consequence of not capturing learning-related actions as they develop over time and co-occur with each other is a loss of statistical power, in the sense that the original data is highly compressed when aggregated to just a few variables. This makes any statistical analysis that is concerned with predicting learning outcomes based on learning actions potentially less sensitive (Shaffer & Serlin, 2004; Kapur 2011). This reduced power could further mean that existing relationships between variables may remain undetected by traditional quantitative techniques based on coding-and-counting (this phenomenon is demonstrated both with real data and in a simulation study in Collier et al. 2016).

Researchers are also often interested in visualizing the results of an analysis. Frequency bars are typically reported in coding-and-counting-based analyses (see Fig. 1), but because frequency data does not account for temporality, such bar graphs are not very informative in this respect. More generally, traditional coding-and-counting-based approaches do not provide the opportunity to visualize how discourse is developing over time (Dyke et al. 2012; Hmelo-Silver et al. 2013; Suthers and Medina 2011).

These limitations of traditional coding-and-counting-bases approaches show that ignoring the temporally developing relationships between socio-cognitive actions of learning may affect the analysis of verbal data. As noted above, it is true that recoding data or introducing a hierarchical coding scheme may be a solution to some limitations of traditional coding-and-counting-based analyses, but it is not a general solution in the sense that it does not directly address the core problem of representing temporality (and the socio-cognitive effects of temporality) in verbal data.

Thus, we argue that CSCL research should look for and develop further methods for analyzing verbal data to address this core problem. Such methods should (1) account for the temporal development of learning actions, (2) address the interdependence between learning actions, (3) afford more powerful quantitative analyses of learning actions, and (4) visualize the dynamics of learning in an insightful manner.

Beyond traditional coding-and-counting-based analyses: Epistemic network analysis as a method to analyze temporal connections of learning activities

CSCL research has developed several methods and tools for the analysis and visual representation of verbal data that take temporality information into account. For example, *sequential*

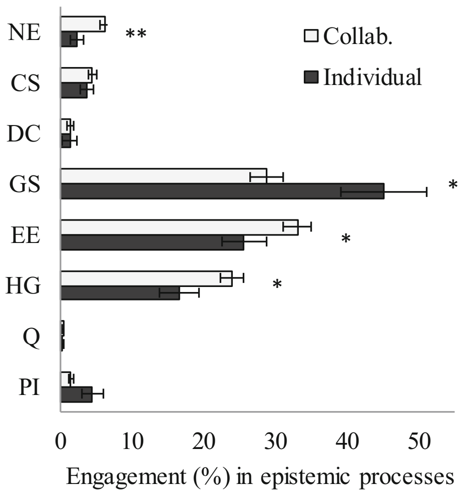


Fig. 1 Proportional frequency (expressed in %) of engagement in the socio- cognitive activities. * $p < .05$, ** $p < .01$

analysis (Cress and Hesse 2013; Jeong 2005; Kapur 2011) recognizes that many actions do not occur independently from each other in verbal data. In fact, an action such as hypothesis generation may be more likely to be followed by a certain action such as evidence generation rather than by another action such as drawing conclusions. The likelihood of such transitions between action pairs is called transition probability. Such transition probabilities can define a "most likely" pattern or sequence of actions across verbal protocols. This way, sequential analysis affords the analysis of activity patterns in the form of sequences.

Verbal data in CSCL has also been analyzed using *process mining* (Reimann 2009). Process mining is not a single method, but rather an approach for developing models and mining data to (a) gain empirical models or to (b) confirm existing models. These models would represent processes of learning, such as self-regulation, including actions of, e.g., planning or progress monitoring (e.g., Bannert et al. 2014).

These methods have several limitations, however. First, the number of possible sequences of actions is extremely large. As a result, these methods require very large data sets, which are not always available (Bakeman and Gottman 1997; Reimann and Yacef 2013). Second, such models are often difficult to interpret, resulting in a set of specific sequences of actions—sometimes a quite large number of such sequences—that are statistically different between one group and another. Yet, explaining the significance of these differences is often a challenge. Moreover, although there are visualizations that are used to represent sequential data, such as transition state diagrams, it is difficult to compare such representations visually.

There are also representational tools in CSCL research that can visualize temporal patterns of verbal data. For example, CORDTRA (Hmelo-Silver et al. 2011) allows a researcher to visually investigate how different activities occur over time in relation to each other. However, such visualization tools do not provide a quantitative measure of the resulting differences, and do not afford analyzing the systematic temporal relationships of events in a larger corpus of data.

For these reasons, in our investigation we chose to use *epistemic network analysis* (ENA; see Marquart et al. 2018; Shaffer and Ruis 2017; Shaffer et al. 2016; Shaffer 2006), a modeling technique that can (1) capture, (2) visualize, (3) quantitatively compare patterns of learning activities across conditions, and (4) be used with smaller datasets. ENA allows researchers to model temporal co-occurrences between socio-cognitive actions, visualize those co-occurrences, and conduct statistical comparisons between different groups of learners with respect to those models.

The theoretical background of ENA is Epistemic Frame Theory (Shaffer 2017; Shaffer 2007). Epistemic Frame Theory assumes that learning cannot be reduced to isolated components such as specific actions in the learning process. Rather, learning is the transformation of an individual's epistemic network: a set of relationships that connects skills, knowledge, and values that a learner uses to make sense of and take action on the world. This network is expressed in discourse and changes over time during the learning process (Shaffer 2012). ENA has been used in diverse research settings, including (a) surgery trainees' operative performance during a simulated procedure (Ruis et al. 2018), (b) gaze coordination during collaborative work (Andrist et al. 2015), (c) communication among health care teams (Sullivan et al. 2018), and more generally in situations where researchers want to analyze the integration of interconnected skills in contrast to the "isolationist" methodological approach that is often used in traditional coding-and-counting-based analyses (Arastoopour et al. 2016; Collier et al. 2016; Eagan and Hamilton 2018).

Like analyses following a traditional coding-and-counting approach, ENA begins with verbal data that has been segmented and coded. However, instead of computing the mere frequencies of single codes, ENA analyzes the data segment-by-segment to identify if certain actions occur either within the same segment or in neighboring segments. (More detailed explanations of the mathematics of ENA can be found in Shaffer (2017) as well as in Shaffer and Ruis (2017); Siebert-Evenstone et al. (2017); Shaffer et al. (2016).

The researcher can therefore identify how far the actions she is interested in may fall from each other in the discourse. For example, in Table 3, action pairs were marked for actions in one-step-distance from each other. ENA can analyze connections with different scopes, however: for example, activities that fall within a *window* of any size in the data, such as two, or five, or eight steps from each other (Siebert-Evenstone et al. 2016). Table 3 shows a window size of two (each event is analyzed in the context of the event that immediately preceded it), meaning that ENA counts occurrences of pairs of adjacent events. If two events occur repeatedly over time in the discourse, the connection between these events is stronger in the ENA model.

After analyzing all neighboring segments in a verbal protocol, a researcher might find several connections that organize themselves into a pattern: that is, into an epistemic network where some pairs of events are strongly connected (they co-occurred more often in the discourse) while others are weaker (they co-occurred less often in the discourse). Epistemic networks can be aggregated into *mean networks* across individuals, such as the mean network of all individual problem solvers or the mean network of all dyadic problem solvers in our data. And networks can be compared by subtracting their connection weights in one network from the weights in the other. The resulting *subtracted network* represents the differences between two epistemic networks. So, for example, by subtracting the mean network for individual problem solvers from the mean network for dyads in our data, it is possible to visualize and quantify the differences between collaborative and individual reasoning.

ENA also provides a method for performing statistical tests on epistemic networks. First, a high-dimensional space is generated where the dimensions represent each pair of possible connections within the networks. Through a dimensional reduction method, that is, *single value decomposition* (Shaffer et al. 2016), the space of networks can be reduced to a simpler, one-, or two-dimensional *projected ENA space*. (See Shaffer et al. 2016 for a more detailed explanation of the mathematics; see Arastoopour et al. 2016 and Sullivan, 2018 for examples of this kind of analysis). Then the resulting one- or two-dimensional values representing different networks in the projected ENA space can be included in further quantitative analysis. For example, the networks of dyadic reasoners and those of individual reasoners can be compared to see if they the differences between them are statistically significant: that is, if dyadic networks are quantitatively different from individual networks.

A key feature of an ENA model is that networks are visualized using network graphs, where nodes correspond to the codes, and edges reflect the relative frequency of co-occurrence, or connection, between two codes. But critically, the positions of the network graph nodes are fixed, meaning all networks in a given ENA space have the same node locations. In addition, *those positions are determined by an optimization routine* that minimizes the difference, for any given network, between the point that represents that given network in the projected ENA space and the *centroid* or center of mass of the same network, computed from the weights of the connections in the network. In other words, the optimization minimizes the difference between the point that represents a network in the projected ENA space and the network centroid for every units in the set. Thus, if two learners' epistemic networks show similar patterns of connections, their centroid values will fall close to each other in the projected ENA space (Shaffer et al. 2016).

Optimizing the position of network nodes in this way creates a *co-registration* of network graphs and projected ENA space from the dimensional reduction. As a result, the positions of the network graph nodes—and the connections they define—can be used to interpret the dimensions of the projected space and explain the positions of plotted points in the space. This makes it possible to conduct quantitative comparisons between two sets of networks in the projected ENA space (in our case, dyadic vs individual networks), and then *interpret* the differences between networks using their corresponding network graphs.

Research questions

Following concerns raised in the CSCL literature (e.g, Kapur 2011; Reimann 2009; Shaffer 2017), we have argued that traditional coding-and-counting-based analyses are limited for in-depth quantitative analysis of verbal data, both regarding the analytical process and the visualization of verbal data. ENA may add to such purely frequentist approaches as it accounts for these limitations. In what follows, we compare these approaches empirically and set the following research questions:

> RQ1: Which technique provides the best explanation of group differences with respect to learners' engagement in different learning actions?

To investigate this question, we analyze to what extent dyads and individuals differ from each other with respect to the learning activities they engage in during their problem solving. We first conduct a traditional coding-and-counting-based analysis, followed by an ENA analysis, and then compare the outcomes of the two approaches. We hypothesize that the results of ENA will reveal information that the more traditional coding-and-counting-based approach did not capture with respect to the way learners engage in learning activities.

> RQ2: To what extent are the results from RQ1 due to systematic temporal co-occurrences between learning actions?

To investigate this question, we compare the epistemic networks resulting from the analysis of RQ1 with epistemic networks generated by randomized version of the same data. Randomizing within each verbal protocol removes temporality from the data by mixing up the order of coded segments. We hypothesize that the ENA results on RQ1 will differ from those of RQ2, showing that the findings from the ENA analysis in RQ1 cannot be explained by the frequency distributions of learning actions, but also reflect the temporality information in the original data.

Method

Participants and design

The data analyzed in this study, the coding procedure and its outcomes come from a previous work (Csanadi et al. 2016). In the original study, pre-service teachers ($N = 76$; 59 female,

$M_{\mathrm{Age}} = 21.22$, $SD_{\mathrm{Age}} = 3.98$) solved an educational problem case from their future professional field (teaching) in one of two between-subject conditions: either as individuals ($N = 16$) or as dyads ($N = 30$ dyads).

Data

In Csanadi et al. (2016), verbal problem-solving data (think aloud data from participants in the individual condition and discourse data of participants in the dyadic condition) were audio-recorded and transcribed.

The transcriptions were segmented into propositional units (Coder 1 agreed on 85.09% of the segments of Coder 2; Coder 2 agreed on 79.73% of the segments of Coder 1; Strijbos et al. 2006). Coding was based on a coding scheme developed by Csanadi et al. (2015), following a typology of epistemic actions, that is, *epistemic processes* of scientific reasoning that was suggested by Fischer et al. (2014). Based on this taxonomy, they distinguished between eight different epistemic processes: problem identification (PI), questioning (Q), hypothesis generation (HG), generating solutions (GS), evidence generation (EG), evidence evaluation (EE), drawing conclusions (DC), and communicating and scrutinizing (CS), as well as non-epistemic propositions (NE). It is important to note that the two independent coders often found it problematic to distinguish between cases when evidence was used to support a claim (EE) or was used for another epistemic purpose (EG). As a result, the two coders had most of their misclassification between evidence evaluation (EE) and evidence generation (EG). Thus, they merged these two codes into one: evidence evaluation (EE). Using this scheme, two raters independently coded 10 % of the material was randomly chosen and both coders independently applied the coding scheme to the material ($\kappa = 0.68$). Afterwards, a single rater coded the remaining data.

To answer RQ2, in the present study we created a *randomized dataset* of the previously segmented propositions (see Table 4) within each dyad and individual participant.

Table 4 illustrates our randomized dataset. This table contains the randomized version of the first six lines from Table 1. Note, that the frequencies of codes in case of Table 1 and Table 4 are the same. Yet, in Table 1, the codes follow each other in the temporal order of their occurrence, while in Table 4, such temporality does not exist.

Analysis

To answer RQ1, we applied five frequency-based inferential statistics (MANOVA, ANOVAs, Welch-test, Chi-square test, and correlations) and ENA to compare the outcomes of the two methodological approaches.

Table 4 Example of the randomized dataset

Line	Excerpt from segmented transcript	Code
4	If it can be her learning method, //	HG
1	Well, I would first inform myself, //	EE
6	but then she always has exam-anxiety. //	HG
2	what can be the reason, //	Q
5	or perhaps she learns well //	HG
3	that she is not so good at the exams. //	EE

For the ENA model, we used a window size of two. We chose a dimensional reduction that maximized the difference between the mean of units (participants or dyads) in the two conditions. The value of each network in the projected ENA space (described above) was included as dependent variable to compare dyadic and individual epistemic networks of scientific reasoning. Mean networks were computed for both the dyadic and the individual reasoning conditions, respectively, and we constructed a subtracted network by subtracting the mean connection strengths for participants in the dyadic condition from the mean connection strengths for participants in the individual conditions. The resulting subtracted network showed what connections contributed to the differences between the two conditions.

To answer RQ2, we used the randomized dataset selecting the same parameters and performing the same ENA analysis as in RQ1. We compared the outcomes of this analysis with the ENA results from RQ1.

Results

> RQ1: Which technique provides the best explanation of group differences with respect to learners' engagement in different socio-cognitive actions?

Coding-and-counting

We compared the frequency of engagement in the different socio-cognitive actions across the two conditions (collaborative vs. individual). The difference in frequency of events between the two conditions was statistically significant overall (Pillai's trace = .40, $F(5,40) = 5.26$, $p < .001$, partial $\eta^2 = .40$). Subsequent univariate comparisons showed that participation in the collaborative condition resulted in a significantly higher engagement in *hypothesis generation* ($M = .24$, $SD = .09$) in contrast to the individual condition ($M = .17$, $SD = .11$), $F(1, 44) = 6.06$, $p < .05$, partial $\eta^2 = .12$. Also, engagement for dyads was higher in *evidence evaluation* ($M = .33$, $SD = .11$) in contrast to the individual condition ($M = .26$, $SD = .13$) $F(1, 44) = 4.28$, $p < .05$, partial $\eta^2 = .09$. Similarly, the odds to engage in *drawing conclusions* were 5.43 times higher for the collaborative than for the individual condition, $\chi^2(1) = 4.51$, $p < .05$. At the same time, collaboration led to a significantly lower engagement in *generating solutions* ($M = .29$, $SD = .13$) in comparison to the individual condition ($M = .45$, $SD = .24$), Welch's $F(1, 19.79) = 6.56$, $p < .05$, partial $\eta^2 = .17$.

We also correlated the frequency of epistemic processes within each condition. In the dyadic condition, *generating solutions* was negatively correlated with *evidence evaluation* ($r = -.65$, $p < .001$) and *hypothesis generation* ($r = -.43$, $p < .05$), and *problem identification* was negatively correlated with *communicating and scrutinizing* ($r = -.38$, $p < .05$). In the individual condition, *generating solutions* was negatively correlated with *evidence evaluation* ($r = -.87$, $p < .0001$), *hypothesis generation* ($r = -.70$, $p < .01$) and *non-epistemic activities* ($r = -.50$, $p < .05$). *Problem identification* and *non-epistemic activities* were positively correlated ($r = .52$, $p < .05$).

ENA

Q4

As described above, we compared dyadic and individual networks. The mean centroid value for individuals' epistemic networks ($M = .21$, $SD = .32$) was significantly different from the mean centroid value for dyads' epistemic networks ($M = -.11$, $SD = .21$, $t(44) = 3.65$, $p < .01$, $d = 1.32$). So, as with the traditional coding-and-counting approach, there were differences between individuals and dyads.

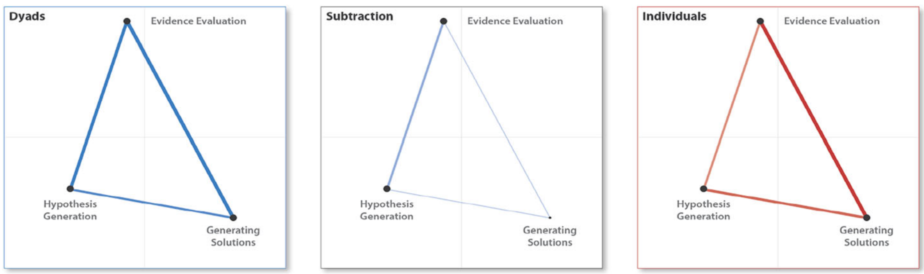
However, the mean network graphs for both groups (Fig. 2) showed different relationships among epistemic practices than the traditional coding-and-counting-based analytical approach. For dyadic conversations, *evidence evaluation* was central to the problem solving process: it formed connections with *hypothesis generation*, *communicating and scrutinizing*, *generating solutions* and *non-epistemic propositions*. In case of the more traditional coding-and-counting-based strategy, in contrast, only one of these four connections was visible: between evidence evaluation and generating solutions. Furthermore, while the correlation analysis indicated a strong negative association between *evidence evaluation* and *generating solutions* overall, ENA showed that there was temporal co-occurrence between the two (Fig. 2). Finally, while correlations indicated that *communicating and scrutinizing* is negatively correlated with problem identification, ENA showed temporal co-occurrence between *communicating and scrutinizing* and *evidence evaluation*.

In the case of individuals, ENA showed that *evidence evaluation* was strongly connected to *hypothesis generation* and *generating solutions*. But in contrast with the dyadic condition, individual networks did not have a clear central node. Rather, each of the three most frequent (Fig. 2) epistemic practices were connected to each other. In the case of individual problem solvers, the more traditional coding-and-counting-based analysis identified relationships between *non-epistemic propositions* and *problem identification* as well as between *non-epistemic propositions* and *generating solutions*. However, this coding-and-counting-based strategy did not show the connection between *hypothesis generation* and *generating solutions*.

Subtracting individual from dyadic networks revealed that, in comparison to individuals, dyadic conversations were strongly characterized by the connections between *evidence evaluation* and *communicating and scrutinizing* as well as between *evidence evaluation* and *hypothesis generation*. However, these relationships were not significant for the correlational findings from the coding-and-counting-based approach. In contrast, connections from *solution generation* to both *hypothesis generation* and *evidence evaluation* were stronger for individuals than for dyads.

In general, ENA showed that *evidence evaluation* was more central to the problem solving of dyads than individuals. The correlations that we found via the traditional coding-and-counting-based approach we used, however, suggested the opposite: evidence evaluation was negatively correlated with other epistemic processes.

On the one hand, the traditional coding-and-counting-based approach and ENA found the same activities to be central for each condition: evidence evaluation for dyads and generating



Q5 **Fig. 2** Epistemic networks of dyads (blue, left), individuals (red, right) and the difference between their networks (center) using the *original dataset* solutions for individuals. However, our coding-and-counting-based analyses did not show the same structure of associations between different epistemic processes, particularly in the case of dyads.

> RQ2: To what extent are the results from RQ1 due to systematic temporal co-occurrences between learning activities?

To determine if the epistemic networks resulting from the analysis of RQ1 are due to temporal connections between socio-cognitive actions and not merely their frequency distribution in the data, we compared epistemic networks resulting from the analysis of RQ1 with epistemic networks resulting from the analysis of the randomized dataset.

The quantitative outcomes with the *randomized dataset* showed that the mean centroid value for the individuals' epistemic networks ($M = .17$, $SD = .26$) still significantly different from the mean centroid value for dyads' epistemic networks ($M = -.09$, $SD = .20$), $t(44) = 3.35$, $p < .01$, 95%, $d = 1.15$), which prior work suggests is not surprising, as data with differences in frequencies of codes will also show differences in connections between codes if the data is randomly ordered (Collier et al. 2016).

However, the mean epistemic networks from the two conditions in the randomized data (Fig. 3) showed that in both conditions, participants only made connections among the three most frequent socio-cognitive activities (*hypothesis generation*, *solution generation* and *evidence evaluation*: compare to Fig. 1). Thus, dyadic and individual networks showed no structural differences from each other. These results were, therefore, in clear contrast with the results on the original dataset (Fig. 2) where dyadic and individual networks showed different structures of association between epistemic processes. The randomized data set suggests that there is no central epistemic practice in either dyadic or individual conditions. Finally, the subtracted network model in Fig. 3 consists of only blue lines (representing connections for dyads), indicating that ENA identified more connections among the highly frequent codes for dyads than for individuals. The results thus show that the epistemic networks captured from the analysis of RQ1 cannot be reduced only to the frequency distributions of epistemic practices. They reflect temporality information in the original data.

Discussion

CSCL research often compares different groups with respect to their learning, including quantitative analysis of verbal process data. CSCL researchers often conduct such analyses

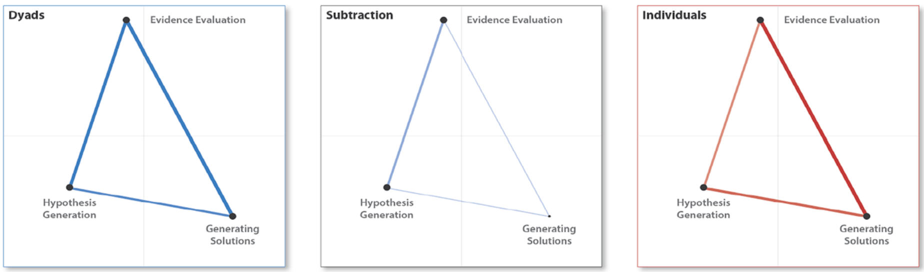


Fig. 3 Epistemic networks of dyads (blue, left), individuals (red, right) and the difference between their networks (center) using the *randomized dataset*

using one or more coding-and-counting strategies, such as summarizing frequencies of occurrences, and conducting ANOVAs or correlational analyses (e.g., Vogel and Weinberger 2018). However, using such traditional coding-and-counting-based techniques is often a suboptimal choice because it does not account for temporality in verbal data (Reimann 2009).

The present work aimed to (a) summarize the main limitations of traditional coding-and-counting-based approaches, (b) survey methodological solutions that account for temporality, and (c) empirically test ENA as a methodological addition to traditional coding-and-counting-based analyses to identify temporal structure of relationships between codes in learning activities.

Our analyses show that ENA revealed relationships in the data that were not found by analyses that were based on a traditional coding-and-counting approach. Through ENA, we were able to identify temporal patterns between socio-cognitive events in verbal problem-solving protocols. ENA helped us to (a) model patterns in the temporal co-occurrence between socio-cognitive events over time, (b) visualize the structure of those temporal co-occurrences in the form of epistemic networks, (c) quantify those patterns, (d) statistically compare our two conditions (individual and dyadic), and (e) use co-registered network visualizations to interpret how the patterns we identified differ from one another.

Compared to more traditional analyses based on coding-and-counting, ENA showed that *evaluating evidence* was a central epistemic practice for dyads but not for individual problem solvers. More specifically, *evaluating evidence* was associated with all of the other epistemic practices. This suggests that collaborating partners argued in a more evidence-focused manner than individuals did. Dyads referred to hypotheses and evidence more frequently in temporal proximity, and made more temporal connections between communicating and scrutinizing in the problem-solving process.

The outcomes on RQ2 further show that these results can be attributed to temporality in the data. The pairwise frequency comparisons based on traditional coding-and-counting (RQ1) showed that *evidence evaluation* was the most frequent learning action in case of dyads compared to individuals. To test whether the frequency of *evidence evaluation* alone made *evidence evaluation* a central epistemic practice for dyads, we compared our results in RQ1 to an ENA model of data where temporality information was removed through randomization of the data. The model with randomized data did not show the same pattern of connections as the original model, demonstrating that the frequency of *evidence evaluation* alone does not explain the connections between learning activities for dyads. To put it simply: Temporality mattered.

A second significant finding is that correlation-based analyses of coding-and-counting showed different relationships between learning activities than ENA captured. A correlation-based analysis did not show a relationship between *hypothesis generation* and *evidence evaluation* for dyads, which was the strongest temporal connection identified by the ENA model. This is notable because earlier qualitative data analysis on the same dataset showed that the relationship between hypothesis generation and evidence evaluation is a particularly important feature of dyadic conversations in this setting (Csanadi et al. 2016). Thus, the ENA findings quantify a salient feature of the qualitative data that is not accounted for by a traditional coding-and-counting correlation analysis. We argue, therefore, that these results further demonstrate the power of an analytical approach that accounts for temporality. The outcomes of the present article thus suggest that for collaborative reasoning, temporal models provide a better account of problem solving than code frequency models.

Finally, what can we conclude from the results about scientific reasoning in groups in comparison to individuals? The learning activities we investigated in this study were developed (Csanadi et al. 2015) based on a theoretical framework on scientific reasoning (Fischer et al. 2014). This theoretical framework proposed eight activities of scientific reasoning. However, the framework did not propose a theoretical model with respect to what patterns of engagement in these activities were more effective than others. Thus, the work here also represents an initial step toward exploring patterns with respect to learners' engagement in such epistemic practices of scientific reasoning in the context of collaborative and individual problem solving. The epistemic networks resulting from our study are quantitative empirical models that describe temporal enactment epistemic practices of scientific reasoning during problem solving.

In sum, through ENA we were able to identify complex temporal relationships in verbal data. Using ENA to model temporal co-occurrences between socio-cognitive events allowed us to (a) build quantitative models of our data, (b) visualize those models in an interpretable manner, and (c) quantitatively compare models between different groups of learners. This comparison (d) provided more detail than the traditional coding-and-counting-based approach in modeling the structure of connections between socio-cognitive events, and (e) was aligned with the findings from earlier qualitative analyses of the same data (Csanadi et al. 2016). We thus suggest that ENA in particular—and models that incorporate the temporal structure of discourse more generally—should be an important part of the toolkit of CSCL researchers, and provide a powerful addition to the widely-used approach of purely frequency-oriented coding-and-counting.

Limitations and conclusions

There are, of course, limitations to this study, as there are in any study. Four of them are mentioned in the following:

First, this work focused on the analysis of process data, but we did not include learning outcomes in the analysis. Often in CSCL research, process data is used to predict learning outcomes, that is to see how learning processes moderate the effects of learning conditions and learning outcomes. Therefore, our findings are limited in that respect that we cannot say much about how different epistemic network models may or may not have led to potentially better learning outcomes. Thus, future studies should more directly address this question.

Second, we could have used more advanced measures that would be based on a traditional coding-and-counting strategy, for example, hierarchical modelling. However, our aim was to focus on analyses that are typically applied in coding-and-counting-based research (specifically ANOVAs and correlations). To triangulate our analyses, we simulated the results of an a-temporal analysis by modeling randomized data with ENA. The results led to similar outcomes as the more traditional coding-and-counting approach: in ENA, the most frequently-occurring events formed connections in the case of the randomized dataset. Thus, despite any limitations of our coding-and-counting models, we demonstrated that temporality captures significant connections between socio-cognitive events beyond those found in a frequency-based analysis.

Third, our ENA-analysis modeled events occurring in direct succession to one another. However, it might be that other window sizes would capture different problem-solving processes. Other studies have looked at how to identify the optimal scope of analysis when analyzing verbal data with ENA (Csanadi et al. 2017; Ruis et al. 2018).

Fourth, our findings were limited by the fact that we were comparing two different data collection methods (recordings of discussion versus recordings of a think-aloud protocol) as well as different problem-solving conditions (dyadic versus individual). While prior research has shown that such comparison may not compromise the results in theory (Csanadi et al. 2016; Mullins et al. 2011; Teasley 1995), the differences between ENA and a more traditional coding-and-counting approach were more pronounced for dyads than for individual problem solvers. There are examples of analyzing temporality in the context of individual learning (e.g., Bannert et al. 2014), but many studies that use temporal patterns focus on collaborative learning contexts (Jeong 2005; Kapur 2011). Thus, more studies are needed to see to what extent temporality is less of a factor in individual problem solving as recorded in think-aloud protocols in comparison to dyadic problem solving as captured through discussion.

Despite these limitations, our results suggest that traditional coding-and-counting-based approaches are limited in their ability to model temporality in verbal process data. We do not suggest that such coding-and-counting approaches should be abandoned, because counting frequencies of occurrences may reveal important information to a researcher. However, we do argue that any analysis that aims to understand how learners engage in activities and how such engagement contributes to learning needs to also use analytical approaches that account for temporal characteristics of data. Temporal analyses are not a luxury that CSCL researchers might choose to enjoy or not; rather, they are an analytical necessity for researchers interested in generating meaningful analyses of collaborative learning. Based on our research, we can say that ENA is a powerful means to perform such analyses.

Acknowledgements This research was funded in part by the following grants: the Elitenetzwerk Bayern (K-GS-2012-209); the National Science Foundation (DRL-1661036, DRL-1713110), 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

Andrist, S., Collier, W., Gleicher, M., Mutlu, B., & Shaffer, D. W. (2015). Look together: Analyzing gaze coordination with epistemic network analysis. *Frontiers in Psychology, 6*(1016).

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.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York, NY: Cambridge University Press.

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and learning, 9*(2), 161–185.

Bause, I. M., Brich, I. R., Wesslein, A. K., & Hesse, F. W. (2018). Using technological functions on a multi-touch table and their affordances to counteract biases and foster collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning, 13*(1), 7–33.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6*(3), 271–315.

Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes: Dynamic multilevel analysis. *Small Group Research, 36*(5), 600–631.

Collier, W., Ruis, A. R., & Shaffer, D. W. (2016). Local versus global connection making in discourse. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016, volume 1* (pp. 426–433). Singapore: International Society of the Learning Sciences.

Cress, U., & Hesse, W. (2013). Quantitative methods for studying small groups. In C. A. Hmelo-Silver, C. Chinn, C. Chan, & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 93–111). New York, NY: Routledge.

Csanadi, A., Kollar, I., & Fischer, F. (2015). *Internal scripts and social context as antecedents of* teacher *students' scientific reasoning.* Paper presented at the 16th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI), Limassol, Cyprus.

Csanadi, A., Kollar, I., & Fischer, F. (2016). Scientific reasoning and problem solving in a practical domain: Are two heads better than one? In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016, volume 1* (pp. 50–57). Singapore: International Society of the Learning Sciences.

Csanadi, A., Eagan, B., Shaffer, D., Kollar, I., & Fischer, F. (2017). Collaborative and individual scientific reasoning of pre-service teachers: New insights through epistemic network analysis (ENA). In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer-Supported Collaborative Learning (CSCL) 2017, volume 1* (pp. 215–222). Philadelphia: International Society of the Learning Sciences.

Dyke, G., Kumar, R., Ai, H., & Rosé, C. P. (2012). Challenging assumptions: Using sliding window visualizations to reveal time-based irregularities in CSCL processes. In J. van Aalst, K. Thompson, M. J. Jacobson, & P. Reimann (Eds.), *The future of learning: Proceedings of the 10th international conference of the learning sciences (ICLS) 2012* (Vol. 1, pp. 363–370). Sydney: International Society of the Learning Sciences.

Eagan, B., & Hamilton, E. (2018). *Epistemic Network Analysis of an International Digital Makerspace in Africa, Europe, and the US. Paper presented at the annual meeting of the American education research association.* New York: NY.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., & Hussmann, H. (2014). Pekrun, R.,…Eberle, J. *Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. Frontline Learning Research, 5*, 28–45.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*(2), 316–344.

Hmelo-Silver, C. E., Jordan, R., Liu, L., & Chernobilsky, E. (2011). Representational tools for understanding complex computer-supported collaborative learning environments. In.: Puntambekar S., Erkens G., Hmelo-silver C. (Eds). *Analyzing Interactions in CSCL. Computer-Supported Collaborative Learning, 12*, 83–106.

Hmelo-Silver, C. E., Jordan, R., & Sinha, S. (2013). Seeing to understand. Using visualizations to understand learning in technology-rich learning environments. In R. Luckin, S. Puntambekar, P. Goodyear, B. Grabowski, J. Underwood, & N. Winters (Eds.), *Handbook of Design in Educational Technology* (pp. 457–471). New York, NY: Routledge.

Jeong, A. (2005). A guide to analyzing message–response sequences and group interaction patterns in computer-mediated communication. *Distance Education, 26*(3), 367–383.

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.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1–48.

Lämsä, J., Hämäläinen, R., Koskinen, P., & Viiri, J. (2018). Visualising the temporal aspects of collaborative inquiry-based learning processes in technology-enhanced physics learning. *International Journal of Science Education, 40*(14), 1697–1717.

Marquart, C. L., Hinojosa, C., Swiecki, Z., & Shaffer, D. W. (2018). Epistemic Network Analysis (Version 0.1.0) [Software]. Available from http://app.epistemicnetwork.org

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences, 17*(1), 33–59.

Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer-Supported Collaborative Learning, 6*(3), 421–443.

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.

Reimann, P., & Yacef, K. (2013). Using process mining for understanding learning. In R. Luckin, S. Puntambekar, P. Goodyear, B. Grabowski, J. D. M. Underwood, & N. Winters (Eds.), *Handbook of design in educational technology* (pp. 472–481). New York, NY: Routledge.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69–97). Berlin: Springer.

Ruis, A.R., Rosser, A.A., Quandt-Walle, C., Nathwani, J.N., Shaffer, D.W., & Pugh, C.M. (2018). The hands and head of a surgeon: Modeling operative competency with multimodal epistemic network analysis. American Journal of Surgery.

Schwaighofer, M., Bühner, M., & Fischer, F. (2017). Executive functions in the context of complex learning: Malleable moderators? *Frontline Learning Research, 5*(1), 58–75.

Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers & Education, 46*(3), 223–234.

Shaffer, D. W. (2007). *How computer games help children learn*. New York, NY: Palgrave Macmillan.

Shaffer, D. W. (2012). Models of situated action: Computer games and the problem of transfer. In C. Steinkuehler, K. Squire, & S. Barab (Eds.), *Games learning, and society: Learning and meaning in the digital age* (pp. 403–433). Cambridge, UK: Cambridge University Press.

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

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics (pp. 175–187)* Society for Learning Analytics Research.

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

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. L., Arastoopour, G., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2016). In search of conversational grain size: Modeling semantic structure using moving stanza windows. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016, volume 1* (pp. 631–638). Singapore: International Society of the Learning Sciences.

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.

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2012). Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instructional Science, 40*(2), 297–323.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. (2006). Content analysis: What are they talking about? *Computers & Education, 46*(1), 29–48.

Sullivan, S. A., Warner-Hillard, C., Eagan, B. R., Thompson, R., Ruis, A. R., Haines, K., & Jung, H. S. (2018). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery, 163*(4), 938–943.

Suthers, D. D. (2005). Technology affordances for intersubjective learning: A thematic agenda for CSCL. In T. Koschmann, D. Suthers, & T. W. Chan (Eds.), *Computer supported collaborative learning 2005: The next 10 years* (pp. 662–671). Mahwah, NJ: Lawrence Erlbaum Associates.

Suthers, D., & Medina, R. (2011). Tracing interaction in distributed collaborative learning. In.: Puntambekar S., Erkens G., Hmelo-silver C. (Eds). *Analyzing Interactions in CSCL. Computer-Supported Collaborative Learning, 12*, 341–366.

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31*(2), 207–220.

Vogel, F., & Weinberger, A. (2018). Quantifying qualities of collaborative learning processes. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences*. New York, NY: Routledge.

Wegerif, R., & Mercer, N. (1997). Using computer-based text analysis to integrate qualitative and quantitative methods in research on collaborative learning. *Language and Education, 11*(4), 271–286.

Wise, A. F., & Chiu, M. M. (2011). Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning, 6*(3), 445–470.

JrnlID 11412_ArtID 9288_Proof# 1 - 07/11/2018

Affiliations

Andras Csanadi[1] • Brendan Eagan[2] • Ingo Kollar[3] • David Williamson Shaffer[2] • Frank Fischer[4]

[1] Bundeswehr University of Munich, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany Q1

[2] University of Wisconsin-Madison, Madison, WI, USA

[3] University of Augsburg, Augsburg, Germany

[4] Ludwig Maximilian University of Munich, Munich, Germany

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES.

Q1. Please check if the affiliations are presented correctly.
Q2. Ref. "Kollar, Fischer, & Slotta, 2007" is cited in the body but its bibliographic information is missing. Kindly provide its bibliographic information in the list.
Q3. Ref. "Shaffer & Serlin, 2004" is cited in the body but its bibliographic information is missing. Kindly provide its bibliographic information in the list.
Q4. Please check if the section headings are assigned to appropriate levels.
Q5. Figures 2-3 contains text below the minimum required font size of 6pts inside the artwork, and there is no sufficient space available for the text to be enlarged. Please provide replacement figure file.