Are We on the Same Page? Modeling Linguistic Synchrony and Math Literacy in Mathematical Discussions

Yukyeong Song
y.song1@ufl.edu
University of Florida
Gainesville, Florida, USA

Wanli Xing
wanli.xing@coe.ufl.edu
University of Florida
Gainesville, Florida, USA

Xiaoyi Tian
tianx@ufl.edu
University of Florida
Gainesville, Florida, USA

Chenglu Li
li.chenglu@ufl.edu
University of Florida
Gainesville, Florida, USA

ABSTRACT

Mathematical discussions have become a popular educational strategy to promote math literacy. While some studies have associated math literacy with linguistic factors such as verbal ability and phonological skills, no studies have examined the relationship between linguistic synchrony and math literacy. In this study, we modeled linguistic synchrony and students' math literacy from 20,776 online mathematical discussion threads between students and facilitators. We conducted Cross-Recurrence Quantification Analysis (CRQA) to calculate linguistic synchrony within each thread. The statistical testing result comparing CRQA indices between high and low math literacy groups shows that students with high math literacy have a significantly higher Recurrence Rate (RR), Number of Recurrence Lines (NRLINE), and the average Length of lines (L), but lower Determinism (DET) and normalized Entropy (rENTR). This result implies that students with high math literacy are more likely to share common words with facilitators, but they would paraphrase them. On the other hand, students with low math literacy tend to repeat the exact same phrases from the facilitators. The findings provide a better understanding of mathematical discussions and can potentially guide teachers in promoting effective mathematical discussions.

CCS CONCEPTS

• **Computing methodologies** → **Discourse, dialogue and pragmatics**; • **Applied computing** → *Education*.

KEYWORDS

mathematical discussion, math literacy, linguistic synchrony, cross-recurrence quantification analysis

ACM Reference Format:
Yukyeong Song, Wanli Xing, Xiaoyi Tian, and Chenglu Li. 2023. Are We on the Same Page? Modeling Linguistic Synchrony and Math Literacy in Mathematical Discussions. In *LAK23: 13th International Learning Analytics and Knowledge Conference (LAK 2023), March 13–17, 2023, Arlington, TX, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3576050.3576082

LAK 2023, March 13–17, 2023, Arlington, TX, USA

ACM ISBN 978-1-4503-9865-7/23/03…$15.00
https://doi.org/10.1145/3576050.3576082

1 INTRODUCTION & BACKGROUND

Math literacy is an essential ability to apply math concepts and knowledge to solve real-world problems and communicate solutions with others [30]. One effective strategy to promote math literacy is through mathematical discussions [19, 35]. Engaging students in mathematical discussions allows them to explain their thinking processes, exchange possible solutions, and justify their reasoning [4].

Prior research has suggested a strong link between students' math literacy and their linguistic skills [31, 40] because math literacy is not just about knowing numbers and symbols but also about understanding the meaning surrounding them and communicating with others using them [10]. A large body of research has focused on the discourse analysis of students' dialogues in mathematical discussions, such as analyzing the complexity of the language in online question-and-answer math forums [9], modeling students' math performance with text cohesion, lexical sophistication, and sentiment [11], and understanding cohesive cues from online math discussion boards [2].

Despite the aforementioned studies associating students' math literacy with a series of linguistic factors, there is no research analyzing the relationship between linguistic synchrony and math literacy. Linguistic synchrony, also referred to as interactive alignment [15] or shared language use [3], is the degree to which two or more interlocutors reciprocally share linguistic properties, such as lexical choices in dialogue [6, 7, 32]. Linguistic synchrony is known to be correlated with better collaborative task performance [15, 28], common knowledge building [3], and learning gains [36, 37] because it represents the language convergence on a shared understanding [14].

In this study, we aimed to model linguistic synchrony along with students' math literacy in mathematical discussions. Motivated by prior works that interlocutors tend to adapt to each other's linguistic attributes [15, 36], we hypothesized that students with higher math literacy would show higher linguistic synchrony with others in mathematical discussions. We used 20,776 discussion threads from Math Nation[1], a middle and high school online math

[1] https://www.mathnation.com/

learning platform supporting asynchronous discussions. First, we categorized students' math literacy into low and high based on the context of each discussion thread. We manually annotated 743 threads and then trained transformer-based models to automatically classify the remaining threads. Next, we measured the linguistic synchrony between students and facilitators within each discussion thread by carrying out Cross-Recurrence Quantification Analysis (CRQA) and generating recurrence plots. This study specifically addressed the following two research questions (RQs):

- RQ1. What are the characteristics of linguistic synchrony in mathematical discussions?
- RQ2. What is the relationship between linguistic synchrony and math literacy?

To answer RQ1, we reported the descriptive statistics of the five indices from CRQA analysis, including Recurrence Rate (RR), Determinism (DET), Number of Recurrence Lines (NRLINE), the average length of line structure (L), and normalized Entropy (rENTR). To answer RQ2, we applied the Mann-Whitney U test to compare CRQA indices between the high math literacy students and the low math literacy group. Also, to provide a qualitative assessment, we included two representative recurrence plots of high and low-literacy student groups. To the best of our knowledge, this study is the first attempt to investigate the relationship between linguistic synchrony and students' math literacy in mathematical discussions. This study contributes to the current line of research to use discourse analysis to understand the dynamics of group discussions. Our discourse analysis of linguistic synchrony holds the potential to guide teachers in promoting effective mathematical discussions.

2 METHODS

2.1 Data and Context

We used the discussion data among students and facilitators from the question-and-answer board of Math Nation, an online math learning platform for middle and high school students. We used 20,776 threads that had more than three replies from the student who started the thread out of 316,352 threads from 2015-06-01 to 2022-03-01. In each thread, a student posts questions related to math problems, and facilitators scaffold them to solve the problems by replying to the posts. Note that there is only one "student" per thread, which is the person who starts a thread; all the other users who replied under the thread are considered as "facilitators".

2.2 Automatic Assessment of Students' Math Literacy

We assessed the students' math literacy based on their conversations in the discussions. To do this, we developed the initial coding scheme in a top-down way based on previous studies about math literacy, including three constructs: *Knowledge and Content, Competency and Skills*, and *Application in the Real World* [25, 27, 29, 33]. The math literacy level of each construct was categorized into high (1) and low (0). The description of the constructs and examples used as a rubric is in Table 1.

To assess students' math literacy in each thread, we first manually coded a small dataset of 743 discussion threads. The following is the procedure of manual coding. First, two doctoral students reviewed the math literacy coding scheme, coded 50 threads independently, and calculated the initial inter-rater reliability (IRR) using Cohen's Kappa. The IRR for the first 50 posts turned out to be 0.75 for *Knowledge and Content*, 0.80 for *Competency and Skills*, and 0.54 for *Application in the Real World.* After reviewing the disagreement, they coded 200 more threads independently and calculated the IRR again. The IRR for the next 200 threads were 0.85 for *Knowledge and Content*, 0.87 for *Competency and Skills*, and 0.93 for *Application in the Real World.* Once they achieved the desirable IRR scores for each construct, one annotator coded 493 more threads, forming 743 hand-labeled threads to be used as a training dataset for automatic text classification.

After the manual annotation, we trained automatic text classification models using both traditional machine learning and deep neural network models. We modeled the three constructs of math literacy as separate binary prediction problems (0 – low vs. 1 - high) using Transformer models, benchmarked with Random Forest (RF) and Support Vector Machine (SVM). The Transformer is a deep neural network architecture that allows researchers to utilize transfer learning with models pre-trained by large-scale text data for downstream tasks [21]. We selected RF as the benchmark model because studies using traditional machine learning consistently showed that it could achieve better or comparable performance than others [23]. We chose SVM due to its advantage in handling high-dimensional data (e.g., vectors of texts) and could well address over-fitting through regularization [8].

We trained a set of models separately for each construct. We first split the data into training (70%, n = 520) and evaluation (30%, n = 223) sets. We examined these two models using raw texts and NLP-enhanced data (stop-word removal, lemmatization, tf-idf). We utilized 5-fold cross-validation with grid search to find the most robust and accurate models of RF and SVM. For Transformer models, we examined Bidirectional Encoder Representations from Transformer (BERT) [12] and LongFormer [5], which have been shown to be robust text classifiers in educational settings [38]. We further examined LongFormer because it is specifically designed to handle lengthy documents, which can fit well in our context as we tried to classify long discussions with multiple turns. We trained these two Transformer models with raw texts directly as they were pre-trained with human-readable documents.

For the evaluation, we used accuracy, F1 score, and Area Under the receiver operating characteristic Curve (AUC) [18]. Table 2 shows the evaluation results of models where Transformer models consistently reached the best performance. Notably, RF and SVM showed the same and poor performance regarding the construct of *Application in the Real World.* This is due to the models' incapability to capture the relationship between data and labels, whereas Transformer models such as BERT could conduct the correct inferences more effectively. After evaluation, we selected the best-performing models for each construct to predict the rest of the dataset (n = 20,003).

2.3 Measurement of Linguistic Synchrony in Mathematical Discussions

To calculate linguistic synchrony in Mathematical discussions, we applied Cross-Recurrence Quantification Analysis (CRQA). This

Table 1: Math Literacy Evaluation Rubric and Examples

Construct	High (1)		Low(0)	
	Descriptions	Examples	Descriptions	Examples
Knowledge and Content	Students show an understanding of math concepts and give correct answers.	"Box plots use an IQR while dot plots use standard deviation." "Triangles equal 180."	Students show no or limited under -standing of math concepts and give wrong answers.	"What is the quadratic formula? I don't remember." "What does an equivalent equation mean?"
Competency and Skills	Students show the ability to communicate using math terms and formulas.	"5x+14=3y-10 will turn into 5x+24=3y." "So, 10 to the 4 power is 10000."	Students show no or limited ability to communicate about math.	"I am confused because the book from school says the equivalent equation and we learned using different words"
Application in the Real World	Students connect math concepts to interpret the real-world phenomena.	"I got stuck because that went into a negative num -ber, which means they are paying you to buy their potato salad."	Students do not connect math concepts to the real-world.	N/A*

* We did not include examples of *Application in the Real World (low)* because any conversation that lacks evidence of this construct was coded as low.

Table 2: Math Literacy Automatic Classification Results. Bold texts indicate the best-performed model of a construct.

Constructs	Knowledge and Content			Competency and Skills			Application in the Real World		
	Accuracy	F1	AUC	Accuracy	F1	AUC	Accuracy	F1	AUC
BERT	0.91	0.91	0.92	0.78	0.70	0.69	**0.87**	**0.75**	**0.74**
LongFormer	**0.93**	**0.93**	**0.93**	**0.85**	**0.79**	**0.78**	0.82	0.45	0.50
RF + raw text	0.89	0.89	0.89	0.75	0.53	0.55	0.83	0.50	0.52
RF + NLP	0.91	0.91	0.91	0.76	0.57	0.58	0.82	0.47	0.51
SVM + raw text	0.90	0.90	0.90	0.73	0.45	0.52	0.82	0.56	0.56
SVM + NLP	0.91	0.91	0.91	0.74	0.47	0.52	0.82	0.45	0.50

subsection describes the text preprocessing steps and the CRQA indices.

2.3.1 *Preprocessing*.
Before computing recurrence between a student and facilitators in each thread, we applied text preprocessing to the corpus. First, there might be multiple facilitators who replied to one post. These multiple consecutive posts from facilitators were joined as if they were a single post, considering that they all replied directly to the student (e.g., a thread containing five total posts {S, F^1, F^2, S, F^1}was combined as {S, F, S, F}; S denotes posts from the student, F^n denotes posts from facilitators). Then, because the discussions often involve different forms of math formulas or equations (e.g., y = 2 + 3x), for the scope of this analysis, we treated these math formulas universally across the corpus and converted them to a unique word. Finally, we processed the corpus using standard text preprocessing techniques, including lowercasing, expanding contractions, removing punctuation and stopwords, stemming, and tokenization.

2.3.2 *Cross-Recurrence Quantification Analysis (CRQA)*.
CRQA is an extended family of Recurrence Quantification Analysis (RQA), a data analysis methodology that quantifies the repetition of data points within a time-series or numerical data [26]. CRQA investigates the repetition of two sequential data [20, 41]. In this context, the two sequential data sources are students' and facilitators' posts in the same thread.

The major outputs of CRQA are the recurrence plot and the CRQA indices. The cross-recurrence plot requires an equal length of posts from both parties (students and facilitators) to draw a square-shaped plot. However, in real-world discussions, the aggregated lengths of the posts from two parties may not be equal. To equalize the post length to produce the cross-recurrence plots we added artificial tokens to the shorter post between each pair of post exchanges (one student's post and one facilitator's post). We equalized the post length for pairs of posts because discussion posts are sequential, which means one post can only synchronize with the previous posts, not the following ones. The equalized word tokens were later converted to their corresponding numeric values. We adopted the *crqa* package in *R* [34] for our analysis.

CRQA indices include Recurrence Rate (RR), Determinism (DET), Number of Recurrence Lines (NRLINE), the average length of line structures (L), and normalized Entropy (rENTR). RR is the density of recurrence points in a recurrence plot [41]. In our context, RR is calculated as the number of repeated words divided by the product of the student's and facilitators' conversation lengths. Below is an example:

> Student: *What do similar polygons have the same number of?*

Facilitator: *If they are similar, they should have the same number of sides and angles.*

After pre-processing and length-equalization, the word tokens from the student are: ['similar,' 'polygon,' 'number,' '$'] [2], and from the facilitator are: ['similar,' 'number,' 'side,' 'angle']. Two words, 'similar' and 'number' are repeated among all possible word combinations (4x4) [3]. Thus, the RR is 12.5%.

DET is the percentage of recurrent points that makes diagonal lines [13]. In our context, DET is the percentage of the times that a student and facilitators use the same phrases out of all the recurrence points. NRLINE is the total number of lines in the recurrence plot [1]. High NRLINE implies more sequences of words overlapping in the conversation. L is the average length of the sequence of recurrence points [41]. High L indicates they share longer sequences of words. Lastly, ENTR is determined by the diversity of the diagonal line lengths in the recurrence plot [24]. High ENTR implies that the overlapping words or phrases used by the student and facilitators are of more variety of lengths. We used normalized Entropy (rENTR), normalized by the number of lines observed in the plot to make it easier to compare across different conditions [39].

2.4 Examining the Relationship between Math Literacy and Linguistic Synchrony

To examine the relationship between math literacy levels and linguistic synchrony between students and facilitators, we classified all threads into high literacy and low literacy (overall and each of the three constructs) and calculated CRQA indices of each thread. A student's *overall* literacy score was summed across all three constructs in which a summed score of 0 and 1 were classified as low *overall* literacy, and 2 and 3 were classified as high *overall* literacy. We compared the difference of CRQA indices between high and low math literacy students using the Mann-Whitney U test. The Mann-Whitney U test does not assume the homogeneity of variance, which makes it suitable for the data used in this paper. To address the potential increase in false discoveries for multiple comparisons, we applied the Benjamini–Hochberg procedure [17] to adjust the statistical significance threshold.

3 RESULTS

3.1 The Characteristics of Linguistic Synchrony in Mathematical Discussions

To characterize linguistic synchrony, we reported five commonly adopted CRQA indices: RR, DET, NRLINE, L, and rENTR. Table 3 shows the general characteristics of CRQA indices in our corpus. As shown in Table 3, the *conversation length* means the total number of words from both students' and facilitators' posts in a thread after text processing (e.g., removing stop words). The *conversation length* varies drastically and is right-skewed. RR is also right-skewed with a mean of 0.73 and mostly (70%) below 1. It is notable that DET and rENTR contain smaller sample sizes (N) compared to the other three indices because there are many missing values. This is because DET is calculated by dividing the number of recurrence points that construct diagonal lines by RR [41], and rENTR is calculated with Shannon entropy [22] based on the probability that a diagonal line has an exact length. Therefore, for threads that demonstrate 0 RR values or no diagonal lines, the results of DET and rENTR are null. NRLINE and L are related to the number of diagonal lines and their length. Since many conversations did not include diagonal lines (no sequence of overlapping words), both data are right-skewed.

3.2 The Relationship between Linguistic Synchrony and Students' Math Literacy

Table 4 shows the results of the Mann-Whitney U test comparing the CRQA indices between high and low math literacy groups.

The Mann-Whitney U test results revealed that the high *overall* literacy group had significantly higher RR, NRLINE, and L (positive Z scores) and significantly lower DET and rENTR compared to the low *overall* literacy group (negative Z scores). We found the same pattern consistently to all three constructs of math literacy.

Fig. 1 presents examples of recurrence plots along with a part of excerpts of two discussions assessed to be high (Fig. 1(a)) and low (Fig. 1(b)) math literacy respectively. The dots in both figures represent recurrence points in which students and facilitators share the same words in their posts. The denser points in Fig. 1(a) indicate a higher Recurrence Rate (RR). The number of marked points constructing diagonal lines (NRLILNE, example lines circled in red) in Fig. 1(a) is greater than that of Fig. 1(b). Both figures demonstrate similar levels of deterministic patterns, although DET indices reveal that Fig. 1(b) shows a bigger DET value than Fig. 1(a). In addition, both figures include about 2 to 3 points constructing each line (L), and have a similar variance in the number of points (rENTR).

4 DISCUSSION

This study investigated linguistic synchrony in mathematical discussions and its relationship with students' math literacy. Next, we will discuss the results with regard to the two research questions.

4.1 RQ1. What are the Characteristics of Linguistic Synchrony in Mathematical Discussions?

In the CRQA indices of our mathematical discussion data, the average RR was 0.73, meaning that in average conversations, students' words would include less than 1% overlap with the facilitators'. Conversations with an RR value of higher than 10% were relatively short, with less than 10 words (after text processing) in the discussion thread (*Mean conversation length* = 5). Besides those cases, most of the conversations in our data showed low recurrence rates. The average DET was 36.39 with a high variance (*std* = 26.59). This was also a relatively low value compared to the previous study that presented DET ranging from 58 to 63 in the collaborative group projects conversations [16]. This indicated that speakers in mathematical discussions would be less likely to duplicate each other's sequence of words. However, the high variance also suggested that this might be highly dependent on the characteristics of each thread.

[2]The fourth token is an artificial token for equalizing the length with the facilitator's utterance

[3]Other words such as 'have,' 'the,' and 'same' are also repeated in the original conversation, but these words were common as English stopwords and were removed during pre-processing.

Table 3: Descriptive Statistics of Overall CRQA Indices

	N	Mean	Std	Min	25%	50%	75%	Max
Conversation length	20,671	59.77	74.56	1	17	35	71	1266
RR (%)	20,671	0.73	1.33	0	0.11	0.38	0.82	50
DET	9,005	36.39	26.59	0.74	15.38	28.57	50	100
NRLINE	20,671	2.71	10.47	0	0	0	2	434
L	20,671	0.96	1.16	0	0	0	2	25
rENTR	2,514	0.74	0.23	0.10	0.55	0.81	0.92	1

N: sample size, RR: recurrence rate, DET: determinism, NRLINE: number of recurrence Line, L: average length of line structures, rENTR: normalized entropy.

Table 4: Comparison of CRQA indices between the high and low literacy group

Math Literacy	CRQA Indices	High Literacy				Low Literacy				Mann-Whitney *U*	
		N	Mean	Median	Std	N	Mean	Median	Std	*U*	*Z*
Overall	**RR**	2,435	0.64	0.51	0.54	18,236	0.72	0.35	1.40	26,771,876 **	16.52
	DET	1,964	20.10	16.61	14.12	7,041	40.94	33.33	27.46	1,382,069 **	-54.30
	NRLINE	2,435	10.04	3.00	21.58	18,236	1.72	0.00	7.34	34,570,086 **	44.72
	L	2,435	1.78	2.00	0.93	18,236	0.85	0.00	1.14	32,634,208 **	37.72
	rENTR	920	0.65	0.65	0.24	1,594	0.79	0.86	0.22	165,366.0 **	-32.39
KC	**RR**	6,324	0.55	0.40	0.59	14,347	0.79	0.36	1.54	47,957,554 **	6.56
	DET	4,121	26.30	20.51	19.50	4,884	44.90	40.00	28.71	6,068,849 **	-32.50
	NRLINE	6,324	5.25	1.00	15.35	14,347	1.57	0.00	7.06	62,333,429 **	42.92
	L	6,324	1.43	2.00	1.09	14,347	0.75	0.00	1.13	60,399,063.5 **	38.03
	rENTR	1,469	0.71	0.72	0.23	1,045	0.78	0.84	0.22	620,371.0 **	-8.21
CS	**RR**	2,928	0.72	0.53	0.77	17,743	0.71	0.34	1.40	32,353,726 **	21.32
	DET	2,257	19.63	15.79	14.44	6,748	42.00	35.13	27.36	3,542,511 **	-38.09
	NRLINE	2,928	10.63	3.00	22.72	17,743	0.00	0.00	5.52	39,665,783.5 **	45.76
	L	2,928	1.70	2.00	0.99	17,743	0.83	0.00	1.14	37,483,038 **	38.47
	rENTR	1,081	0.26	0.00	0.32	1,433	0.80	0.92	0.20	470,349 **	-16.88
APP	**RR**	1,838	0.70	0.41	1.24	18,833	0.71	0.37	1.33	18,089,757.5 *	3.20
	DET	984	33.28	25.00	24.64	8,021	36.77	28.57	26.80	3,674,019.0 **	-3.54
	NRLINE	1,838	4.55	1.00	12.75	18,833	2.51	0.00	10.20	19,764,734.0 **	10.06
	L	1,838	1.22	2.00	1.40	18,833	0.93	0.00	1.13	19,579,518.5 **	9.30
	rENTR	369	0.70	0.72	0.23	2,145	0.74	0.81	0.23	353,862.0 *	-3.25

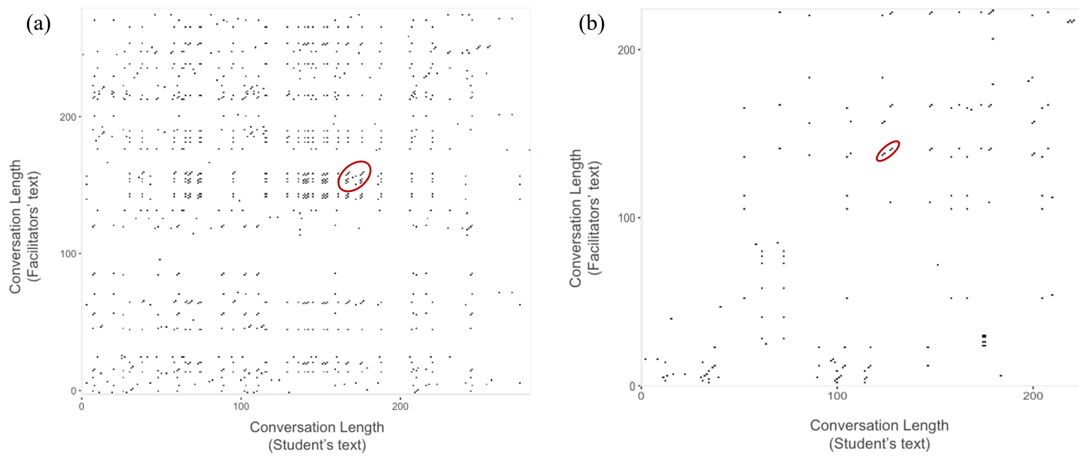
Note 1. * p <.01, ** p <.001 (after the Benjamini–Hochberg correction).
Note 2. KC: Knowledge and Content, CS: Competency and Skills, APP: Application in the Real World

4.2 RQ2. What is the Relationship between Linguistic Synchrony and Math Literacy?

The Mann-Whitney *U* test result revealed that high math literacy students show higher RR, NRLINE, and L while showing lower DET and rENTR. We found a consistent pattern in all math literacy constructs. The pattern suggested that these five CRQA indices could be useful to model linguistic synchrony and further be divided into the following three sub-groups to represent different sub-concepts of linguistic synchrony: (1) RR, NRLINE, and L could mean *linguistic concurrence*, which indicates the degree to which two interlocutors use the same words or phrases. (2) DET could correspond to *predictability*, which means the percentage of the overlapping sequence of words out of all the recurrence points. (3) rENTR indicates *complexity*, which refers to how complex the recurring patterns are (the variance of the recurrence lines).

First, the high math literacy student group showed higher 1) RR, 2) NRLINE, and 3) L. These three indices are closely related to the *linguistic concurrence*. Higher RR means that the student and the facilitators use the same words more often in the conversation. Higher NRLINE implies that they share more sequences of words in the discussion, while higher L indicates that they share longer lines of those words. Thus, we can conceptualize higher values of these three indices to suggest the *higher linguistic concurrence*. This could be because students with high math literacy might be more likely to use the standard math terms to communicate with others than low literacy students.

On the other hand, we found the high literacy group showed lower 1) DET and 2) rENTR. These two indices are related to *predictability* and *complexity* of the conversation. Conversations with low determinism are less predictable because it includes a less regular frequency of the same word sequences shared among speakers.



...
F: Yup! And when you multiply their ages you get 3192, then you do c(c+1) and set it equal to 3192.
S: Ohhhh interesting.
F: Do you understand now?
S: Yes, you would multiply because their ages have a product of 3192, right?
...

...
S: I THINK IT IS C ...
F: Yes, it is C ...
F: Because he is consuming 3 a minute there must be a multiple of three as well as a subtraction.
S: Oh lol. So, how's your freshman year going?
S: and THANKS GUYS.
...

Figure 1: Recurrence plots for conversations of students (S) with facilitators (F) assessed to be of high (a) and low (b) math literacy.
CRQA indices for (a): RR= 1.22, DET= 27.13, NRLINE= 121, L= 2.16, rENTR= 0.43.
CRQA indices for (b): RR= 0.32, DET= 31.01, NRLINE= 23, L= 2.13, rENTR= 0.32.

That is, students with higher math literacy are less likely to keep repeating the exact same word sequences from the facilitators. This result aligns with the previous research about students' writing that identified that high-quality writing had a lower DET than average-quality writing works [24], along with another study that revealed DET had a negative correlation with students' holistic essay scores [1].

rENTR could represent *complexity* of the conversation. Conversations with low rENTR are less complex and rather more stable because the two interlocutors share more uniform lengths of diagonal lines. That is, students with high math literacy would use more uniform numbers of overlapping words rather than speaking varied lengths of words that overlap with the facilitators. For example, the overall high literacy group had an average line length (L) of 1.78, with a standard deviation of 0.93 and a maximum of 8.75. On the other hand, the overall low literacy group showed an average line length (L) of 0.85, with a bigger standard deviation of 1.14 and a maximum of 25. These indices imply situations like a student with high math literacy would only repeat certain terms that are composed of a few words (e.g., 1.78 words), while another student with low literacy would either repeat the whole sentence (e.g., 25 words) or neglect using necessary terms in the discussion. This result also aligns with a previous study in which rENTR was negatively correlated with the students' writing quality [24].

Previous research also highlighted the importance of investigating the recurrence plots in context [24]. By examining the actual texts from the conversation of a high *overall* literacy group (Figure 1 (a)), we found that students and facilitators constantly talked about the math problems without going off-task and they shared some common math languages, such as "multiply" or "equation". They also used many formulas to communicate. In addition, the student tended not to repeat the same phrases or sentences from facilitators; instead, they paraphrased them or switched symbols to words (e.g., "-" to "subtract"). In contrast, an inspection of actual text from the conversation of a low *overall* literacy group (Fig. 1 (b)) revealed that they often deviated from the math topic to random topics, such as school life. Also, the student wrote out the options in the multiple choice problem directly (e.g., "I think it is C") rather than discussing their reasoning on how to solve the problem. This would have led this thread to show low RR and NRLINE.

5 CONCLUSION AND FUTURE DIRECTIONS

The goals of this paper were to investigate the linguistic synchrony of mathematical discussions through CRQA indices and to discover if linguistic synchrony is related to students' math literacy levels. The CRQA analysis revealed that the CRQA indices could indicate linguistic synchrony in three aspects: *linguistic concurrence*, *predictability*, and *complexity*. In addition, statistical testing of CRQA indices between high and low math literacy groups suggested that students with high math literacy showed higher *linguistic concurrence* (indicated by higher RR, NRLINE, and L), lower *predictability* (indicated by lower DET) and lower *complexity* (indicated by rENTR) with facilitators in math discussions. This result implies that students with high math literacy are more likely to use the overlapping words with facilitators, but they paraphrase or elaborate on these overlapping words, while students with low math literacy tend to repeat the exact same phrases or sentences from facilitators.

This study highlights several future directions. First, it is important to design and deploy a monitoring dashboard reflecting linguistic synchrony to guide teachers in understanding the dynamics of ongoing group discussions. For example, teachers can use the recurrence plots and CRQA indices to gauge how students share the standard language needed for effective communication. Second, more research needs to be done to investigate teaching strategies to promote linguistic synchrony and effective discussions. Last, to provide teachers with more insights into the discussions, further analysis is needed to examine how linguistic synchrony varies depending on whether students engage in math-related or off-task conversations such as social interactions.

REFERENCES

[1] Laura Kristen Allen, Aaron D Likens, and Danielle S McNamara. 2017. Recurrence quantification analysis: A technique for the dynamical analysis of student writing. In *The Thirtieth International Flairs Conference*.

[2] Michelle Banawan, Jinnie Shin, Renu Balyan, Walter L Leite, and Danielle S McNamara. 2022. Math Discourse Linguistic Components (Cohesive Cues within a Math Discussion Board Discourse). In *Proceedings of the Ninth ACM Conference on Learning@ Scale*. 389–394.

[3] A Burcu Bayram and Vivian Ta. 2020. Measuring common knowledge: Latent semantic analysis, linguistic synchrony, and common knowledge in international relations. *International Relations* 34, 2 (2020), 180–203.

[4] Clare V Bell and Stephen J Pape. 2012. Scaffolding students' opportunities to learn mathematics through social interactions. *Mathematics Education Research Journal* 24, 4 (2012), 423–445.

[5] Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150* (2020).

[6] Holly P Branigan, Martin J Pickering, and Alexandra A Cleland. 2000. Syntactic co-ordination in dialogue. *Cognition* 75, 2 (2000), B13–B25.

[7] Susan E Brennan and Joy E Hanna. 2009. Partner-specific adaptation in dialog. *Topics in Cognitive Science* 1, 2 (2009), 274–291.

[8] Kwok Tai Chui, Dennis Chun Lok Fung, Miltiadis D Lytras, and Tin Miu Lam. 2020. Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Computers in Human Behavior* 107 (2020), 105584.

[9] Scott Crossley, Tiffany Barnes, Collin Lynch, and Danielle S McNamara. 2017. Linking Language to Math Success in an On-Line Course. *International Educational Data Mining Society* (2017).

[10] Scott Crossley, Ran Liu, and Danielle McNamara. 2017. Predicting math performance using natural language processing tools. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. 339–347.

[11] Scott Crossley, Jaclyn Ocumpaugh, Matthew Labrum, Franklin Bradfield, Mihai Dascalu, and Ryan S Baker. 2018. Modeling Math Identity and Math Success through Sentiment Analysis and Linguistic Features. *International Educational Data Mining Society* (2018).

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[13] Muhterem Dindar, Iman Alikhani, Jonna Malmberg, Sanna Järvelä, and Tapio Seppänen. 2019. Examining shared monitoring in collaborative learning: A case of a recurrence quantification analysis approach. *Computers in Human Behavior* 100 (2019), 335–344.

[14] Riccardo Fusaroli, Bahador Bahrami, Karsten Olsen, Andreas Roepstorff, Geraint Rees, Chris Frith, and Kristian Tylén. 2012. Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological science* 23, 8 (2012), 931–939.

[15] Riccardo Fusaroli and Kristian Tylén. 2016. Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive science* 40, 1 (2016), 145–171.

[16] Jamie C Gorman, Nancy J Cooke, Polemnia G Amazeen, and Shannon Fouse. 2012. Measuring patterns in team interaction sequences using a discrete recurrence approach. *Human Factors* 54, 4 (2012), 503–517.

[17] Yosef Hochberg and Yoav Benjamini. 1990. More powerful procedures for multiple significance testing. *Statistics in medicine* 9, 7 (1990), 811–818.

[18] Jin Huang and Charles X Ling. 2005. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering* 17, 3 (2005), 299–310.

[19] Eva Jablonka. 2003. Mathematical literacy. In *Second international handbook of mathematics education*. Springer, 75–102.

[20] Andrew P Knight, Deanna M Kennedy, and Sara A McComb. 2016. Using recurrence analysis to examine group dynamics. *Group Dynamics: Theory, Research, and Practice* 20, 3 (2016), 223.

[21] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942* (2019).

[22] Christophe Letellier. 2006. Estimating the Shannon entropy: Recurrence plots versus symbolic dynamics. *Physical review letters* 96, 25 (2006), 254102.

[23] Jiajun Liang, Jian Yang, Yongji Wu, Chao Li, and Li Zheng. 2016. Big data application in education: dropout prediction in edx MOOCs. In *2016 IEEE second international conference on multimedia big data (BigMM)*. IEEE, 440–443.

[24] Aaron D Likens, Kathryn S McCarthy, Laura K Allen, and Danielle S McNamara. 2018. Recurrence quantification analysis as a method for studying text comprehension dynamics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. 111–120.

[25] France Masilo Machaba. 2017. Pedagogical demands in mathematics and mathematical literacy: A case of mathematics and mathematical literacy teachers and facilitators. *Eurasia Journal of Mathematics, Science and Technology Education* 14, 1 (2017), 95–108.

[26] Norbert Marwan, M Carmen Romano, Marco Thiel, and Jürgen Kurths. 2007. Recurrence plots for the analysis of complex systems. *Physics reports* 438, 5-6 (2007), 237–329.

[27] Hannele Niemi, NİU Shuanghong, Marianna Vivitsou, and Lİ Baoping. 2018. Digital storytelling for twenty-first-century competencies with math literacy and student engagement in China and Finland. *Contemporary Educational Technology* 9, 4 (2018), 331–353.

[28] Utku Norman, Tanvi Dinkar, Barbara Bruno, and Chloé Clavel. 2022. Studying Alignment in a Collaborative Learning Activity via Automatic Methods: The Link Between What We Say and Do. *Dialogue & Discourse* 13, 2 (2022), 1–48.

[29] Department of Basic Education. 2011. *Curriculum and Assessment Policy Statement GRADES 10-12*. Department of Basic Education.

[30] Bobby Ojose. 2011. Mathematics literacy: Are we able to put the mathematics we learn into everyday use. *Journal of mathematics education* 4, 1 (2011), 89–100.

[31] Peng Peng, Xin Lin, Zehra Emine Ünal, Kejin Lee, Jessica Namkung, Jason Chow, and Adam Sales. 2020. Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin* 146, 7 (2020), 595.

[32] Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27, 2 (2004), 169–190.

[33] PISA. 2022. Overview. https://pisa2022-maths.oecd.org/ca/index.html

[34] R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[35] John Richards. 1991. Mathematical discussions. In *Radical constructivism in mathematics education*. Springer, 13–51.

[36] Tanmay Sinha and Justine Cassell. 2015. We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In *Proceedings of the 1st workshop on modeling INTERPERsonal SynchrONy And influence*. 13–20.

[37] Caitlin Snyder, Nicole Hutchins, Gautam Biswas, Mona Emara, Shuchi Grover, and Luke Conlin. 2019. Analyzing students' synergistic learning processes in physics and CT by collaborative discourse analysis. In *Proceedings of the 13th International Conference on Computer-Supported Collaborative Learning (CSCL)*.

[38] Shannon H Sung, Chenglu Li, Guanhua Chen, Xudong Huang, Charles Xie, Joyce Massicotte, and Ji Shen. 2021. How does augmented observation facilitate multimodal representational thinking? Applying deep learning to decode complex student construct. *Journal of Science Education and Technology* 30, 2 (2021), 210–226.

[39] Sofia Tancredi, Rotem Abdu, Dor Abrahamson, and Ramesh Balasubramaniam. 2021. Modeling nonlinear dynamics of fluency development in an embodied-design mathematics learning environment with Recurrence Quantification Analysis. *International Journal of Child-Computer Interaction* 29 (2021), 100297.

[40] Rose K Vukovic and Nonie K Lesaux. 2013. The relationship between linguistic skills and arithmetic knowledge. *Learning and Individual Differences* 23 (2013), 87–91.

[41] Sebastian Wallot and Giuseppe Leonardi. 2018. Analyzing multivariate dynamics using cross-recurrence quantification analysis (crqa), diagonal-cross-recurrence profiles (dcrp), and multidimensional recurrence quantification analysis (mdrqa)–a tutorial in r. *Frontiers in psychology* 9 (2018), 2232.