Education and Information Technologies (2025) 30:9215–9240
https://doi.org/10.1007/s10639-024-13195-9

Investigating dialogic interaction in K12 online one-on-one mathematics tutoring using AI and sequence mining techniques

Deliang Wang[1] · **Dapeng Shan**[2] · **Ran Ju**[3] · **Ben Kao**[2] · **Chenwei Zhang**[1] · **Gaowei Chen**[1]

Received: 27 August 2024 / Accepted: 25 November 2024 / Published online: 29 November 2024

Abstract

Online one-on-one tutoring serves as a supplementary approach to traditional classroom instruction. It has been shown to enhance personalized learning and academic performance. However, the dynamics of dialogic interactions within this educational setting are not fully understood. Thus, we present a computational analysis of dialogic interactions in online one-on-one mathematics tutoring. Specifically, we devised a coding scheme tailored to online tutoring sessions and leveraged advanced artificial intelligence techniques to construct an automated model for annotating dialog acts. We then investigated the basic characteristics and interaction patterns in a dataset encompassing online one-on-one tutoring dialogs within K-12 mathematics education and obtained insightful findings. First, tutors were found to often apply both didactic and other effective teaching strategies. Second, off-task chatting accounted for a significant proportion of tutoring sessions. Third, high school students exhibited greater engagement and cognitive abilities than primary and middle school students through their more active participation and superior reasoning skills. Primary school students, despite their less active participation, responded positively when engaged by tutors. The findings highlight the importance of optimizing strategies applied by tutors and students to create a more dynamic and effective learning environment and provide valuable insights into the nature of online one-on-one tutoring.

Keywords Dialogic interaction · Online one-on-one tutoring · Interaction patterns · Artificial intelligence · Lag sequential analysis

---

Deliang Wang and Dapeng Shan contributed equally to this work.

Extended author information available on the last page of the article

1 Introduction

Language is a key medium for transmitting information between individuals and is a significant method of facilitating teaching and learning (Howe et al., 2019). A large body of research has advocated for the careful use of language as an instructional tool within educational contexts and developed a range of pedagogical approaches related to dialog (Kim & Wilkinson, 2019). Empirical studies have consistently identified the pivotal role of dialog in learning, such as developing learners' communication skills and promoting critical thinking and reasoning abilities in language- and science-related disciplines (Hu et al., 2023; Tao & Chen, 2023a; Teo, 2019). Researchers have investigated teacher-student and student-student interaction patterns to identify the key characteristics of dialog in supporting teaching and learning (e.g., Jacobs & Renandya, 2021; Ong et al., 2023). Compared with other pedagogical methods, dialogic teaching has been found to create more opportunities for students to express themselves and elaborate on their ideas, which then further improves their language and reasoning skills (Al-Adeimi & O'Connor, 2021; Skovsmose, 2020).

In the digital age, dialogic practices are increasingly being adapted for online environments, thus presenting new opportunities and challenges for educators and learners. A key development is the rise of online one-on-one tutoring, which has been shown to be more effective than traditional classroom lecturing (Booth et al., 2024; Nickow et al., 2024) because it provides more opportunities for interaction, which enhances learning outcomes. This also provides valuable data for researchers investigating the teaching and learning process. Unlike traditional face-to-face teaching, online one-on-one tutoring can overcome the limitations of space and time. Students can connect with a wider range of tutors, thereby promoting a more personalized and flexible learning experience. This flexibility is particularly crucial in the wake of the COVID-19 lockdown, as many students have turned to online tutoring as a supplement to formal classroom learning (Yusuf, 2021). While the benefits of online tutoring have been established (Gortazar et al., 2024), there is still a lack of understanding of the specific dynamics of student-teacher interactions in the context of online tutoring, particularly in online synchronous mathematical discussions (Wallach & Kontorovich, 2024).

Thus, we explore the interaction patterns within online mathematics tutoring sessions aimed at enhancing online teaching and learning practices. However, our investigation faces two challenges. First, current research methods for analyzing educational dialog are primarily tailored toward traditional classroom settings and cannot be easily generalized to the unique dynamics of online one-on-one tutoring. For example, coding schemes like the academically productive talk framework and the scheme for educational dialog analysis were primarily developed for analyzing classroom conversations (Tao & Chen, 2023a) and may not be appropriate for online dialog. Second, the volume of dialog data derived from online tutoring is substantial, and traditional manual coding approaches to its analysis are inconvenient and laborious (Wang et al., 2023b). To tackle these challenges, we devised a coding scheme that builds on other frameworks and well-established learning theories for our analysis of online one-on-one tutoring dialog. We leveraged advanced artificial intelligence techniques to construct a model for automatically annotating online dialog, which improves the laborious process of manual coding. With the help of the newly developed coding scheme and the automatically annotated data, we aim to investigate dialogic interactions within the context of online one-on-one mathematics tutoring among different educational levels. The research questions guiding our study were as follows, which are designed to inform stakeholders, such as tutors and tutoring companies, about dialogic features in this context. We aim to suggest tailored tutoring strategies based on grade levels and specific features.

- **RQ1**: What are the basic characteristics of tutor teaching and student learning behaviors in online one-on-one mathematics tutoring, such as the similarities and differences of primary, middle, and high school levels?
- **RQ2**: What are the sequential patterns of dialogic behaviors in online one-on-one mathematics tutoring, such as tutors' teaching patterns, students' learning patterns, and teacher-student interactive patterns, among primary, middle, and high school levels?

2 Related work

2.1 Classroom dialog and dialogic teaching

The transformative potential of language in personal growth and development has motivated researchers to develop various teaching methods related to dialog, aimed at empowering teachers to effectively utilize specific forms of talk to enhance learners' knowledge acquisition and cognitive development (Cui & Teo, 2021; Tao & Chen, 2023a). These pedagogical approaches are primarily grounded in socio-cultural theory, which emphasizes the role of social interactions in learning processes (Lantolf et al., 2021). In the classroom context, "interaction" refers to the ongoing dialog between teachers and students and among students themselves. Unlike didactic instruction, which primarily involves teachers' authority, in dialogic interaction teachers can acknowledge students' voices, and space is created for student responses, thereby fostering their active participation and collective construction of knowledge (Skidmore, 2019; Tong & Ding, 2024).

Over the past 50 years, scholars and educators have extensively investigated classroom dialog and dialogic teaching from various perspectives, including theoretical clarifications, dialogic patterns, and the effects of dialogic teaching methods. For example, five principles of dialogic teaching were proposed, respectively collective, reciprocal, cumulative, supportive, and purposeful. These principles encourage teachers and students to jointly address learning tasks (collective), listen to each other (reciprocal), build on their own and others' ideas to create coherent lines of thinking (cumulative), articulate ideas without fear of wrong answers and support one another (supportive), and plan and facilitate teaching with clear educational goals (purposeful). Kim and Wilkinson (2019) provided a thorough explanation of the distinction between pedagogies related to dialog. In terms of dialogic patterns, classroom interactions predominantly followed the initiation-response-follow-up/evaluation (IRF/E) pattern (Song et al., 2019), as characterized by teacher questions, student responses, and subsequent teacher feedback or evaluation. In terms of the effects of dialogic teaching, empirical studies have concluded that students who received dialogic instruction demonstrated greater learning gains than others (e.g., Alexander, 2020; Chen et al., 2020; Resnick et al., 2018), which were not only sustained over time but also transferrable to other subjects. Additionally, such approaches have been shown to stimulate interest, foster critical thinking, enhance understanding, promote lifelong learning and democratic engagement, and support problem-solving and innovative exploration (e.g., Gillies, 2019; Howe et al., 2019; Hu et al., 2023).

As Howe et al. (2019) pointed out, productive dialog mainly centers on open-ended questions, extensions of previous contributions, reasoning and elaboration, and metacognition. These elements are also reflected in online tutoring, such as in the context of mathematics education. However, compared with the extensive literature on classroom dialog, relatively limited attention has been given to online tutoring. The remote nature of these interactions and the use of digital tools also complicate the dialogic process of online tutoring. Thus, it is necessary to investigate the optimization of online tutoring tools.

2.2 Face-to-face tutoring and online tutoring

Research on tutoring has historically been rooted in face-to-face contexts. Tutoring strategies, interactive patterns, and dialogic behaviors that promote the effectiveness of teaching and learning have been investigated (e.g., Chi et al., 2001; Cukurova et al., 2022; Graesser et al., 1995; Lin et al., 2024; Zhang et al., 2023). Scaffolding is a critical process in effective face-to-face tutoring (Chi et al., 2001), as it enables tutors to accurately assess students' understanding and elicit constructive responses. This adaptive approach to support requires tutors to adjust their scaffolding behaviors according to the student's level of mastery. When tutors note incorrect reasoning or answers, they can encourage students to explain their reasoning and intervene as needed. The teaching of less procedural and more declarative content can lead tutors to provide in-depth explanations to facilitate comprehension (Evens & Michael, 2006). The five-step model proposed by Graesser et al. (1995) encapsulates these strategies: (1) the tutor asks an initiating question; (2) the student provides an answer; (3) the tutor gives minimal feedback on the student's answer; (4) the tutor scaffolds to improve the student's answer; and (5) the tutor checks the student's understanding of the answer.

Technological advancement has enabled tutoring to expand beyond physical boundaries to the realm of online environments. Scaffolding is applied as a strategy in both online and face-to-face tutoring (Chi et al., 2001; Rus et al., 2017), but online tutoring has distinct characteristics. For example, the sessions are generally shorter than face-to-face sessions. The methods tutors use to engage students through virtual communication also vary. In online environments, unsolicited hints may prove less effective than engaging in task-based discussions or providing guidance only when students request it. Digital tools can also be effectively applied to enhance the tutoring process and can lead to improved learning outcomes (Zhang et al., 2023).

Despite the growing body of research on online tutoring, few studies have explored interaction patterns, particularly in the context of one-on-one tutoring. Therefore, it is crucial to examine the unclear dialogic interaction between pedagogical strategies and student behaviors to enhance the effectiveness of online one-on-one tutoring practices.

2.3 Automated dialog act annotation in education

An analysis of educational dialog can provide valuable insights for educators, who can make adjustments to their teaching practices accordingly. Researchers typically record and transcribe the dialog and then use a coding scheme to capture and examine any features and patterns (Wang et al., 2024). However, the detailed coding procedure, typically performed by skilled professionals, is both labor-intensive and time-consuming, thus limiting its scalability for studies involving an enormous amount of data (Song et al., 2019). To address this issue, researchers have applied automated annotation by utilizing natural language processing and machine learning techniques to classify dialog components.

Early studies in this area applied traditional machine learning models such as naive Bayes, decision trees, and Bayes nets, combined with various linguistic features. These models achieved moderate results but were not reliable or accurate. For example, Rus et al. (2017) applied traditional machine learning models to the first few tokens in each message of an online tutorial dataset, achieving an accuracy of 65.73% with a coding scheme involving 16 codes. Researchers later applied more advanced deep learning models, such as convolutional neural networks (e.g., Ma et al., 2021), recurrent neural networks (e.g., Lileikyte et al., 2022), and long short-term memory (LSTM) networks (e.g., Park et al., 2021). These models improved accuracy but were quickly surpassed by pre-trained language models that leverage large-scale datasets. Transformer-based pre-trained models, which can use the knowledge learned in large-scale pre-trained data, have recently been adopted for educational dialog act annotation (Song et al., 2021; Suresh et al., 2021; Wang et al., 2023b). These have been shown to outperform previous deep learning models in both classroom and tutoring contexts.

Despite these recent advancements, automated dialog annotation in online tutoring still faces two significant challenges. First, due to privacy and ethical concerns, access to online tutoring dialog datasets is frequently restricted, with few being publicly available. Second, the quality of the dialog dataset is closely linked to the performance of automated models. For example, the size and distribution of the training data can significantly affect model accuracy. To overcome these challenges and create a robust automated model for annotating online tutoring dialog, we collaborated with an international educational company to collect dialog data and applied a ChatGPT-based data augmentation method to enhance the quality of the dataset.

3 Method

3.1 Dataset

We obtained a deidentified mathematics tutorial dataset from a digital one-on-one tutoring platform. The dataset was derived from the unique and multifaceted educational system of Singapore and represents student-tutor interactions across grade levels from 1 to 12. In the process delineated by the platform, students are required to photograph the query they wish to address and designate accompanying specifications, such as the subject and the academic level of the question. Based on the information, the platform pairs the student with a suitable tutor. Through pedagogical dialog, the tutor then assists the student in resolving the query.

The collected dataset comprised dialog data from September 2018 to August 2020, including data from students who have free trials or who make token and subscription payments. As different payment methods may affect how students interact with tutors (e.g., students receiving free tutoring show different behavior from those receiving tutoring by subscription), we only selected dialog data from students with regular subscription payments. To facilitate a comparative analysis of dialog across different educational stages, we categorized the data into three levels: primary school (grades 1–6), middle school (grades 7–10), and high school (grades 11–12). The Singaporean educational system provides students with various options after the completion of primary school, but the most common is to proceed to middle school, which typically spans four years, although a five-year option is available. After middle school, students can opt to attend higher-level secondary institutions or pre-university schools. Students must indicate the grade level of their queries, so we considered it reasonable to group questions from grade 10 and above as reflective of a level equivalent to "high school" in other educational frameworks. Table 1 presents a basic description of the selected data, including the number of dialog sessions, total messages, and tutor and student messages.

3.2 Coding scheme design

To fully comprehend the online tutoring process, the dialog acts of each utterance in the dialogs must be annotated for further analysis. However, few established coding schemes specifically designed for online one-on-one tutoring dialog are currently available (Hrastinski et al., 2023). Thus, we followed the recommendations of Hennessy et al. (2020) and developed our own coding framework by integrating several well-established coding schemes, including the interactive, constructive, active, and passive (ICAP) framework (Chi & Wylie, 2014), the scheme for educational dialog

Table 1 Overview of the online one-on-one math tutoring dataset

School Stage	Number of Sessions	Number of Messages	Number of Tutor Messages	Number of Student Messages
Primary	4,238	38,317	25,652	12,665
Middle	8,581	89,834	61,029	28,805
High	977	21,822	12,407	9,415

analysis (Hennessy et al., 2016), and the Cambridge dialog analysis scheme (Hennessy et al., 2020).

Our scheme utilizes a two-tier structure consisting of six categories and 15 dialog acts. The ICAP framework classifies learning behavior into the four distinct categories of interactive, constructive, active, and passive, based on the level of learner engagement. The first two groups of our dialog act categories, *Question asking* and *Evaluation and feedback*, are representative of the interactive behaviors specified in the ICAP framework. The groups *Constructive participation* and *Active participation* are designed to mirror the constructive and active categories of ICAP, respectively. Note that we deliberately excluded the passive learning category, as behaviors within this category are difficult to directly discern from tutoring conversations.

We acknowledged the frequent instances of off-topic banter in our dataset by establishing a unique group, *Off-topic*. We also incorporated a *Metacognition* group to account for the instances of higher-level metacognitive behavior observed during the tutoring sessions. To accommodate the nuanced information inherent in each dialog act group, we subdivided the six primary categories into 15 dialog act behaviors. For example, we subdivided the *Question asking* group into three categories: *Constructive questions* (asking for opinions); *Procedural questions* (inquiring about solution procedures); and *Reasoning questions* (seeking explanations).

Some dialog acts are exclusive to either tutors or students whereas others may be used by both parties. For example, only students acknowledge others' expressions, termed as *Acknowledgment*, while *Knowledge-sharing and instruction* can only be enacted by tutors. However, dialog acts such as asking *Reasoning questions* or providing feedback (e.g., being agreeable, neutral, or negative) can be manifested by both tutors and students. The detailed structure of our revised coding scheme is given in Table 2.

3.3 Automatic annotator

3.3.1 Data annotation

To analyze the online one-on-one tutoring dialog, we recruited two experienced researchers with backgrounds in education to annotate a randomly selected subset of the whole dataset. This consisted of 1,000 tutoring sessions with a total of 9,257 messages. The researchers underwent a rigorous three-round training process to ensure they understood the designated coding scheme and to ensure their annotations were consistent. After they annotated the subset, we calculated Cohen's kappa coefficient and achieved a final value of 0.83, indicating substantial agreement between the two researchers. Any discrepancies in labeling were resolved through post-annotation discussion.

3.3.2 Data augmentation

The annotated data exhibited a distribution imbalance, with some codes represented fewer than 50 times. These accounted for only 0.5% of the annotated subset, demonstrating the need for augmentation for codes such as T.REQ, T.PQ, T.CQ, T.FB,

Table 2 The coding scheme for online one-on-one tutoring dialog

Group	Name	Code	Role	Description
Question asking	Constructive question	CQ	S&T	Ask for ideas and opinions that can lead to task resolution
	Procedural question	PQ	S&T	Inquire precisely regarding the method, technique, or sequence of actions required to address a task or attain a particular result
	Reasoning question	REQ	S&T	Ask for logical reasoning to prove, justify, or infer a cause-and-effect relationship with evidence or examples
Evaluation and feedback	Evaluation inducing question	EVQ	S&T	Ask the listener to evaluate the speaker's expressions
	Agreement	A	S&T	Affirm agreement or verify correctness
	Neutral and negative feedback	FB	S&T	Provide neutral or negative feedback to the listener's expression
Constructive participation	Constructive expression	C	S	Offer thoughts or opinions that help the solution of the task
	Solution procedure	P	S	Offer the method or steps involved in reaching a complete or partial solution and attaining specific results.
	Reasoning	RE	S	Elaborate on logical reasoning, justifications, or explanations supported by evidence or examples
	Knowledge sharing and instruction	KS	T	Provide information or instruction, share knowledge or explain solutions
Active participation	Acknowledgment	AP	S	Simple acknowledgment without additional constructive action
Metacognition	Reflection inducing question	RFQ	S&T	Students are asked to reflect on their understanding, performance, or progress
	Other reflection	RF	S&T	Reflection on understanding, performance, or progress
	Reflection of unknowing	RFN	S&T	Reflecting a partial or a lack of understanding
Off-topic	Off-topic chatting	OT	S&T	Off-topic expressions or questions

Note We will use S.xx and T.xx to represent the dialog act of an utterance. S and T refer to the role of student and tutor respectively, and xx refers to the code of dialog act

T.RFN, T.RF, S.EVQ, S.RFQ, S.RE, S.P, and S.FB. We addressed this imbalance by designing a data augmentation pipeline to develop an augmented dataset, as illustrated in Fig. 1. We utilized the cutting-edge generative AI model ChatGPT to generate 10 analogous contextual utterances for each utterance within the underrepresented codes, ensuring that the generation of a single, specific utterance was aligned with the chosen code but also the generation of an appropriate contextual conversation (Shan et al., 2023c). To preclude overfitting in the training of the automatic annotator, a Self-BLEU threshold was used as a filter, rejecting any generated short conversations exhibiting a similarity score above 0.5. The Self-BLEU is a metric used to evaluate the diversity of generated text. Finally, the original annotated data and generated data were used to train a more reliable automatic model for annotating the remaining dialog data.

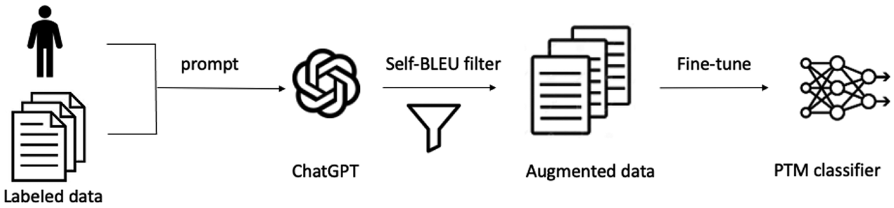

Fig. 1 The data augmentation pipeline for augmenting selected dialogic behavior and fine-tuning the PTM classifier

3.3.3 Model construction

To build an automatic annotator, we used Roberta, a transformer-based pre-trained model, which has been found to perform well in sequence classification tasks (e.g., Suresh et al., 2021). We applied the five-fold cross-validation approach for training and testing and achieved an average F1 score of 0.89. As a metric commonly used in classification tasks to gauge a model's accuracy by balancing precision and recall, the F1 score was regarded as the most appropriate measure of the multiple code categories, as it provides a more precise reflection of the annotation than other metrics. This F1 score surpassed that of the same automatic annotator model when solely trained on the original data, which had a value of 0.86. The score exceeded the performance of the human annotators, thus indicating higher accuracy.

We then applied our automatic annotator to the remaining unannotated data for further dialogic pattern analysis. We examined 100 randomly selected sessions to validate its reliability and verified the machine-annotated labels, with an F1 score of 0.87. This confirms the effectiveness of the automatic model in annotating the data.

3.4 Data analysis

For RQ1 to investigate the characteristics of tutor teaching and student learning behaviors in online one-on-one mathematics tutoring, we first used basic descriptive statistics to identify the aggregated quantity of messages and the proportion of each code usage throughout the tutoring sessions across primary, middle, and high school levels. This provided valuable insights into the teaching approaches of the tutors and the learning preferences of students. We then assessed these preferences in more detail using basic descriptive statistics and one-way analysis of variance (ANOVA) to ascertain and compare the instances of each unique code within the individual expressions of tutors and students across the three school stages.

For RQ2, we used the lag sequential analysis (LSA) method to assess the dialogic sequential patterns across the three school stages in online one-on-one mathematics tutoring. LSA is mainly used to compute the probability that an action is followed by another action and evaluate whether this probability is statistically significant (Wang et al., 2023). This sequence mining technique can decipher the sequential characteristics of intricate processes by estimating the correlations between data process streams and graph structures. It has been successfully used to analyze learners' behavior in various educational settings (e.g., Huang et al., 2019). We calculated the conditional probability and the adjusted residual (z-score) for each sequential pattern. The conditional probability indicates the likelihood of transitioning from the initial dialogic behavior to the target dialogic behavior. The adjusted residual (z-score) was used to evaluate the statistical significance of the patterns. We selected dialogic behavior patterns with conditional probabilities of above 5% and adjusted z-scores of over 1.96 (corresponding to a p-value of less than 0.05), which enabled us to identify practical and statistically significant sequential relationships. Accordingly, we constructed four diagrams representing tutor-to-tutor (T-T), student-to-student (S-S), tutor-to-student (T-S), and student-to-tutor (S-T) interactions, providing a comprehensive overview of the potential interactive dynamics within the dialogic tutoring sessions.

4 Results

4.1 Basic characteristics of dialogic behavior

4.1.1 Characteristics common to all three school stages

Table 3 outlines key statistics from online one-on-one math tutoring sessions, including message count and teacher-student message ratio across educational levels. A trend reveals tutors contributed a larger message portion, with primary, middle, and high school levels at 74.26%, 77.02%, and 63.05% tutor message contribution respectively, indicating a general tutor dominance in online discourse.

Next, we examine the distribution of dialogic code ratios within each speaker's individual messages, to gain a more nuanced understanding of teaching and learning preferences.

In terms of tutors' dialogic behavior, Fig. 2 shows that off-task interactions (T.OT) constituted more than 30% of the tutors' total behavior. Task-related behaviors, specifically those concerning didactic instruction and knowledge sharing (T.KS), constituted more than 40% of all of the behaviors observed. The second most common task-related behavior involved tutors checking students' understanding (T.RFQ). Interestingly, tutors dedicated little effort to evaluating students' expressions, with positive feedback (T.A) and neutral or negative feedback (T.FB) together constituting less than 5% of their total behavior.

We found that off-topic discussions constituted more than 30% which is also a significant proportion of students' dialogic behavior, as Fig. 3 shows. Students consistently devoted a higher proportion of their interactions to posing questions(S. CQ, S.PQ, S.REQ) but the total ratio is still less than 10%. We also observed they expressed a larger number of constructive claims (S.C) than procedural solutions

Table 3 Key descriptive statistics of the online one-on-one math tutoring dataset

School Stage	Messages Count		Teacher Message Ratio		Student Message Ratio	
	Mean	SD	Mean	SD	Mean	SD
Primary	9.041	12.26	74.26%	20.08%	25.74%	20.08%
Middle	10.469	15.697	77.02%	20.09%	22.98%	20.09%
High	22.336	32.972	63.05%	18.48%	36.95%	18.48%

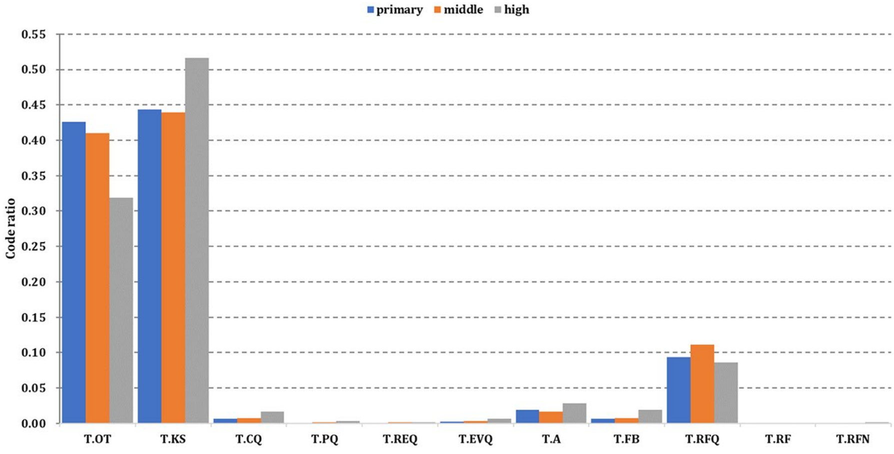

Fig. 2 The distribution of tutors' dialogic code ratios within their own individual messages

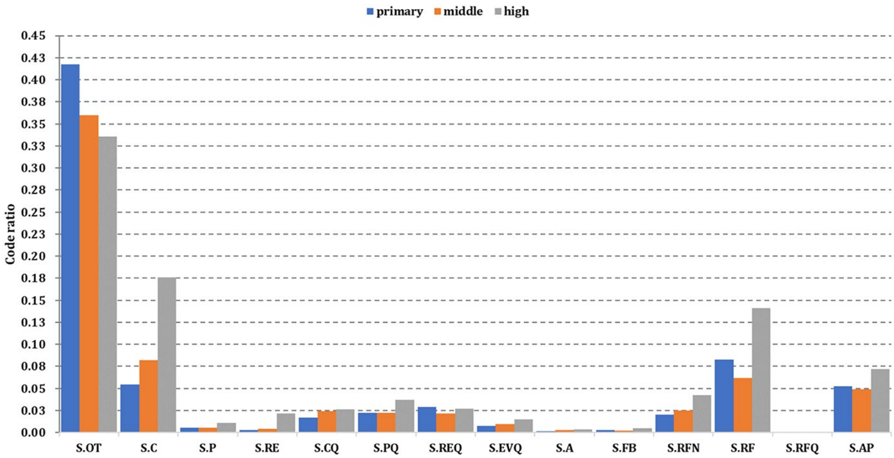

Fig. 3 The distribution of students' dialogic code ratios within their own individual messages

(S.P) and reasoning (S.RE). Regarding students' metacognition, they reflected on their lack of knowledge (S.RFN) to a lesser degree than on other areas (S.RF).

4.1.2 Different characteristics in three school stages

Table 3 shows the average messaging frequency between teachers and students across educational levels. High school tutoring sessions had significantly more messages compared to primary or middle school sessions, with mean message count being over twice. High school sessions also displayed a more balanced interaction, with a notably smaller proportion of tutor messages than in the earlier stages.

These findings presented in Figs. 2 and 3 indicate that the dialogic behaviors of tutors and students at different educational levels may vary. We conducted one-way

ANOVA tests on the ratio of dialogic behaviors within their own expressions to explore these differences further. The results, summarized in Table A1 and A2 in the appendix, generally indicated significant differences between the school stages, with the student reflection question (S.RFQ), tutor reflection (T.RF), and the tutor reasoning question (T.REQ) being exceptions.

Table A1 shows that high school tutors engaged in off-task chats (T.OT) significantly less than middle and primary school tutors. They provided more task-related instruction (T.KS) and posed more questions (T.CQ, T.PQ, TREQ). Middle school tutors checked understanding (T.RFQ) and posed more evaluation questions (T.EVQ) than primary tutors. High school tutors gave significantly more positive (T.A) and neutral or negative feedback (T.FB) than the other tutors.

Regarding student behavior, Table A2 illustrates that primary students engaged more in off-task chatting (S.OT) while high school students exhibited more constructive behaviors (S.C, S.P, S.RE), and asked more evaluation-inducing (S.EVQ) and procedural questions (S.PQ). Middle school students made more low-level constructive claims (S.C) than primary students. High school students also provided more evaluation feedback (S.FB, S.A), expressed more acknowledgement (S.AP), and reflected significantly on their understanding (S.RFN, S.RF) compared to the other groups, with primary students reflecting more than middle school students.

4.2 The sequential structure of dialogic behavior

To examine the dialogic patterns in online one-on-one mathematics tutoring, we conducted lag sequential analysis for each of the primary, middle, and high school levels. The sequential transition diagrams in Figs. 4, 5 and 6, and 7 are based on patterns with a z-score exceeding 1.96 and a conditional probability value higher than 0.05. These represent the sequential patterns of tutor pedagogical behavior, student learning behavior, tutor-initiated and student-followed behavior, and student-initiated and tutor-followed behavior, respectively. Nodes represent dialogic behaviors, edges indicate significant transitions between these behaviors, and arrows display preceding and following behaviors. The numbers on the arrows and in parentheses denote the z-scores and conditional probabilities of the dialogic behaviors, respectively. Black arrows indicate shared patterns across all three stages, while red arrows highlight patterns specific to individual educational levels.

4.2.1 Tutor consecutive teaching pattern

Figure 4 shows a recurring pedagogical pattern, denoted as T.FB -> T.KS -> T.RFQ, across all three educational stages. This reflects the path of a tutor offering neutral or negative feedback (T.FB) in response to a student's previous expression, followed by sharing instruction or knowledge (T.KS), and subsequently ascertaining whether the student comprehends the imparted information (T.RFQ). Various questioning patterns were also evident across the stages: primary school tutors commonly asked evaluative questions (T.EVQ), middle school tutors frequently requested procedural solutions (T.PQ), while high school tutors often posed reasoning (T.REQ) and constructive claim questions (T.CQ).

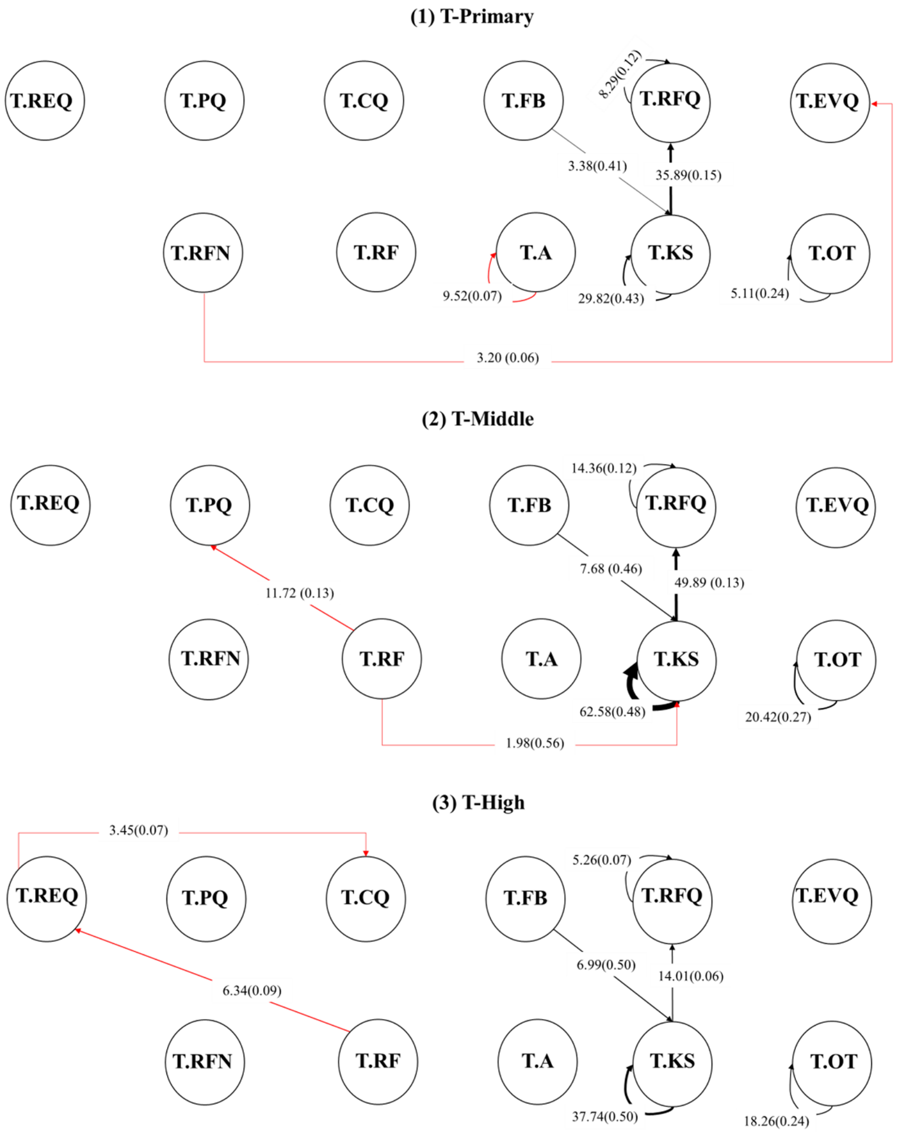

Fig. 4 The sequential pattern of tutors' consecutive dialogic behaviors

4.2.2 Student consecutive learning patterns

Figure 5 shows that the learning pattern S.CQ -> S.C was prevalent across all stages, which indicates a student making a constructive claim (S.C) following their own question (S.CQ). Compared with primary school students, middle and high school students more frequently demonstrated behavior patterns connected to a constructive claim (S.C). For example, higher cognitive questions (S.PQ and S.REQ) and advanced constructive engagements (S.P and S.RE) were often followed by S.C.

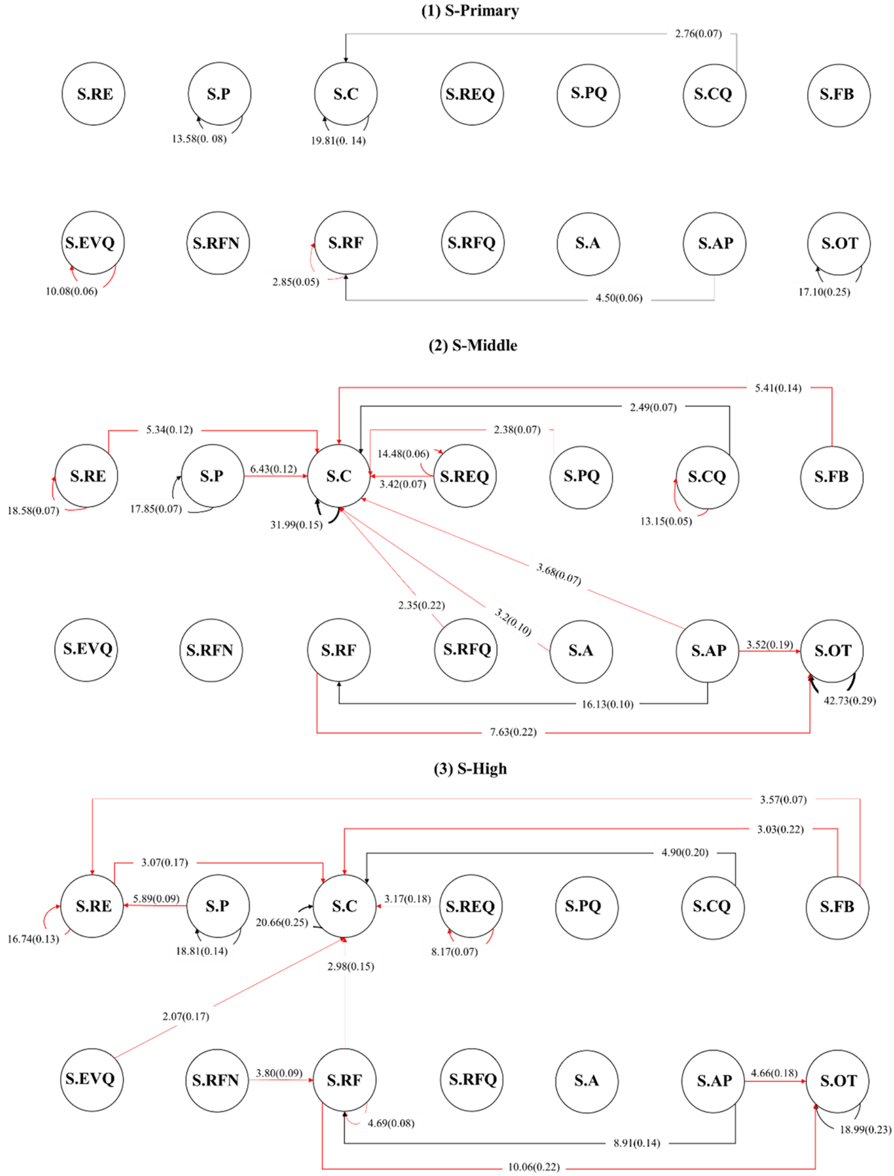

Fig. 5 The sequential pattern of students' consecutive dialogic behaviors

Reasoning (S.RE) often followed a procedural solution (S.P) or neutral/negative feedback (S.FB) for high school students. Students' behaviors sometimes led to off-task chatting. For example, self-reflection (S.RF) and acknowledgment (S.AP) transitioned into off-task conversation (S.OT) for secondary and high school students.

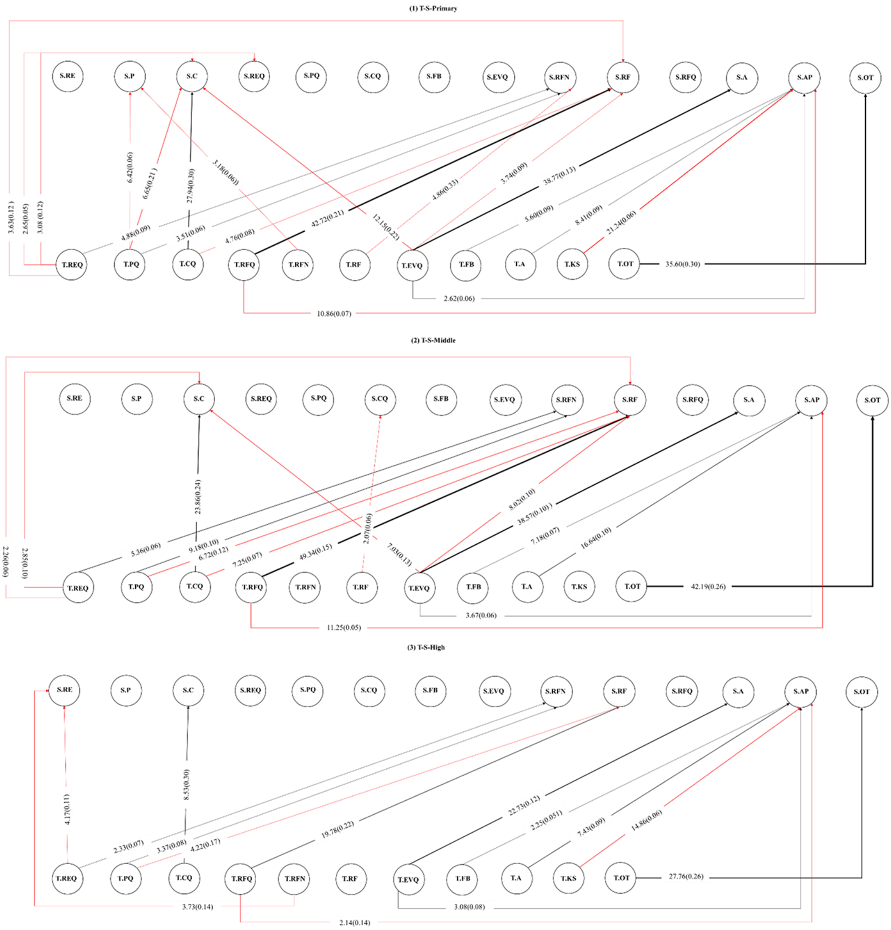

Fig. 6 The sequential pattern of teacher-initiated-student-followed dialogic behaviors

4.2.3 Interactive patterns: tutor-initiated behaviors and student responses

Figure 6 highlights the interplay between tutors' instructional behaviors and students' responses across primary and high schools. The transmission of knowledge and instruction by the tutors (T.KS) did not appear to effectively encourage any constructive student behaviors (S.C, S.P, S.RE) in either primary or high school contexts. Student acknowledgment (S.AP) was instead the main response. This diagram also identifies other common patterns across the school levels. For example, lower-level constructive questions posed by the tutor (T.CQ) typically elicited constructive claims (S.C), while tutor-led evaluation invitation questions (T.EVQ) lead to affirmative acknowledgments (S.A) from the students. Interestingly, tutors' higher-order cognitive questions, i.e., procedural solution questions (T.PQ) and reasoning questions (T.REQ), predominantly resulted in students expressing a lack of comprehension (S.RFN). In terms of metacognition, tutors' reflection-inducing questions

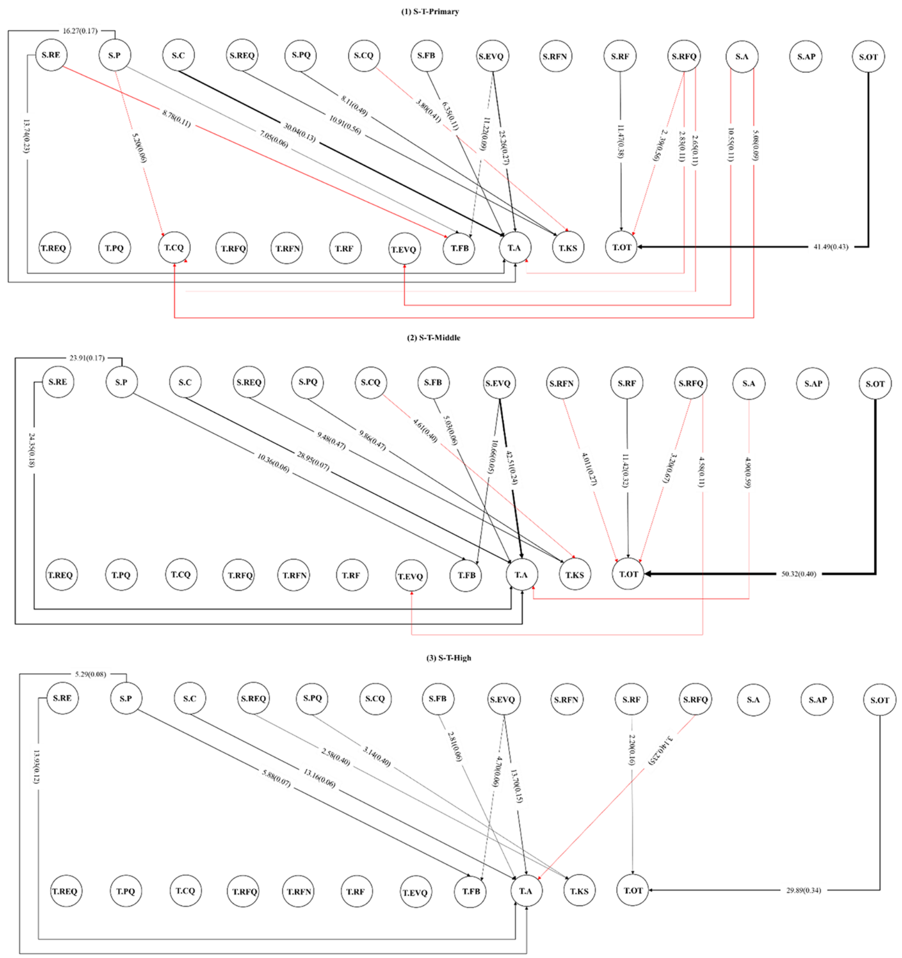

Fig. 7 The sequential pattern of student-initiated-teacher-followed dialogic behaviors

(T.RFQ) was found to lead to student reflection (S.RF), but not reflection on misunderstandings (S.RFN).

Distinct interaction patterns emerged at each educational stage. In the primary school sessions, the responses to tutors' high-level constructive questions were typically in the form of expressions or inquiries. For example, only in primary schools did procedural questions (T.PQ) lead to procedural claims (S.P). Additionally, students tended to pose reasoning questions (S.REQ) in response to the tutor's similar inquiry. Lower-level responses to high-level queries represent a notable pattern, as T.PQ -> S.C and T.REQ -> S.C indicate. Constructive claims (S.C) also provided students with the opportunity to pose evaluation questions to the tutor (T.EVQ). The middle school sessions revealed fewer unique interaction patterns than the primary school sessions. The patterns identified generally overlapped with those identified in primary schools. The one distinct pattern identified involved a low-level constructive claim question (S.CQ) following the tutor's reflection (T.RF). High school sessions were the least likely to demonstrate patterns common to other levels. Their interactions primarily led to students' reasoning (S.RE) in response to tutors' reasoning questions (T.REQ) and reflections on misunderstandings (T.RFN). A similar pattern was found in primary schools, where tutors' reflections on misunderstandings (T.RFN) prompted a procedural solution response from students (S.P).

4.2.4 Interactive patterns: student-initiated behaviors and tutor responses

Figure 7 reveals several prevalent patterns of interactions between student-initiated behaviors and tutor responses. For example, high-level constructive questions posed by students (S.PQ and S.REQ) generally elicited instructions and knowledge sharing (T.KS) from tutors. Similarly, when students posed evaluation invitation questions (S.EVQ), tutors tended to respond with evaluation feedback, denoted as T.A and T.FB. Tutors also gave positive feedback in response to all forms of students' constructive expressions, which encompassed S.C, S.P, and S.RE. However, a procedural solution (S.P) offered by a student may trigger negative or neutral feedback (T.FB) from the tutor. Surprisingly, negative or neutral feedback from students (S.FB) was likely to be met with positive feedback (T.A) from tutors. Finally, both student reflection (S.RF) and off-task chatting (S.OT) generally led to off-task chatting from the tutor.

Distinct interaction patterns were observable at each educational stage. As the Tutor–Student diagram shows, primary schools exhibited the greatest number of unique patterns, whereas high schools demonstrated the least. Lower-level constructive questions (S.CQ) from the students typically received a T.KS response in primary and middle schools. The emergence of a constructive follow-up question from the tutor (T.CQ) after a student's procedural solution claim (S.P) revealed a noteworthy pattern exclusive to primary schools. Primary school tutors also tended to pose low-level constructive questions (T.CQ) and evaluation-inducing questions (T.EVQ) in response to students expressing positive feedback (S.A).

5 Discussion

5.1 The shared characteristics of online one-on-one math tutoring

5.1.1 Tutor dialogic behavior characteristics

Firstly, in our investigation, tutors constitute a significant portion of tutoring sessions, delivering over 63% of communications, primarily following a 'provide-and-explain' model, aligning with findings from offline tutoring research (Chi et al., 2001; Graesser et al., 1995). They often used reflection-inducing queries for comprehension assessment, reflecting a consistent feedback-instruction-understanding assessment pattern across all educational level. This model, criticized for not fostering advanced cognitive thinking, was also corroborated by our findings where student responses were mainly acknowledgment of instruction. Tutors' questions didn't stimulate higher cognitive activities, and the lack of student self-reflection on misconceptions or understanding gaps was notable. Due to self-doubt or inadequate self-evaluation confidence, students struggle to discern their understanding, hence reflection-inducing questions may not help identify knowledge deficiencies.

Secondly, Evens and Michael (2006) report that experienced tutors use fewer comprehension assessment questions than novice tutors, instead inferring understanding through knowledge-based questions, requiring a thorough subject matter grasp and awareness of common student errors. In online settings, time constraints may hinder tutors from accurately assessing student understanding. Pattern of tutors' use of lower cognitive constructive claim questions followed by student's constructive response might suggest its ability to boost student confidence and self-efficacy, yet rarely posed are cognitively challenging questions demanding procedural solutions and reasoning. Effectively evaluating responses to such questions requires envisioning multiple problem-solving approaches and understanding each solution's logic (Iwuanyanwu, 2020), expertise often lacking in online tutors. This lack, and fear of inadequately explaining student responses, may explain the rare use of cognitively demanding questions, potentially detrimental to student learning. Not posing these questions can leave student misconceptions unaddressed.

Thirdly, after student post-confusion, tutors didn't exhibit significant pedagogical behavior. Effective tutors, according to Evens and Michael (2006), should clarify doubts and guide students towards problem-solving, with encouraging student reasoning suggested as good scaffolding pedagogy. However, online tutors in our study seemed to fail to act in this way. Chi et al. (2001) suggest novice tutors could attain the goal by posing questions without providing answers, possibly a more feasible approach for online tutors given the time constraints in mastering the subject matter in one-on-one online tutoring.

5.1.2 Student dialogic behavior characteristics

Firstly, a pattern of off-task conversation in online one-on-one tutoring sessions reveals a propensity for distraction among students. However, such casual interactions, often reflective in nature, can build rapport, enhance comfort, and inform tutor's teaching approaches (Ong & Quek, 2023).

Secondly, student queries, constituting about 10% of interactions, surpass those of tutors, with a balanced distribution across types. Tutors address constructive questions with instructions and evaluations, showcasing the benefits of student-initiated inquiries. Notably, an answer self-question pattern (S.CQ -> S.C) across all school stages indicates self-initiative in expression, suggesting the potential for enhancing critical thinking by encouraging students to answer their own questions.

Thirdly, higher-level constructive expressions are less frequent compared to lower-level claims, possibly due to a scarcity of tutor-posed questions or lack of student initiative. This gap could impede cognitive processes and learning gains (Graesser et al., 1995). Tutor reflections on unknowing elicited high-level constructive expressions in primary and high schools, hinting at a pedagogical approach for enhancing self-efficacy. The data also suggests that posing higher-level questions could elicit a broader response range and stimulate higher thinking levels.

Fourthly, frequent student self-reflection in response to tutor questions demonstrates resilience and metacognitive skill practice. This behavior aids tutors in understanding student comprehension (Chi et al., 2001), facilitating personalized guidance and intervention strategies, thus fostering a conducive learning environment.

5.2 Different characteristics of primary and high school students

The interaction dynamics in online tutoring sessions varied significantly between high school and primary school students. High school students demonstrated distinct characteristics and patterns from those of, and notably higher engagement than, primary and middle school students. We identified more than double the average number of utterances per session and a significantly greater proportion of expressions in their tutoring conversations. Their proportions of constructive expressions, questions, evaluative feedback, and metacognitive reflections were markedly higher. Over 25% of primary school and over 30% of middle school tutoring conversations were "passive" modes of engagement, with students contributing no expressions (Chi & Wylie, 2014). This rate significantly decreased to 8.2% in high school. High school students also exhibited less off-task chatter, suggesting higher levels of focus and engagement during online math tutoring sessions. Despite consistent pedagogical tutoring approaches online, the high school tutors demonstrated fewer patterns than primary and middle school tutors when responding to students' expressions, with their generally didactic teaching style. The minimal level of feedback following students' expressions in high school further supports this observation. The higher levels of interactions in high school tutoring dialogs may therefore be primarily student-driven. This is unlike traditional offline tutoring, in which the tutor primarily guides the tutoring process and adheres to IRF interactive patterns. In addition, we identified more patterns related to reasoning in high school students' self-consecutive behavior and tutor–student interactions. This reasoning followed student feedback and procedural assertions, which illustrates the students' proactive approach to offering additional explanations while solving tasks or providing feedback to tutors. Following the tutor–student pattern, high school students were the only group to commonly respond to tutors' behavior with reasoning. The ability to generate "self-explanations" is one characteristic that distinguishes good from poor learners, and we found that high school students exhibited such behaviors better during tutoring. The capacity to provide reasoning is also considered to be a higher cognitive ability, indicating that high school students possess better cognitive abilities and thus higher cognitive questions are less likely to affect their self-efficacy.

Primary school students also exhibited unique characteristics. They demonstrated the highest rate of absenteeism and the least diverse self-consecutive patterns, implying that they had a lower level of initiative. A high rate of off-task interactions suggests that primary students may be more prone to distractions. However, when opportunities are provided (e.g., when tutors pose questions), these students were more responsive, as illustrated by the more varied patterns in response to tutors' questions. They engaged with tutors' queries and even reciprocated with questions of their own. While primary students may struggle to answer higher-level constructive questions, they attempted to respond with lower-level responses. These observations imply that although primary students may begin tutoring sessions with low confidence, given the opportunity they are more inclined to participate. This highlights the importance of tutors' questions in teaching primary school students, suggesting that these young learners require a secure and comfortable learning environment to develop their confidence and self-efficacy. However, as the online tutors in this study regrettably employed a didactic knowledge-telling teaching style, they did not generally encourage students to participate and articulate their thoughts and reasoning. This can result in fewer interactions and decreased learning gains during primary school tutoring sessions.

6 Conclusion

To better understand and improve teaching and learning in the context of online tutoring, we present an analysis of the dialog in online one-on-one, task-based math tutoring sessions across primary, middle, and high school levels. We devised a coding scheme and utilized advanced AI techniques to build a dialog annotation model. With the annotated data, we used statistical methods and lag sequential analysis to reveal the characteristics of online tutoring dialog.

We found that a significant proportion of the tutoring involved off-task chatting, a characteristic unique to the context of online one-on-one tutoring. The tutoring strategies were predominantly didactic, and they were characterized by extensive instruction and imparting of knowledge, with minimal elicitation of questions and limited evaluation from the tutor. These strategies were found to negatively impact student learning gains and self-efficacy, as demonstrated by the lack of constructive expressions in the interactive patterns. However, we also identified several effective teaching strategies, including some that were unexpected such as tutors' reflections on their own misunderstandings leading to students' constructive expressions. These strategies can contribute to the corpus of effective pedagogies. We also explored the differences in student behavior across different school stages during online one-on-one mathematics tutoring sessions. High school students demonstrated higher engagement and cognitive abilities than primary and middle school students. They participated more actively, exhibited superior reasoning skills, and their interactions were more self-led. However, although primary school students were less active, they gave positive responses when engaged by their tutors. This highlights the need for a supportive learning environment and interactive teaching styles, particularly for younger students, as it can enhance their participation and improve their learning outcomes.

Notwithstanding the insightful discoveries, this investigation is accompanied by several limitations. First, it was executed utilizing data from a singular country, sourced from a specific online one-on-one tutoring platform that exhibited an imbalanced distribution across various school stages. This aspect potentially constrains the broad applicability of our findings. Second, our coding scheme is not exhaustive and may be potentially unable to capture all teaching and learning behaviors. Third, while the AI model developed for dialog annotation exhibited an accuracy higher than humans, it could still have biases and may not be completely reliable. The sequence mining technique of lag sequential analysis only captures a lag of one (i.e., two consecutive dialogic behaviors), ignoring other potentially insightful patterns in other lengths (e.g., three or more consecutive dialogic behaviors). Future studies can address these limitations by diversifying the data sources, refining the coding scheme, and improving the reliability of the annotator model. However, despite these limitations, we believe that our findings provide valuable insights into the nature of online one-on-one tutoring for future research in this area.

Appendix

Table A1 The ANOVA results comparing the tutor dialogic code ratio in tutors' self-expression

Dialogic Code	(I)	(J)	F	*p*	Mean Difference(I-J)
T.RF	Primary	Middle	2.927	0.054	0.000
	Primary	High			0.000
	Middle	High			0.000
T.EVQ	Primary	Middle	12.632	0.000	-0.001**
	Primary	High			-0.004***
	Middle	High			-0.003**
T.RFN	Primary	Middle	3.528	0.030	0.000
	Primary	High			-0.001
	Middle	High			-0.001
T.RFQ	Primary	Middle	33.741	0.000	-0.018***
	Primary	High			0.008
	Middle	High			0.026***
T.REQ	Primary	Middle	0.999	0.368	0.000
	Primary	High			-0.001
	Middle	High			0.000
T.A	Primary	Middle	13.759	0.000	0.002
	Primary	High			−0.010***
	Middle	High			−0.012***
T.PQ	Primary	Middle	8.594	0.000	0.000
	Primary	High			-0.002***
	Middle	High			−0.002**
T.FB	Primary	Middle	21.939	0.000	-0.001
	Primary	High			-0.013***
	Middle	High			−0.0117***
T.OT	Primary	Middle	91.848	0.000	0.0164**
	Primary	High			0.108***
	Middle	High			0.091***
T.CQ	Primary	Middle	17.913	0.000	0.000
	Primary	High			−0.010***
	Middle	High			−0.009***

Table A1 The ANOVA results comparing the tutor dialogic code ratio in tutors' self-expression

Dialogic Code	(I)	(J)	F	p	Mean Difference(I-J)
T.KS	Primary	Middle	50.524	0.000	0.004
	Primary	High			−0.073***
	Middle	High			−0.077***

*$p<0.05$, **$p<0.01$, ***$p<0.001$

Table A2 The ANOVA results comparing the tutor dialogic code ratio in students' self-expression

Dialogic Code	(I)	(J)	F	p	Mean Difference(I-J)
S.A	Primary	Middle	5.906	0.003	-0.001
	Primary	High			-0.002**
	Middle	High			-0.001
S.OT	Primary	Middle	34.699	0.000	0.058***
	Primary	High			0.082***
	Middle	High			0.024
S.EVQ	Primary	Middle	8.352	0.000	-0.002
	Primary	High			-0.007***
	Middle	High			-0.005**
S.PQ	Primary	Middle	10.076	0.000	0.001
	Primary	High			-0.014***
	Middle	High			-0.015***
S.RFN	Primary	Middle	18.209	0.000	-0.005*
	Primary	High			-0.023***
	Middle	High			-0.017***
S.FB	Primary	Middle	4.977	0.007	0.000
	Primary	High			-0.002**
	Middle	High			-0.003
S.AP	Primary	Middle	11.041	0.000	0.004
	Primary	High			-0.019***
	Middle	High			-0.023***
S.CQ	Primary	Middle	10.399	0.000	-0.008***
	Primary	High			-0.009**
	Middle	High			-0.002
S.RE	Primary	Middle	35.241	0.000	-0.001
	Primary	High			-0.019***
	Middle	High			-0.018***
S.REQ	Primary	Middle	6.376	0.002	0.007**
	Primary	High			0.002
	Middle	High			-0.006
S.RF	Primary	Middle	66.698	0.000	0.021***
	Primary	High			-0.058***
	Middle	High			-0.0791***
S.C	Primary	Middle	156.46	0.000	-0.027***
	Primary	High			-0.121***
	Middle	High			-0.094***
S.RFQ	Primary	Middle	0.412	0.662	0.000
	Primary	High			0.000
	Middle	High			0.000

Table A2 The ANOVA results comparing the tutor dialogic code ratio in students' self-expression

Dialogic Code	(I)	(J)	F	*p*	Mean Difference(I-J)
S.P	Primary	Middle	6.405	0.002	0.000
	Primary	High			-0.005**
	Middle	High			-0.005**

*$p<0.05$, **$p<0.01$, ***$p<0.001$

Acknowledgements This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contribution Deliang Wang and Dapeng Shan: Conceptualization, Methodology, Investigation, Formal Analysis, Writing - Original Draft, Writing - Reviewing & Editing; Ran Ju, Ben Kao, Gaowei Chen, and Chenwei Zhang: Methodology, Reviewing & Editing.

Data availability The data will be available upon reasonable request from the first author.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Al-Adeimi, S., & O'Connor, C. (2021). Exploring the relationship between dialogic teacher talk and students' persuasive writing. *Learning and Instruction*, *71*, 101388.

Alexander, R. (2020). A dialogic teaching companion *(1st ed.)*. Routledge. https://doi.org/10.4324/9781351040143

Booth, B. M., Jacobs, J., Bush, J. B., Milne, B., Fischaber, T., & DMello, S. K. (2024). Human-tutor coaching technology (htct): Automated discourse analytics in a coached tutoring model. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 725–735).

Chen, G., Zhang, J., Chan, C. K., Michaels, S., Resnick, L. B., & Huang, X. (2020). The link between student-perceived teacher talk and student enjoyment, anxiety and discursive engagement in the classroom. *British Educational Research Journal*, *46*(3), 631–652.

Chi, M. T., & Wylie, R. (2014). The icap framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243.

Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, *25*(4), 471–533.

Cui, R., & Teo, P. (2021). Dialogic education for classroom teaching: A critical review. *Language and Education*, *35*(3), 187–203.

Cukurova, M., Khan-Galaria, M., Millán, E., & Luckin, R. (2022). A learning analytics approach to monitoring the quality of online one-to-one tutoring. *Journal of Learning Analytics*, *9*(2), 105–120.

Evens, M., & Michael, J. (2006). *One-on-one tutoring by humans and computers*. Psychology.

Gillies, R. M. (2019). Promoting academically productive student dialogue during collaborative learning. *International Journal of Educational Research*, *97*, 200–209.

Gortazar, L., Hupkau, C., & Roldán-Monés, A. (2024). Online tutoring works: Experimental evidence from a program with vulnerable children. *Journal of Public Economics*, *232*, 105082.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialog patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, *9*(6), 495–522.

Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A. M., Maine, F., Ríos, R. M., García-Carrión, R., Torreblanca, O., & Barrera, M. J. (2016). Developing a coding scheme for analyzing classroom dialog across educational contexts. *Learning Culture and Social Interaction*, *9*, 16–44.

Hennessy, S., Howe, C., Mercer, N., & Vrikki, M. (2020). Coding classroom dialog: Methodological considerations for researchers. *Learning Culture and Social Interaction*, *25*, 100404.

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher–student dialog during classroom teaching: Does it really impact on student outcomes? *Journal of the Learning Sciences*, *28*(4–5), 462–512.

Hrastinski, S., Stenbom, S., Saqr, M., Jansson, M., & Viberg, O. (2023). Examining the development of K-12 students' cognitive presence over time: The case of online mathematics tutoring. *Online Learning*, *27*(3), 252–271.

Hu, L., Chen, G., & Wu, J. (2023). Teaching talk for thinking: The efficacy of a peer talk teaching program for improving group thinking. *Thinking Skills and Creativity*, *48*, 101291.

Huang, C. Q., Han, Z. M., Li, M. X., Jong, M. S., & Tsai, C. C. (2019). Investigating students' interaction patterns and dynamic learning sentiments in online discussions. *Computers & Education*, *140*, 103589.

Iwuanyanwu, P. N. (2020). Nature of problem-solving skills for 21st century STEM learners: What teachers need to know. *Journal of STEM Teacher Education*, *55*(1), 27–40.

Jacobs, G. M., & Renandya, W. A. (2021). Any questions? Ideas for encouraging more and better student questions. *LLT Journal: A Journal on Language and Language Teaching*, *24*(2), 349–363.

Kim, M. Y., & Wilkinson, I. A. (2019). What is dialogic teaching? Constructing, deconstructing, and reconstructing a pedagogy of classroom talk. *Learning Culture and Social Interaction*, *21*, 70–86.

Lantolf, J. P., Xi, J., & Minakova, V. (2021). Sociocultural theory and concept-based language instruction. *Language Teaching*, *54*(3), 327–342.

Lileikyte, R., Irvin, D., & Hansen, J. H. (2022). Assessing child communication engagement and statistical speech patterns for American English via speech recognition in naturalistic active learning spaces. *Speech Communication*, *140*, 98–108.

Lin, J., Raković, M., Li, Y., Xie, H., Lang, D., Gašević, D., & Chen, G. (2024). On the role of politeness in online human–human tutoring. *British Journal of Educational Technology*, *55*(1), 156–180.

Ma, Y., Wiggins, J. B., Celepkolu, M., Boyer, K. E., Lynch, C., & Wiebe, E. (2021). The challenge of noisy classrooms: Speaker detection during elementary students' collaborative dialogue. *Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part I*, 268–281.

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The promise of tutoring for PreK–12 learning: A systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal*, *61*(1), 74–107.

Ong, S. G. T., & Quek, G. C. L. (2023). Enhancing teacher–student interactions and student online engagement in an online learning environment. *Learning Environments Research, 26*(3), 681–707.

Park, K., Sohn, H., Mott, B., Min, W., Saleh, A., Glazewski, K., Hmelo-Silver, C., & Lester, J. (2021). Detecting disruptive talk in student chat-based discussion within collaborative game-based learning environments. *Proceedings of the 11th International Learning Analytics and Knowledge Conference*, 405–415.

Resnick, L. B., Asterhan, C. S., Clarke, S. N., & Schantz, F. (2018). Next generation research in dialogic learning. *Wiley handbook of teaching and learning*, pp. 323–338.

Rus, V., Maharjan, N., & Banjade, R. (2017). Dialogue act classification in human-to-human tutorial dialogues. *Innovations in Smart Learning. Lecture notes in Educational Technology*. Springer. https://doi.org/10.1007/978-981-10-2419-1_25., *et al.Popescu*, *E.*.

Shan, D., Wang, D., Zhang, C., Kao, B., & Chan, C. K. (2023). Annotating Educational Dialog Act with Data Augmentation in Online One-on-One Tutoring. In Wang, N., Rebolledo-Mendez, G., Dimitrova, V., Matsuda, N., Santos, O.C. (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky. AIED 2023. Communications in Computer and Information Science* (pp. 472–477). Cham: Springer Nature Switzerland.

Skidmore, D. (2019). Dialogism and education. *The Routledge International Handbook of Research on Dialogic Education*, pp. 27–37.

Skovsmose, O. (2020). Dialogic teaching and learning in mathematics education. *Encyclopedia of Mathematics Education*, pp. 195–197.

Song, Y., Chen, X., Hao, T., Liu, Z., & Lan, Z. (2019). Exploring two decades of research on classroom dialog by using bibliometric analysis. *Computers & Education*, *137*, 12–31.

Song, Y., Lei, S., Hao, T., Lan, Z., & Ding, Y. (2021). Automatic classification of semantic content of classroom dialog. *Journal of Educational Computing Research*, *59*(3), 496–521.

Suresh, A., Jacobs, J., Lai, V., Tan, C., Ward, W., Martin, J. H., & Sumner, T. (2021). Using transformers to provide teachers with personalized feedback on their classroom discourse: The talkmoves application. *arXiv Preprint arXiv*:210507949.

Teo, P. (2019). Teaching for the 21st century: A case for dialogic pedagogy. *Learning Culture and Social Interaction*, *21*, 170–178.

Tao, Y., & Chen, G. (2023). Coding schemes and analytic indicators for dialogic teaching: A systematic review of the literature. *Learning, Culture and Social Interaction, 39*, 100702.

Tong, Y., & Ding, Y. (2024). Productive classroom dialogue and its association with student achievement in knowledge-building environments. *Language and Education*, 1–20. https://doi.org/10.1080/09500782.2024.2323207

Wallach, M. N., & Kontorovich, I. (2024). When Learner-centered teaching and learning goes online: Zooming into Linear Algebra tutorials during the pandemic. *International Journal of Research in Undergraduate Mathematics Education*, 1–21.

Wang, D., Tao, Y., & Chen, G. (2023). The Talk Move Patterns of Teacher-Student Interactions in K-12 Mathematics Lessons: An Analysis Based on the TalkMoves Dataset. In *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023* (pp. 1811–1812). International Society of the Learning Sciences.

Wang, D., Shan, D., Zheng, Y., Chen, G. (2023). Teacher Talk Moves in K12 Mathematics Lessons: Automatic Identification, Prediction Explanation, and Characteristic Exploration. In Wang, N., Rebolledo-Mendez, G., Matsuda, N., Santos, O.C., Dimitrova, V. (Eds.), *Artificial Intelligence in Education. AIED 2023*. *Lecture Notes in Computer Science* (pp. 651–664). Springer, Cham. https://doi.org/10.1007/978-3-031-36272-9_53

Wang, D., Tao, Y., & Chen, G. (2024). Artificial intelligence in classroom discourse: A systematic review of the past decade. *International Journal of Educational Research*, *123*, 102275.

Yusuf, N. (2021). The effect of online tutoring applications on student learning outcomes during the covid-19 pandemic. *Italienisch*, *11*(2), 81–88.

Zhang, L., Pan, M., Yu, S., Chen, L., & Zhang, J. (2023). Evaluation of a student-centered online one-to-one tutoring system. *Interactive Learning Environments, 31*(7), 4251–4269.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Deliang Wang[1] · **Dapeng Shan**[2] · **Ran Ju**[3] · **Ben Kao**[2] · **Chenwei Zhang**[1] · **Gaowei Chen**[1]

✉ Ben Kao
kao@cs.hku.hk

Deliang Wang

wdeliang@connect.hku.hk

Dapeng Shan
sdpwsb0927@gmail.com

Ran Ju
juran@connect.hku.hk

Chenwei Zhang
chwzhang@hku.hk

Gaowei Chen
gwchen@hku.hk

[1] Faculty of Education, The University of Hong Kong, Hong Kong, China

[2] Faculty of Engineering, The University of Hong Kong, Hong Kong, China

[3] Faculty of Architecture, The University of Hong Kong, Hong Kong, China