Teaching and Teacher Education 110 (2022) 103584

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

Research paper

Toward the automated analysis of teacher talk in secondary ELA classrooms☆

Meghan E. Dale [a, *], Amanda J. Godley [a], Sarah A. Capello [b], Patrick J. Donnelly [c], Sidney K. D'Mello [d], Sean P. Kelly [e]

[a] Department of Teaching, Learning and Leading, University of Pittsburgh, PA, USA
[b] School of Teacher Education and Leadership, Radford University, Radford, VA, USA
[c] School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA
[d] Institute of Cognitive Science & Department of Computer Science University of Colorado Boulder, Boulder, CO, USA
[e] Department of Educational Foundations, Organizations, and Policy, University of Pittsburgh, Pittsburgh, PA, USA

HIGHLIGHTS

- Automated Speech Technology can help researchers reliably detect the frequency of teacher talk features.
- Secondary English Language Arts (ELA) teacher talk varies widely but reflects dialogic instruction.
- Features of dialogic instruction in secondary ELA classrooms were co-related.

ARTICLE INFO

Article history:
Received 18 May 2020
Received in revised form
18 October 2021
Accepted 15 November 2021
Available online 4 December 2021

Keywords:
Classroom instruction
Literacy
Teacher education
Technology
Discourse
Automated speech recognition

ABSTRACT

Qualities of teacher talk strongly affect student learning, yet opportunities for teachers to receive feedback on their talk face time and cost limitations. To address this, we developed a semi-automated method of analyzing teacher talk using Automated Speech Recognition technology, compared it to traditional methods, and used it to analyze teacher talk features in 127 secondary English Language Arts lessons. We found the new method aligned with traditional methods, allowing us to reliably identify interrelated but distinct talk features. These features often reflected dialogic instruction. Our findings contribute to future studies of instructional discourse and technology-driven teacher learning opportunities.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Decades of research have shown that the characteristics and quality of classroom talk shape students' learning opportunities and understanding of disciplinary content (Applebee et al., 2003; Howe et al., 2019; Nystrand & Gamoran, 1991, 1997; Resnitskaya & Gregory, 2013). Important characteristics of classroom talk in the United States and elsewhere include the quality of teachers' questions (Alexander, 2008; Chisholm & Godley, 2011; Howe et al., 2019; Juzwik et al., 2013), responses to students' ideas (Cazden, 2001; Bloome et al., 2004), and students' expressions of multiple perspectives (Howe et al., 2019). Teacher talk, specifically, has been shown to be a strong indicator of instructional quality and a driver of student learning (Applebee et al., 2003; Danielson, 2011; Molinari & Mameli, 2013).

In response, many efforts to increase instructional quality in the past 20 years have focused on improving classroom talk through teacher professional development (Juzwik et al., 2013; Kucan, 2007; Stein & Matsumura, 2009; Sedova et al., 2016). These learning opportunities include individual coaching on classroom talk (Matsumura et al., 2012), self-reflection (Juzwik et al., 2013; Kucan, 2007), and expert feedback on observations of teaching

☆ This research was supported by a grant from the National Science Foundation (#1735785). Authors have no competing interests to declare.
* Corresponding author. University of Pittsburgh, 5500 Wesley W. Posvar Hall, 230 S. Bouquet Street, Pittsburgh, PA 15260, USA.
E-mail address: med78@pitt.edu (M.E. Dale).

https://doi.org/10.1016/j.tate.2021.103584
0742-051X/© 2021 Elsevier Ltd. All rights reserved.

(Danielson, 2011; Hill & Grossman, 2013). Numerous studies have shown that teachers gain insight into and improve their instruction through these activities (Juzwik et al., 2013; Kucan, 2007; Sedova et al., 2016).

However, improving classroom talk can be challenging. As Hill and Grossman (2013) note, the "grain size" of information shared with teachers for feedback and improvement of classroom discussions must be extremely detailed to be beneficial (p. 375). Another challenge is that opportunities to engage in a critical examination or reflection of one's classroom talk may occur infrequently for teachers because of time and cost (Juzwik et al., 2012). To be effective, teachers' reflections on their classroom talk must be timely, frequent, facilitated by an expert, and grounded in transcripts or audio/videorecordings (Sedova et al., 2016). However, such detailed documentation and analyses of teacher talk is both time-consuming and expensive. Transcription of a single 45-min class can take human transcribers from two to 4 h, and coding a transcript for effective features of teacher talk can take hours more. These constraints often mean that teachers rarely have the opportunity to see transcribed examples of or receive information about their classroom talk in a timely manner.

To overcome the limitations of time-consuming and expensive analyses of teacher talk, recent research efforts have drawn on advances in technology, such as automatic speech recognition, machine learning methods, and natural language processing to train computers to record and analyze teacher discourse in real-world classrooms (Clarke et al., 2015; Gerritsen, 2018; Kelly et al., 2018). Automatic Speech Recognition (ASR) refers to computer systems that transcribe human speech, such as the systems that run Amazon's Alexa and Google's Home. Machine Learning is a branch of Artificial Intelligence (AI) that allows computerized systems to learn by identifying patterns in data. In turn, Natural Language Processing (NLP) is a branch of machine learning that uses computer systems to detect, categorize, and analyze features of natural human speech, either spoken or written. Recent advances in NLP have made it possible to capture complexities of natural human language, such as speakers' attitudes and coherence of topics, and to detect these patterns in group conversations (Lugini et al., 2018; Rahimi et al., 2015; Rosé & Tovares, 2015).

Such advances in technology have the potential to allow teachers to receive information about their classroom talk quickly and frequently. Technological advances have helped scholars build computer-driven systems that can detect significant features of teacher talk with accuracy and convey information to teachers quickly and clearly (Gerritsen, 2018; Jensen et al., 2020, 2021, pp. 302–312; Wang et al., 2014). However, these technological advances are just beginning to be applied to the field of K-12 literacy teaching and learning.

In the present study, we used a semi-automated approach that combined ASR for audio transcription combined with human coding of the transcripts to investigate several features of high-quality teacher talk in the classrooms of 16 English Language Arts (ELA) teachers in the United States. The purpose of our study was twofold: (1) to investigate how semi-automatic methods compared with traditional methods of analyzing features of secondary ELA teachers' classroom talk, and (2) to provide a snapshot of the frequency of important features of teacher talk in secondary ELA classrooms (e.g., Applebee et al., 2003; Nystrand & Gamoran, 1991, 1997). Additionally, given growing research on the importance of classroom talk in shaping student outcomes and the inclusion of "quality of teacher talk" as a measure in classroom observation frameworks (e.g., Danielson's framework), we believe that an updated investigation of teachers' classroom talk—one that specifically builds on pivotal studies of ELA classroom discourse (e.g., Nystrand & Gamoran, 1997)—is timely and relevant to current scholarly debates about classroom talk. Our study was guided by three research questions:

1. How can computer-based Automated Speech Recognition (ASR) be used to analyze the frequency of teacher talk features?
2. What is the frequency of key features of teacher talk across middle and high school ELA classrooms?
3. What is the relationship between features of teachers' instructional talk?

Like other researchers, our aim is to contribute to both methodological and empirical discussions about teacher talk (Hennessy et al., 2016; Lefstein et al., 2015). Scholarly discussions about how frequent certain classroom talk features should be are necessarily influenced by the ways in which the discourse data is segmented and coded. Our perspective is that methodological choices (e.g., segmentation, coding choices, investigating specific terms) and empirical outcomes cannot be disentangled (Hennessy et al., 2020; Lefstein et al., 2015) and that methodological, theoretical, and coding decision-making matters for the analysis of classroom discourse (Song et al., 2020). We aim to contribute to both methodological and empirical scholarly conversations. In this paper, we offer a transparent discussion about our methodological coding choices, a comparison between our coding framework and more traditional frameworks, and an analysis of the prevalence rates of key teacher talk features using our methods.

1.1. Effective teacher talk

Our project is grounded in sociocultural theories of learning, which suggest that language and social interaction play a central role in the learning process (Muhonen et al., 2018; Vygotsky, 1978; Wertsch, 1991). Within classrooms, sociocultural theory posits that learning and development are supported by talk that includes open-ended questions and discussions, disciplinary-based ways of talking, multiple perspectives and elaborating ideas to others (Chisholm & Godley, 2011; Greeno, 2015; Lee, 2006). Effective teaching is thus centered more on facilitating dialogue and the exchange of ideas than on transmitting or lecturing about an established body of knowledge. Classroom discourse that is characterized by open-ended disciplinary questions, authentic sharing of ideas, and the collaborative construction of knowledge is often referred to as *dialogic*, emphasizing the reciprocal dialogue between teachers and students (Alexander, 2008; Nystrand & Gamoran, 1997; Resnick & Schantz, 2015). Empirical research conducted in high schools supports this theory; instructional time spent on student-centered discussions of open-ended disciplinary questions have been linked to gains in students' literacy skills, disciplinary knowledge and engagement (Applebee et al., 2003; Chisholm & Godley, 2011; Nystrand, 2006; Wilson, 2013).

Specific features of teacher talk have been found to positively affect student literacy learning in English classrooms. Authentic questions and uptake, or the incorporation of student ideas in teacher talk and feedback on student ideas, have been shown to enhance students' learning opportunities and are often considered key features of teacher talk in dialogic instruction (Chin, 2006; Nystrand et al., 2003). Other features of teacher talk that contribute to instructional quality include teachers' use of disciplinary terms and concepts (Duke et al., 2012; Grossman et al., 2013) and talk that reflects or requests "high cognitive demand" thinking, such as analyzing (Taylor et al., 2003; Wells & Meija-Arauz, 2006, p. 390). Finally, teachers' communication of specific and clear learning goals and procedures helps students better understand the learning activities they are engaged in (Grossman et al., 2013; Newmann et al., 1992; Shernoff et al., 2016).

1.1.1. Authentic questions

Teacher questions are ubiquitous in classrooms and thus the focus of many studies of productive instructional talk. Boyd (2015) calls questions the "discursive move of choice" for teachers (p. 372). However, questions differ in the learning opportunities they provide to students. Authentic questions, or inquiries that do not have prespecified answers, have been shown to promote student learning in ELA (Juzwik et al., 2013; Nystrand & Gamoran, 1997). Authentic questions allow for more student talk, which is beneficial for all students (Boyd, 2015). Though studies have found that teachers use fewer authentic questions in lower tracked classrooms, across academic tracks, the prevalence of authentic questions is positively related to student verbal participation (van de Pol et al., 2017) as well as student learning (Applebee et al., 2003; Nystrand & Gamoran, 1997).

1.1.2. Teacher feedback

Studies of teacher discourse often investigate teacher feedback within the IRF (teacher initiation/question, student response, teacher feedback) structure. The "F" portion of this sequence refers to teacher talk that may evaluate student responses or repeat, expand, or question student ideas (Sadler, 2013). Studies have shown that teacher feedback supports student learning, especially when it is specific or elaborated (Grossman et al., 2013). Chin (2006) found that teacher feedback was most effective when teachers provided elaborated responses, expanding on students' responses, either with further inquiry or by challenging student thinking. Chin also found that more neutral teacher feedback without a positive or negative valence elicited more high cognitive talk from students. In other studies of dialogic literacy classrooms, the absence of evaluation in teacher feedback encouraged students to expand upon their own thinking (Auckerman, 2007; Howe & Abedin, 2013) and engage in participation (Kelly, 2007).

1.1.3. Uptake of student ideas

Another key feature of productive teacher talk is uptake, or the incorporation of students' ideas into a subsequent line of questioning. Teacher uptake has a positive relationship with student literacy learning (Applebee et al., 2003; Nystrand & Gamoran, 1991) and encourages students to listen to and build upon one another's responses by emphasizing coherence between speakers and ideas (Nystrand et al., 2003). Uptake also encourages students to expand on their own ideas and provide evidence of their perspective (Soter et al., 2008).

1.1.4. High cognitive demand

High cognitive demand, or the promotion of higher order student thinking, is another feature of teacher talk that encourages students to move beyond memorization or procedural understandings to more analytical thinking. In Grossman's observation protocol for instructional quality in secondary English Language Arts classrooms (PLATO), the intellectual demand of questions and tasks are one component of effective teaching practices (Grossman et al., 2013). Taylor et al. (2003) also found that teacher use of cognitively demanding questions was correlated with student reading comprehension and literacy growth. Further, teacher questions that "press" students to explain or expand upon their thinking is one form of high cognitive questioning that is particularly effective for student learning (Witherspoon et al., 2016).

1.1.5. Subject-specific language

Teachers' use of subject-specific language, or vocabulary and other linguistic representations associated with specific academic disciplines, is often viewed as a facet of instructional quality and quality of classroom talk (Ernst-Slavit & Mason, 2011; Grossman et al., 2013; Witherspoon et al., 2016). In Resnick and her colleagues' (2015) concept of "Accountable Talk," teachers and students are not only accountable to each other but also to their representations of the discipline through their discourse. Similarly, Schleppegrell (2004) documents the specific language patterns that are used to build knowledge in disciplines and argues that teachers must explicitly teach "the language of schooling" as part of content instruction.

1.1.6. Clear goals and procedures

Lastly, a long line of research has shown that students learn better when the specific and broader purposes of instruction are made clear and when strategies or processes for engaging in academic work are made explicit (Smith & Feathers, 1983). Grossman et al. (2013) found that secondary ELA teachers' ability to clearly convey the goals of tasks and provide "explicit strategy instruction" were strongly correlated with other measures of teacher quality (p. 459).

Our study built upon this line of research on effective instructional discourse in order to develop a computer-based system to analyze features of teacher talk and to document the frequency of these features across multiple secondary English Language Arts classrooms.

2. Methodology and methods

2.1. Participants

Classroom audio data was collected from the classrooms of 16 teachers in two school districts in Western Pennsylvania (see Table 1): Elm School District and Pine School District (pseudonyms). All participating teachers identified as non-Hispanic and white; eleven identified as female and five as male. Participating teachers averaged 14.6 years of experience (the national average is 14 years [NEA, 2016]), and ranged from one to 27 years of teaching experience, including four teachers in their first ten years of teaching.

Following IRB approval and school district permission, all secondary ELA teachers in the districts were contacted via email and were provided with details about the study, compensation, and a consent form. Participating teachers, students and the schools they taught in were all given pseudonyms. None of the teacher information or data collected was shared with teachers' supervisors or administrators. On average, teachers completed their lesson recordings in 17 days (ranging from 6 to 57 days), depending on scheduling and familiarity with the technology.

2.2. School and classroom context

Data were collected in three schools, all serving predominantly white and middle/upper-middle class students. Elm High School and Elm Middle School were located in the same district. Approximately 7% of students qualified for free and reduced lunch and fewer than 1% were English Language Learners. At Pine High School, 12% of students qualified for free and reduced lunch and less than 1% were English Language Learners.

Information about individual classrooms were reported by participating teachers. The dataset included four 7th grade classrooms, four 8th grade classrooms, seven 9th grade classrooms, six 10th grade classrooms, seven 11th grade classrooms, and four 12th grade classrooms. Classrooms had a mean size of 22 students (SD = 4.8). Two classrooms were reported as low-track, 14 as general track, and 16 as high-track classrooms. Consistent with the demographics of the participating districts, teachers reported that

Table 1
Participant information.

Teacher	Gender Identity	Years of Teaching Experience	School	Grades Taught	Number of Lessons Recorded
Michael	M	27	Elm High School	12	9
Corinne	F	24	Elm High School	11	9
Anne	F	7	Elm High School	9, 12	5
Lisa	F	11	Elm High School	11, 12	7
Derrick	M	1	Elm High School	10	8
Elizabeth	F	18	Pine High School	9	9
Claire	F	12	Pine High School	11	9
Morgan	F	20	Pine High School	10	8
Daniel	M	12	Pine High School	9	8
Charles	M	11	Pine High School	11	8
Catherine	F	16	Pine High School	9	8
Quinn	F	12	Pine High School	10	8
John	M	9	Elm Middle School	8	9
Olivia	F	9	Elm Middle School	8	7
Lindsey	F	19	Elm Middle School	7	6
Alexandra	F	26	Elm Middle School	7	9

most students in their classrooms were white (an average of 93%), with 3.8% Asian (primarily reflecting two classrooms in which 29% and 45% of students were Asian), 2.5% Black (reflecting one classroom in which 22% of students were Black), and less than 1% other race/ethnicity categories.

Teachers also reported the literacy skills taught in each recorded lesson. Teachers reported that 62% of lessons included reading instruction (both using literary and informational texts), 26% contained writing instruction (both creative and expository), 10% contained grammar instruction, 5% included vocabulary instruction, and 6.9% included test review. Most teachers reported teaching a range of literacy skills in the recorded lessons. Approximately 69% of lessons contained small group work or paired activity, 57% contained teacher lecture, 56% contained whole-class discussion, 41% contained individual student activity, 13% contained a dramatic activity, and 6% contained a student-led activity. Note that most lessons included more than one focal literacy skill and multiple activity structures/groupings.

2.3. Data collection

Teachers were first trained by research assistants to independently record their own classroom talk using a provided laptop computer, a Samson AirLine 777 headset with a high-quality microphone, and the audio-recording program Audacity. The recording setup was validated in our previous work (Jensen et al., 2020; Kelly et al., 2018) with respect to producing high-quality teacher audio that can be automatically transcribed. We only focused on recording teacher speech, not student speech, because microphone and ASR technologies have not yet advanced to the point where it is possible to simultaneously record and transcribe multiparty speech with high accuracy without mic'ing individual students.

Teachers were asked to record two different classes four times each for a total of eight recordings per teacher, though a few teachers contributed more recordings. Teachers were asked to record their entire class period, meaning our dataset features a variety of instructional activities including teacher lecture, student-centered discussions, teacher-led question and answer sessions, and independent work. Multiple recordings per classroom allowed for opportunities to investigate teachers' use of talk features across multiple lessons, instructional activities, topics, and texts and allowed for increased reliability in examining each talk feature. Only recordings with high quality audio were retained, leaving us with 127 recordings. Audio files were saved and automatically uploaded to a shared folder with researcher access.

2.4. Data analysis

2.4.1. Automatic speech recognition and segmentation

Classroom audio recordings were automatically transcribed using the IBM Watson automatic speech recognition (ASR) web-based software. To evaluate the quality of the transcriptions, we manually transcribed a random sample of 20 sequential segments per transcript (2 540) and compared them to the automated transcription. The Watson speech recognizer was 72% accurate across all segments of teacher talk. Because we recorded conversational speech in noisy classroom environments, we anticipated that the ASR would not be 100% accurate. For comparison, Blanchard et al. (2015) compared how other ASR technologies accurately transcribed teacher questions in noisy classroom environments. In their study, technologies like Google Speech and Bing Speech had the highest mean accuracy rates of 0.56 and 0.52 respectively. Most speech recognition errors in our ASR-produced transcripts were considered insertion or deletion errors, meaning the ASR tended to miss words or add words to "complete the context," rather than substitution errors. To account for these errors, coders listened to the classroom audio as they coded and did not exclusively rely on the ASR-created transcripts.

We next segmented the transcription into meaningful and logical units of talk. Segmentation criteria are needed in any analysis of spoken language to define units of analysis because speech contains no punctuation that demarcates idea units. Some research on teachers' spoken language uses turns at talk or episodes as units of analysis; however, because the features of teacher talk we examined were micro-level, numerous and overlapping, we needed to define a smaller unit of analysis for spoken discourse for accurate coding and for exploring the relationship between the talk features. We also sought a unit of analysis that could be automated with ASR or some other technology without human coding.

We determined that the best unit of analysis was a segment of teacher speech separated by a 1 s (or more) pause. We explored other ASR segmentation times using 0.5, 1.0, 1.5, and 2.0 s pauses in teacher speech. Pauses shorter than 1.0 s, like the 0.5 s pause, resulted in segmentation that was broken into incomplete clauses (e.g., "you want", "it's not so much about", "so I want to"). Longer pauses, like the 2.0 s pause, resulted in long segments which included multiple features of talk that were difficult to code accurately. After testing various points of segmentation, we found that 1.0 s pauses in teacher speech yielded segments of teacher talk that averaged 22 words in length (SD = 44) and more often represented a single complete statement or question. Table 2 presents a comparison between more traditional coding segmentation (in this

case, the turn at talk) and our segmentation (a 1.0 s or longer pause in teacher speech). The example is pulled directly from an ASR-developed transcript and coders did not edit teacher speech for clarity. The comparison in Table 2 demonstrates that the commonly used unit of analysis the turn-at-talk can yield long segments of teacher talk that contain multiple significant features, such as questions and ELA terminology (two features that are investigated in this study). Segmenting teacher talk at the 1.0 s pause, in contrast, allowed us to investigate and count each significant feature of teacher talk more precisely and at a finer level of detail.

Overall, our dataset consisted of 127 audiorecordings containing 35 142 segments. Given the size of our dataset, we randomly selected 200 sequential segments of teacher talk from each audiorecording for coding. Because some of the transcripts included fewer than 200 segments of teacher talk, our coded dataset included a total of 16 977 segments (an average of 134 segments per transcript) averaging approximately 30 min in duration (ranging from 17 to 57 min).

2.4.2. Codebook and coding

Coding was done by two graduate student research assistants and two undergraduate students supervised by faculty. Each student coded approximately 25% of the data and all four met weekly to discuss and resolve coding differences.

The codebook (Table 3) was developed by the research team and utilized classroom talk features from the existing literature associated with dialogic instruction (e.g., authentic questions, and uptake) (Nystrand & Gamoran, 1997) as well as other teacher talk features associated with student learning and engagement, such as subject-specific terminology. The codebook was revised and improved through a recursive process as coders tested the codebook on prior data and worked towards improving inter-rater reliability. In addition to the range of discourse features we studied, our semi-automated coding process differed from previous episode or utterance-level studies of discourse (e.g. Applebee et al., 2003) in a few ways: (1) all instructional time, not just text-based discussions, was eligible for sampling and coding and thus all teacher speech (both statements and questions) was coded, (2) teacher talk was transcribed using ASR rather than human transcription, and (3) the unit of analysis was the segment of teacher speech (defined by 1.0 s pauses in speech) rather than turn-at-talk, question episode, activity or class session.

The first step in coding process was to exclude segments of talk that were not teacher speech or were short, incomplete "fragments" of teacher speech that could not be coded such as "so." After identifying complete segments of teacher talk, coders then determined whether a segment of teacher speech was a question or statement. Questions were defined as "requests for information," whereas statements were anything else. Given the importance of questions in scholarship on teacher talk, segments that contained both a statement and a question were coded as questions. Segments of teacher talk were then labeled as instructional or non-instructional. Instructional statements or questions related to the lesson and its learning goals while non-instructional talk was irrelevant to the lesson (for example, utterances related to student behavior such as, "Please sit down").

Instructional segments of teacher talk were then further coded for more detailed features of teacher talk. We coded for two aspects of teacher feedback: valence (positive, negative or neutral) and elaboration of feedback (elaborated or unelaborated). Elaborated feedback occurred when the teacher offered a rationale for a student's answer and showed more than cursory consideration of the student response. Unelaborated responses, in contrast, simply provided feedback (e.g., "that's exactly right") and moved onto a subsequent topic. Teacher feedback was coded as either positive (e.g., "that's correct"), negative (e.g., "that's incorrect"), or neutral (e.g., "I wonder about that").

Segments of teacher talk were also coded for uptake, cognitive level, explicit learning goals and procedures, and the presence of ELA terms. We only considered teacher talk to be uptake if it incorporated a student's idea within teacher talk. The simple revoicing of a student response was not considered uptake. For cognitive demand, we used Webb's (2006) Depth of Knowledge (DOK) chart to code segments of teacher talk as low or high. Any question or statement that encouraged students to engage in thinking tasks in sector 1 of Webb's chart were considered low cognitive (e.g., recall) and those within sectors 2 through 4 were considered high cognitive (e.g., predicting, analyzing). The specificity with which teachers explained the goals of the lesson was coded as high, medium, low or none. Low goal specificity included vague lesson goals and procedures, medium goal specificity provided specifics regarding materials or steps, and high goal specificity included goals, procedures, and a rationale for lesson activities. The presence of ELA terminology in each segment of teacher talk was coded simply as "present" or "not present" in each segment. The research team generated the list of ELA terms in the codebook from existing scholarship and curricular materials (Baumann & Graves, 2010; National Governors Association & Council of Chief State School Officer, 2010).

Some codes were only applied to teacher questions. Authentic questions, as noted in our literature review, do not have a pre-specified answer and position students as having knowledge the teacher does not. On the other hand, questions were coded as inauthentic if they had a "right" or expected answer that the

Table 2
Segmentation example.

Turn at Talk	Time Stamp (Mins and Secs)	1.0 Second Pause in Speech
it okay so you see who he is on the moment of crisis taking advantage of people and so forth can't Santonio Holmes said that Walmart okay so obviously in the book we're going to see lots of characters in crisis and you're going to see I think who they really are like their characters will develop further when you see how they handle a crisis okay flip to the back/duty versus desire when if ever should desire be placed before duty what may be the result of such a decision who do we expect to place duty before desire and what happens if they do not so let's start with the first question/and take a few minutes to jot down your ideas first when if ever if ever should desire be placed before duty so desire meaning like something that you want/over a sense of duty	24:57	it okay so you see who he is on the moment of crisis taking advantage of people and so forth can't Santonio Holmes said that Walmart okay so obviously in the book we're going to see lots of characters in crisis and you're going to see I think who they really are like their characters will develop further when you see how they handle a crisis okay flip to the back
	25:23	duty versus desire when if ever should desire be placed before duty what may be the result of such a decision who do we expect to place duty before desire and what happens if they do not so let's start with the first question
	25:43	and take a few minutes to jot down your ideas first when if ever if ever should desire be placed before duty so desire meaning like something that you want
	25:56	over a sense of duty

Table 3
Coding categories.

Code	Subcodes	Definition	Example
Authentic Question		An authentic question is an open-ended question for which the teacher does not have a pre-scripted answer.	(Authentic) What was your reaction to the end of the story?
Feedback	Elaborated or Unelaborated; Positive, Negative or Unclear.	Expression of elaborated consideration or judgment of correctness of a student's response.	(Positive, Elaborated) That's right. You're dying with each breath, and this is what the poet tries to bring to the consciousness of the beloved.
Uptake	Yes, No, or Faux	Uptake is the teacher's incorporation of ideas from a student utterance into a subsequent statement or question.	(Yes) You think he can't get help, can you expand on that?
ELA Terms	Yes or No	The use of English Language Arts disciplinary terms in teacher talk.	(Yes) Ensure that you include a topic sentence in each one of your paragraphs.
Cognitive Demand	High or Low	High cognitive demand talk emphasizes analysis (e.g. compare, interpret, synthesize, etc.). Low cognitive demand talk emphasizes reports or recitation of facts (e.g., define, recall, identify).	(Low cognitive level) A simile is a metaphor that uses "like" or "as."
Goal Specificity	None, Low, Medium, or High	Goal specificity refers to the extent to which the teacher explains the process and end goals of a particular activity. Low specificity gives some details of the learning, Medium specificity provides more concrete details of the learning goals and processes, and High specificity explains details of the learning and processes as well as reasoning for why students are engaging in the learning.	(Medium) Your writing partner should give you three overall comments, before any minor editing stuff.

teacher already knew. Teacher questions that were repeated explicitly or implicitly multiple times were coded as serial questions. Serial questions were given the same codes as the initial question. For example, if the original question was coded as "authentic" and "high cognitive," then all serial questions, despite their syntactic structure, would also be coded as "authentic" and "high cognitive."

We double-coded 13 transcripts to determine inter-rater reliability (see Table 4). Simple percent agreement reflected high interrater agreements on all features of teacher talk at the segment level (78.9–97.4%). We also assessed the inter-rater reliability of each of these features of teacher using Gwet's (2008) AC1 statistic at the segment level and correlation coefficients at the transcript level (e.g., proportion of authentic questions in the lesson). For some features, such as uptake and negative teacher feedback, low prevalence rates along with peaky or skewed distributions contributed to lower inter-rater reliability.

3. Results

In this section, we first compare our semi-automated coding method with traditional coding methods for analyzing teacher talk. Second, we present the distribution of features of ELA teachers' talk as detected by the semi-automated method we employed. Third, we analyze inter-relationships between the features of teacher talk.

3.1. Semi-automated analyses of teacher talk

Given our long-term goal of developing automated methods of analyzing teacher talk, we first wanted to determine how well our semi-automated methods aligned with the traditional methods for coding ELA teacher talk found in related research. Our semi-automated method of coding differed from prior studies primarily in its unit of analysis, a segment of teacher speech separated by pauses of 1.0 s or longer. This unit of analysis could potentially yield different prevalence rates for specific features of teacher talk from prevalence rates reported in existing research of secondary ELA teacher talk (e.g., Applebee et al., 2003), given that previous studies have utilized various units of analysis, such as the turn at talk, question, episode, activity, or lesson.

Thus, to compare our semi-automated coding method with existing methods for coding teacher talk, we segmented and coded a randomly-selected subset of five transcripts of class sessions (1000 teacher talk segments) using both methods and compared the prevalence of each teacher talk feature (Table 5). Note that we used the same codes and definitions for both coding approaches. To implement traditional coding, we used the unit of analysis most frequent in existing research on each teacher talk feature. Authentic questions, cognitive level, uptake, and teacher feedback have typically been coded at the unit of teacher questions (Applebee et al., 2003) or in relation to teacher questions (Wells & Meija-Arauz, 2006). For these particular talk features, our semi-automated method adapted well to comparisons with traditional methods because we were able to compare similar units (i.e., teacher questions). Other teacher talk features were more difficult to compare across coding methods because prior studies either used much larger-grained units of analysis (e.g., the activity or whole class) or used a variety of units of analysis.

Overall, our semi-automated method and traditional methods tended to result in similar prevalence rates and high correlations for most qualities of teacher questions. For authentic questions, we found a high correlation (0.979) between semi-automated and traditional methods of coding and similar prevalence rates: 33% and 36% respectively (as a percentage of all questions). Coding for high cognitive level teacher questions using both methods yielded somewhat different prevalence rates (46% for semi-automated coding, 36% for traditional coding) and slightly lower correlations (0.76). For teacher uptake of student ideas, semi-automated coding resulted in a prevalence rate of 17% and traditional methods resulted in a rate of 32%. Despite this difference in prevalence rates, results of both coding methods were highly correlated (0.984). Although coding for the presence of teacher feedback yielded quite different prevalence rates and correlation was lower (0.63), the two methods of coding resulted in similar prevalence rates and moderate to strong correlations for specific features of teacher feedback, namely elaboration (prevalence rates of 26% and 30%, respectively, correlation 0.912) and positive feedback (prevalence rates of 94.3% and 89.9%, respectively, correlation 0.72).

For some teacher talk features, traditional methods of analysis were so large-grained or disparate that it was hard to find a traditional analytical method to adequately compare to our new method. Still, in many cases, the results of the two methods were highly correlated even when the prevalence rates differed. For example, the prevalence rates of instructional/non-instructional teacher talk have traditionally been calculated in minutes and seconds, or the proportion of class time that focuses on instruction rather than procedural talk (e.g., taking attendance, talking to students about missing work) (Fisher, 2009). In comparison, our

Table 4
Measures of inter-rater agreement.

	Utterance-level reliability statistics		Observation-level reliability
	Simple percent agreement (%)	Gwet's AC1[a]	Pearson correlation
Instructional utterance	90.6	.855	.731
Disciplinary Utterance	82.8	.661	.641
Authentic questions	82.6	.722	.743
Serial questions	79.9	.664	.207
Feedback present	85.2	.770	.666
Elaborated feedback	97.4	.966	.375
Negative valence	78.9	.690	.107
Uptake	95.4	.951	.394
ELA terms present	91.6	.887	.832
High cognitive demand	85.0	.802	.213
High/Medium goal specificity[b]	84.1	.776	.611

[a] The AC1 provides an alternative measure of inter-rater reliability when prevalence rates approach 0, 1 and corresponding percent agreements are high (See Gwet, 2008).
[b] Ratio of T/S was not coded during IRR study.

semi-automated method calculated prevalence rates by calculating the proportion of segments that were non-instructional. The prevalence rates for teachers' non-instructional talk resulting from semi-automated and traditional methods were highly correlated (0.897), even though the prevalence rates differed (23% and 10%, respectively).

For teachers' use of ELA terminology, it was difficult to compare our coding method with a single traditional method. In some studies, teachers' use of subject-specific terminology was assessed using a numerical scale (e.g., 1–3) to represent the quality for each 5-min segment of a transcription (Hill et al., 2008), and in other studies was calculated by lexical density, that is, the ratio of subject-specific words to all words in a transcript (Halliday & Martin, 1993). We chose to compare our semi-automated method with the method that was most similar to ours – lexical density. A comparison between our semi-automated method and a traditional method yielded quite different prevalence rates (3% for traditional coding as compared with 31% using the semi-automated method) but a high correlation (0.914).

Comparing methods for analyzing the quality of teachers' explanations of goals and procedures was the most challenging since traditional methods typically code such explanations at the learning activity or lesson level and report overall quality using numerical scales rather than prevalence rates (Grossman et al., 2013). For traditional coding, we chose to use the activity as the unit of analysis because it was a more fine-grained unit than the lesson. However, even using this unit, across the five transcripts only 22 learning activities were identified for coding, a small number for calculating prevalence rates and correlations. For both semi-automated and traditional approaches, we coded each instance of goal and procedure explanation as high, medium or low specificity. We found the same prevalence rates (79%) for high and medium specificity teacher explanations using both methods but almost no correlation (0.210) between the prevalence rates.

Overall, our semi-automated, ASR-based coding approach aligned considerably with traditional human transcription and coding for most teacher talk features. The two coding methods resulted in similar prevalence rates with high correlations at the transcript level for authentic questions, high cognitive questions, elaborated feedback and positive feedback. For other teacher talk features – uptake, ELA terms, and non-instructional talk – prevalence rates differed but correlations between the two methods were high. The high correlations between the two methods suggests that even when prevalence rates differed, both methods resulted in similar measurements but just on different scales. In other words, much like measuring and comparing temperature in Fahrenheit or Celsius, there was a consistent relationship between prevalence rates that resulted from both methods and their strong correlations suggest that they are measuring the same features of teacher talk.

3.2. Prevalence of teacher talk features

We next report on the results of our human coding of the ASR-transcribed data that we collected in 127 secondary ELA classes (Table 6). These findings offer a snapshot of ELA teachers' instructional talk in current U.S. classrooms and a point of comparison for similar studies conducted in the past. As in the section above,

Table 5
Comparison of prevalence rates and correlations in semi-automated coding and traditional coding.

	Semi-automated coding		Traditional coding		Correlation
	Mean (SD)	X/?[a]	Mean (SD)	X/?	
Authentic questions	.33 (.28)	# of ques	.36 (.37)	# of ques	.979
High cognitive questions	.46 (.25)	# of ques	.36 (.29)	# of ques	.758
Uptake	.17 (.25)	# of ques	.32 (.59)	# of ques	.984
Feedback	.14 (.10)	# of ques	.54 (.31)	# of ques	.630
Positive feedback	.943(.065)	# of utt w/feedback	.898 (.063)	total feedback	.720[c]
Elaborated feedback	.26 (.25)	# of utt w/feedback	.30 (.16)	total feedback	.912
Med/high goal specificity	.79 (.10)	# of utt with any goal spec.	.79 (.30)	# of activities	.210
ELA Terms	.31 (.18)	# of all segments	.03 (.03)	# of teacher words	.914
Frequency of non-instructional talk	.23 (.18)	# of all segments	.10 (.13)	Total teacher talk time[b]	.897

[a] X/? displays the relevant denominator for variable calculation.
[b] Time in seconds of teacher talk (isolated/occasional student utterances removed).
[c] Insufficient variation.

prevalence rates are reported as a percentage of segments, or teacher talk bounded by 1.0 s or longer pauses. The examples of teacher talk included below (punctuated and edited for readability) were pulled from randomly selected teacher classrooms and therefore represent a variety of English Language Arts topics and instructional activities. Unlike similar studies of teacher and classroom talk, our dataset is not made up exclusively of text-based discussions.

Across the dataset, teachers' speech was comprised of 64.6% statements, 31.7% questions, and 3.7% fragments. Most teacher talk was instructional (84%), including 95% of questions.

3.2.1. Teacher questions

We first examined various features of teacher questions that have been associated with instructional quality, such as authenticity and uptake (Nystrand & Gamoran, 1997). In our dataset, 26% of teacher questions were coded as authentic. In one high school discussion of Conrad's (2006) *Heart of Darkness*, the teacher, Charles, asked the class about their perceptions of Kurtz (a central character) and his personality and motivations:

Charles: What did you make of that? What are we supposed to take from that? Or, what did you?

[Student speaking]

Charles: Thanks, Jane.

[Student speaking]

Charles: Right.

[Student speaking]

Charles: Yeah, what's that about? Is he just—has he become— Is this like a game to him, where it's just like, collect all the ivory you can? And—he's kind of lost the ability to communicate on a human level? Or with the outside world? Or is it—what did you make of that?

Charles asked authentic questions (e.g., "What did you make of that?" "What's that about?) numerous times. These questions had no pre-specified answer and encouraged students to share their original ideas.

Conversely, 74% of teachers' questions were not authentic or had a known (often single) answer. For example, John, a middle school teacher, asked students the known-answer question during a lesson about sentence combining, "What is the relationship between these two sentences?" Although the prevalence of known-answer questions was high in our dataset and is not usually considered a desirable feature of teacher talk, this example suggests that asking non-authentic questions can encourage students' learning of some ELA topics and literacy skills, such as grammar and syntax.

Twenty-one percent of teacher questions in our dataset were coded as serial, meaning the same question was asked multiple times to different students. Of those, 35% were authentic, such as this stretch of Charles's talk during the same literary discussion of *Heart of Darkness* (Conrad, 2006) described above:

Charles: Because he's different from how everyone else operates?

[Student continues sharing]

Charles: Like, we're here to do a job, we're not here to make friends or to be part of the community, so to speak? Okay. Melissa? And then we'll come back.

[Student shares response]

Charles: So, you think from Kurtz's perspective it's almost Machiavellian and he was thinking, well one way to get in would be to attend these nightly dances and to kind of act very interested. So it's not coming out of an actual love of these people, but more it helps him. A means to an end. Interesting. Evan?

In the example above, Charles offered multiple students the chance to respond to the same authentic question ("Because he's different from how everyone else operates?"), posing serial questions by calling on individual students and organizing the order in which students were invited to share their thinking (e.g., "Okay. Melissa? And then we'll come back."). Through serial questions, Charles prompted multiple students to share their unique perspectives and interpretations on the same topic.

3.2.2. Teacher feedback and uptake

Classroom discourse research has also typically investigated teacher feedback on student utterances. In our dataset, teacher feedback appeared in 29% of teacher talk. The majority of teacher feedback was unelaborated (62%) and positive (86%) (e.g., "right" or "correct"). Elaborated teacher feedback included statements that expanded on students' responses or thinking. For example, in one high school classroom during a discussion of motifs related to Charles Dickens's (2020) *A Tale of Two Cities*, Morgan asked her students to discuss revenge and justice:

Morgan: Someone had a hand up over there. Yes, Hannah?

[Student talking]

Morgan: Okay, good. So, revenge is more emotion-filled where justice is—back to what Laura said—more fair. And as Ian said, maybe tied to the law. Okay, so then, if we can say that revenge is more a personal satisfaction, is revenge ever justified?

[Student talking]

Morgan: Okay. Luke?

[Student talking]

Morgan: So if you were the one seeking revenge, you're going to feel justified but to everybody else you won't? Okay, that's an interesting perspective.

Morgan's responses to students offered elaborated feedback by unpacking students' answers and providing both positive ("good") and neutral ("interesting perspective") valence. In the same transcript, Morgan also offers an example of unelaborated feedback. Later in this same discussion about themes and motifs, Morgan responded to a student and said, "All right, that's a good example." Here, Morgan positively responded to the student's idea, but offered no further elaboration. Although both forms of feedback (elaborated or unelaborated) indicate that the student's response is heard and, in some cases, confirmed, by the teacher, elaborated feedback encourages the whole class to fully consider the student's response and enriches student learning (Chin, 2006).

Morgan's talk also reflects uptake of students' ideas, a feature of teacher talk that supports student learning and is a key characteristic of dialogic classrooms (Applebee et al., 2003; Nystrand & Gamoran, 1997). Teacher uptake was very rare in our data (7% of total teacher talk and 8.7% of questions), with many class sessions containing no teacher uptake. However, some teachers engaged in moments of genuine uptake, such as in Morgan's second turn at talk above when she incorporates Laura's and Ian's comments to ask a follow up question to the class and her last turn in which she incorporates Luke's comment, calling it, "an interesting perspective."

3.2.3. Disciplinary language

Scholars posit that teacher use of subject-specific language can support students' mastery of the language and knowledge base of a specific discipline (Ernst-Slavit & Pratt, 2017; Townsend, 2015). In our data, 24% of teacher talk contained ELA terminology. For example, in her high school classroom, Elizabeth fielded individual questions from students as they worked on their introductions to an essay on *Romeo and Juliet* (Shakespeare, 2009). After one student's question, Elizabeth responded:

Elizabeth: It could be character and plot. Because if you're talking about what the characters do in the story, that's plot. But if you use what the characters do as a way of introducing them, then

Table 6
Observation-level prevalence rates teacher talk features in secondary ELA classrooms.

	Proportion (mean)	SD	Skew	Kurt	IQR	Min, Max
General Features						
Instructional talk	.841	.109	−1.27	4.96	.796–.914	.398, 1
Disciplinary talk	.692	.143	−.724	3.51	.617–.794	.250, .960
Questions (vs. statements)	.317	.113	−.236	2.51	.238–.395	.064, .547
Features of Disciplinary Talk						
Feedback present	.290	.159	.110	2.62	.168–.398	0, .696
Elaborated feedback[a]	.377	.215	.437	3.09	.235–.491	0, 1
Negative valence	.140	.177	2.29	9.59	0–.192	0, 1
Positive valence	.778	.204	−1.71	6.54	.696–.909	0, 1
Neutral valence	.082	.134	3.77	22.30	0–.1	0, .6
Uptake	.070	.123	2.37	7.88	0–.058	0, .55
ELA terms present	.240	.197	.601	2.64	.066–.385	0, .909
High cognitive demand	.189	.184	.950	3.23	.026–.303	0, .75
Features of Questions/Statements						
Authentic questions	.258	.326	1.13	2.88	0–.485	0, 1
Use of serial questions	.207	.174	.401	4.97	.029–.360	0, .643
Questions with feedback	.255	.178	.545	3.40	.125–.375	0, .783
Elaborated feedback[a]	.349	.281	.780	3.06	.176–.5	0, 1
Negative valence	.088	.172	3.35	16.73	0–.143	0, 1
Uptake in questions	.087	.170	2.90	12.27	0–.078	0, .8
High cog level ques.	.352	.307	.427	1.96	.044–.576	0, 1
High/medium goal specificity[b]	.156	.146	1.42	4.97	.054–.224	0, .667

[a] Feedback properties were only coded if feedback was present.
[b] Coded for statements only.

that's introduction and summary.

[Student speaking]

Elizabeth: An introduction of a character might be something like: Romeo and Juliet are the main characters in *Romeo and Juliet*.

[Student speaking]

Elizabeth: Yeah, unless you're talking about something that is a fact of the book—like something that happened or the situation.

In this example, Elizabeth uses multiple ELA terms such as, "character" and "plot" to illustrate the difference between summarizing the play's plot and introducing the characters. Overall, teachers' use of ELA terminology ranged considerably across the lessons, from 0 to 91% of segments including at least one ELA term.

3.2.4. High cognitive demand

We also analyzed evidence of high-level cognitive demand in teacher talk. In our dataset, high cognitive demand accounted for 18.9% of teacher talk. For example, Corinne posed both high and low cognitive questions during a whole-class discussion about the relationship between Christopher and his father, two characters in Haddon's (2003) *The Curious Incident of the Dog in the Nighttime*:

Corinne: Right [Christopher] thought [his dad] might kill him. Why does [Christopher] think that [his dad] might kill him?

[Student response]

Corinne: Because he killed the dog. Right. So, he's kind of logically, again, if my dad kills a dog that means that my dad could kill me. So, he's afraid of his dad. Good.

Corinne: Okay, so, predictions on what do you think is going to happen at this point in the book with Christopher and his dad. What do you think?

In this example, Corinne asked students to make predictions about both characters' future decisions, which we considered high cognitive questions since they are aligned with Level 2 of Webb's "Depths of Knowledge" chart (Webb, 2006). However, the majority of teacher talk in our data reflected Level 1 questions and statements, which required students to recall or memorize information (Webb, 2006), as illustrated in the example above when Corinne asked her students to simply recall the events of the novel by asking, "Why does [Christopher] think that [his dad] might kill him?" Although we coded this recall question as low-cognitive because it falls under Level 1 of Webb's chart, the question also prompts students to recall the plot in order to prepare for considering more analytical questions about the novel. We note that although high cognitive questions are generally associated with more robust opportunities to learn, low cognitive questions can serve important instructional purposes within the context of ELA teacher talk.

3.2.5. Explanation of goals and procedures

Finally, we looked at how ELA teachers explained the goals and procedures for student learning, drawing on research showing that a clear understanding of tasks, learning goals, and their importance supports student learning and engagement (Grossman et al., 2013; Smith & Feathers, 1983). We coded for three levels of goal specificity (low, medium, and high), or how detailed teachers' explanations of goals and procedures were. Highly specific explanations included details and reasoning for why the lesson's task was important or could benefit students' future learning. For example, in her lesson about argumentative essays about Shakespeare's (2009) *Romeo and Juliet*, Elizabeth introduced the day's writing assignment by explaining:

In your introduction, what you're doing is you're actually preparing your reader for what he or she is going to read about in the body of your essay. So now, your next assignment is going to be to do the outline for that essay. So, that's why we're reviewing what you do in an introduction.

Elizabeth detailed the resources and tasks her students were expected to engage in during class (i.e., writing an outline for the essay) and also explicitly stated how this learning could benefit students in their future writing endeavors (i.e., considering audience needs).

In contrast, explanations that were coded as medium specificity tended to offer some detail about the task but not its future significance or benefit. For instance, in a lesson on poetry and songs that represent the American experience, Claire explained, "Today we're going to look at another type of song. And Walt Whitman wrote a song—a poem—that he felt represented Americans during his time. This poem was written in 1860, right before the Civil War." In this example, Claire provided some specific details about the task

that students would engage in (e.g., reading a Walt Whitman poem), but did not specifically tell students why they were engaging in this task and what they were expected to learn.

In our dataset, 15.6% of teacher statements included explanations of learning goals and procedures that were coded as high or medium. Our findings are difficult to compare to prior research, even though communication of clear learning objectives and task directions are common expectations in most K-12 settings and teacher education programs, because very few studies report teacher explanations of learning goals and procedures as a percentage of teacher talk and instead report quality of explanations at the task or class level (Grossman et al., 2013). However, given the importance of clear explanations and goals in school settings, our findings could be used as a benchmark for future studies.

As a final note, we wish to point out that all the teacher discourse features analyzed here, even ones with particularly low or high prevalence rates, show notable variation across classes (with standard deviations above 0.1). In other words, across our dataset, teachers' talk varied considerably in the frequency with which each discourse feature was voiced. Some class-level variation was likely shaped by the activities and learning goals of that particular lesson. However, this variation also suggests that the teacher talk features we identified are sensitive to individual differences in instructional talk.

3.3. Relationships between teacher talk features

We also investigated interrelationships among the teacher talk features through analyzing correlations at the lesson level (Table 7). Given that multiple lessons were taught by the same teacher and thus each lesson was not truly independent, we corrected the standard errors for clustering in our analysis. With the exception of uptake, which was highly correlated with both authentic questions (0.674) and high cognitive level (0.617), none of the features were so highly correlated with each other to suggest they reflect a unitary dimension of discourse. A number of features associated with dialogic instruction – authentic questions, uptake, and high cognitive demand – were correlated. This reinforces prior research about dialogic classroom discourse, which has described how teacher uptake and authentic questions together reinforce the give-and-take of ideas, with teachers and students alike valuing and learning from students' ideas and providing students with more cognitively demand learning opportunities (Applebee et al., 2003; Juzwik et al., 2013). Authentic questions and the use of serial questions were also highly correlated, which suggests that teachers' serial questions may reinforce dialogic talk in ELA classrooms by explicitly requesting multiple points of view on the same topic or question.

On the contrary, use of ELA terms was negatively associated with the above four features of teacher talk. Although the negative correlations are modest, they are statistically significant in most cases. We hypothesize that ELA terms were used more frequently by teachers in lessons focused on topics such as grammar, writing, and vocabulary – topics that are often taught in traditional, teacher-centered approaches. It is also possible that when teachers focus on ELA concepts such as symbolism and characterization in literary discussions, these narrow learning goals cause teachers to be more likely to value correct responses and less likely to invite students' original ideas through authentic questions and uptake.

Negative feedback was also negatively associated with features of dialogic teacher talk such as authentic questions, uptake and high cognitive level talk, though this relationship was only statistically significant for high cognitive talk. This finding suggests that negative feedback rarely if ever supports high-cognitive, dialogic instruction (Kelly, 2007). Explanations of learning goals and procedures did not seem to have a strong relationship with any of the other talk features other than high cognitive demand. This feature of teacher talk warrants further investigation. Overall, our findings on the relationship between ELA teacher talk features reinforce prior research on dialogic classroom discourse but complicate prior research on the value of setting specific learning goals and using subject-specific vocabulary in instructional talk.

4. Discussion

In this study, we investigated a new semi-automatic method for analyzing secondary ELA teacher discourse with the goals of investigating how this method compared with traditional methods of analyzing teacher classroom talk and providing a snapshot of teacher talk in secondary ELA classrooms.

4.1. Implications for research

We found that our ASR-based methods for analyzing features of teacher talk—which included segmenting talk at 1.0 s pauses in teacher speech—often yielded similar prevalence rates and/or high lesson-level correlations with traditional methods of ELA teacher talk analysis. These results suggest that our semi-automated method is promising and could promote advances in technology that provide useful feedback to teachers on their classroom practices. However, our methods yielded prevalence rates for some features of teacher talk – such as goal specificity and feedback – that were quite different from those reported in studies using traditional methods and had low lesson-level correlations with traditional methods, suggesting fundamental differences in construct representation and the possibility that further advances in technology may be needed before ASR-based methods can accurately capture these specific talk features.

This study also contributes to scholarly discussions about researcher transparency in reporting methodological decision-making with regards to segmenting and coding classroom talk. In our review of related literature, we often had trouble finding consistent units of analysis used to report the frequency of specific teacher talk features. For example, a large body of literature on teacher talk reports the frequency of uptake, cognitive demand, and feedback only within teacher questions, not statements (Applebee et al., 2003; Juzwik et al., 2013). However, our findings indicate that teacher talk is comprised of more statements than questions and feedback and uptake often appear within those statements. Additionally, our results suggest that for ELA terminology and other talk features, using segments as units of analysis rather than words, questions or activities might better capture high-quality instructional talk and thus provide more useful feedback to teachers. Thus, we recommend that the field consider more transparency in the reporting of units of analysis, methodological decision making (e.g., segmentation and choices in units of analysis) and developing shared units of analysis so that the results of individual studies can be more easily compared and aggregated. Additionally, the public sharing of corpora of classroom talk, a growing practice in the field of human language technologies, would allow multiple research teams to analyze the same dataset and compare results using different theoretical and methodological approaches.

Compared to seminal research on secondary ELA teacher talk, particularly dialogic instruction, we found similar prevalence rates of authenticity and uptake in teacher questions in studies with similar datasets (Gamoran & Kelly, 2003; Kelly, 2007), but higher prevalence rates for high cognitive demand teacher questions than found in most studies published in the last two decades (see Kelly et al., 2018). The differences in these prevalence rates suggest that small differences in dataset characteristics and contexts (e.g.,

Table 7
Correlation matrix of teacher talk features.

	1	2	3	4	5	6	7
1. Authentic Questions							
2. High Cognitive Demand	.417**						
3. Uptake	.715***	.632***					
4. Serial Questions	.328***	.252	.236				
5. Negative Feedback	−.198	−.271**	−.220	−.125			
6. ELA Terms	−.276*	−.143	−.219*	−.147*	.044		
7. Goal Specificity	−.097	.212*	.043	−.162	−.050	.123	

*p < .05, **p < .01, ***p < .001.

teachers' awareness of dialogic instruction, age of students) as well as coding methods may affect results and make it difficult to generalize beyond a single study even when analyzing a large dataset. In the case of uptake, a seemingly small but important change in coding (i.e., the trend in more recent studies to use the strict definition we use here, which omits mere revoicing) may affect reported prevalence rates. However, our results also suggest that dialogic instruction is widespread and frequent in secondary ELA classrooms despite the focus in recent educational policy on standardized testing and state-level standards.

Finally, our findings also suggest some features of instructional talk that are widely valued in K-12 settings, such as explaining learning goals and using subject-specific vocabulary, warrant further research to advance methodological approaches, shared definition of codes, and awareness of prevalence rates.

4.2. Implications for teacher development

The features of teacher talk that we analyzed occurred regularly and with enough variation (e.g., use of ELA terms) to be useful topics for preservice and inservice teacher professional development. Additionally, our results suggest that talk features that appeared less often in our data but are considered key features of effective instruction (e.g., uptake) could be emphasized more within ELA teacher professional development efforts in order to improve instruction.

Teachers' reflections on their classroom talk facilitated by an expert, such as an instructional coach, and grounded in an analysis of classroom recordings (audio or video) or transcripts are effective for teacher learning and growth (Kucan, 2007; Matsumura et al., 2016; Sedova et al., 2016) but also difficult to provide at scale. Semi-automated methods for analyzing teacher talk, such as ours, hold great promise as more efficient and economical ways to provide teachers with classroom discourse data that they can analyze and learn from, particularly as part of ongoing efforts to provide low-stakes opportunities for teachers to improve their literacy instruction through reflection and lesson planning.

Matsumura et al. (2016) developed a Cloud Coaching model in which teachers are able to record their classrooms and reflect on their practice virtually with the guidance of an instructional coach. Our results suggest that semi-automated, ASR-driven methods could be used to create similar technologically-driven systems that would: (1) provide more frequent and timely feedback and learning opportunities for teachers who aim to improve their instructional talk independently and with lower stakes and (2) work in conjunction with pre-existing virtual, Cloud Coaching model. ASR-transcripts of teacher talk are versatile, low-cost, and less intensive to develop. Further, they can provide summaries or examples of patterns that can quickly cut across a week or a month in small stretches. We hypothesize that ASR-transcripts can be wrapped in as low-cost way to anchor a discussion about teacher talk between a teacher and a coach, or a group of teachers. Additionally, the efficiency of our semi-automated system allows teachers to view discourse data almost instantaneously, thus encouraging immediate changes to instructional talk and lesson planning.

Currently, some technological challenges limit the potential of semi-automated methods. ASR methods do not produce transcripts that are 100% accurate and thus the transcripts themselves might not be useful for teacher reflection and learning. Further, limitations in microphone and ASR technology do not yet allow for accurate transcription of multiple speakers, such as students. Thus, another limitation of our chosen method is that it cannot yet capture and analyze student talk, which could allow for a more robust understanding of classroom discourse more broadly.

Finally, a limitation of our study is the lack of racial and language diversity represented by participating schools, which were located in privileged, suburban communities and served majority white students whose first language was English. Further, all teacher participants in our sample identified as non-Hispanic and white. In the future, we plan to collaborate with more diverse school districts and teachers.

5. Conclusion

Our results support the goal of developing automated systems that can detect features of teacher talk with acceptable accuracy and provide useful information to teachers (Jensen et al., 2021, pp. 302–312). We have begun developing a smartphone-based app that does this and that provides opportunities for teacher reflection. Given the considerable body of research demonstrating the importance of high-quality teacher talk and the current barriers to providing frequent and timely feedback to teachers on their talk, we believe that research such as ours that draws on advances in technology such as ASR, machine learning, and natural language processing has the potential to provide teachers and researchers with greater insight into the features of instructional talk that impact student learning. Applying such technology to the field of K-12 literacy has the potential to improve teacher talk and thus instructional quality and learning opportunities for all students.

References

Alexander, R. (2008). *Towards dialogic teaching: Rethinking classroom talk* (4th ed.) Dialogos.

Applebee, A. N., Langer, J., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685–730. https://doi.org/10.3102/00028312040003685

Auckerman, M. S. (2007). When reading it wrong is getting it right: Shared evaluation pedagogy. *Research in the Teaching of English, 42*(1), 56–103.

Baumann, J. F., & Graves, M. F. (2010). What is academic vocabulary? *Journal Of Adolescent & Adult Literacy, 54*(1), 4–12. https://doi.org/10.1598/JAAL.54.1.1

Blanchard, N., Brady, M., Olney, A., Glaus, M., Sun, X., Nystrand, M., Samei, B., Kelly, S., & D'Mello, S. K. (2015). A study of automatic speech recognition in noisy classroom environments for automated dialog analysis. In C. Conati, N. Heffernan, A. Mitrovic, & M. Felisa Verdejo (Eds.), *Proceedings of the 17th international conference on artificial intelligence in education* (pp. 23–33). Springer-Verlag. https://doi.org/10.1007/978-3-319-19773-9_3.

Bloome, D., Carter, S. P., Christian, B. M., Otto, S., & Shuart-Faris, N. (2004). *Discourse analysis and the study of classroom language and literacy events: A microethnographic perspective*. Lawrence Erlbaum Associates, Inc.
Boyd, M. P. (2015). Relations between teacher questioning and student talk in one elementary ELL classroom. *Journal of Literacy Research, 47*(3), 370–404. https://doi.org/10.1177/1086296X16632451
Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning* (2nd ed.). Heinemann.
Chin, C. (2006). Classroom interactions in science: Teacher questioning and feedback to students' responses. *International Journal of Science Education, 28*(11), 1315–1346. https://doi.org/10.1080/09500690600621100
Chisholm, J., & Godley, A. J. (2011). Learning about language through inquiry-based discussion: Three bidialectal high school students' talk about dialect variation, identity and power. *Journal of Literacy Research, 43*(4), 430–468. https://doi.org/10.1177/1086296X11424200
Clarke, S. N., Resnick, L., & Rosé, C. (2015). Dialogic instruction: A new frontier. In L. Corno, & E. Anderman (Eds.), *Handbook of educational psychology* (3rd ed., pp. 378–389). Taylor and Francis.
Conrad, J. (2006). *Heart of darkness*. W. W. Norton & Company, Inc.
Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership, 68*(4), 35–39.
Dickens, C. (2020). *A tale of two cities*. East India Publishing Company.
Duke, N., Caughlan, S., Juzwik, M., & Martin, N. (2012). *Reading and writing genre with purpose in a K-8 classroom*. Heinemann.
Ernst-Slavit, G., & Mason, M. R. (2011). "Words that hold us up:" Teacher talk and academic language in five upper elementary classrooms. *Linguistics and Education, 22*, 430–440. https://doi.org/10.1016/j.linged.2011.04.004
Ernst-Slavit, G., & Pratt, K. L. (2017). Teacher questions: Learning the discourse of science in a linguistically diverse elementary classroom. *Linguistics and Education, 40*, 1–10. https://doi.org/10.1016/j.linged.2017.05.005
Fisher, D. (2009). The use of instructional time in the typical high school classroom. *The Educational Forum, 73*(2), 168–176. https://doi.org/10.1080/00131720902739650
Gamoran, A., & Kelly, S. (2003). Tracking, instruction, and unequal literacy in secondary school English. *Stability and change in American education: Structure, process, and outcomes, 109*(1), 109–126.
Gerritsen, D. (2018). *A socio-technical approach to feedback and instructional development for teaching assistants (Unpublished doctoral dissertation)*. Carnegie Mellon University (CMU-HCII-18-102).
Greeno, J. G. (2015). Classroom talk sequences and learning. In L. Resnick, C. Asterhan, & S. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 255–262). American Educational Research Association.
Grossman, P., Loeb, S., Cohen, L., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education, 19*(3), 445–470. https://doi.org/10.1086/669901
Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology, 61*, 29–48. https://doi.org/10.1348/000711006X126600
Haddon, M. (2003). *The curious incident of the dog in the night-time*. Random House, Inc.
Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. The Falmer Press.
Hennessy, S., Howe, C., Mercer, N., & Vrikki, M. (2020). Coding classroom dialogue: Methodological considerations for researchers. *Learning Culture and Social Interaction, 25*. https://doi.org/10.1016/j.lcsi.2020.100404
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 analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction, 9*, 16–44. https://doi.org/10.1016/j.lcsi.2015.12.001
Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511. https://doi.org/10.1080/07370000802177235
Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371–384. https://doi-org.pitt.idm.oclc.org/10.17763/haer.83.2.d11511403715u376.
Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education, 43*(3), 325–356. https://doi.org/10.1080/0305764X.2013.786024
Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher—student dialogue during classroom teaching: Does it really impact on student outcomes? *The Journal of the Learning Sciences, 28*(4–5), 462–512. https://doi.org/10.1080/10508406.2019.1573730
Jensen, E., Dale, M., Donnelly, P., Stone, C., Kelly, S., Godley, A. J., & D'Mello, S. (2020). Toward automated feedback on teacher discourse to enhance teacher learning. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems USA*, 1–13. https://doi.org/10.1145/3313831.3376418
Jensen, E., Pugh, S., & D'Mello, S. K. (2021). *A deep transfer learning approach to automated teacher discourse feedback*. USA: Proceedings of the 11th Learning Analytics & Knowledge Conference. https://doi.org/10.1145/3448139.3448168
Juzwik, M. M., Borsheim-Black, C., Caughlan, S., & Heintz, A. (2013). *Inspiring dialogue: Talking to learn in the English classroom*. Teachers College Press.
Juzwik, M. M., Sherry, M. B., Caughlan, S., Heintz, A., Borsheim-Black, C., Auerbach, S., & Herbst, P. (2012). Supporting dialogically organized instruction in an English teacher preparation program: A video-based, web 2.0-mediated response and revision pedagogy. *Teachers College Record, 114*, 1–42.
Kelly, S. (2007). Classroom discourse and the distribution of student engagement. *Social Psychology Of Education, 10*(3), 331–352. https://doi.org/10.1007/s11218-007-9024-0
Kelly, S., Olney, A. M., Donnelly, P., Nystrand, M., & D'Mello, S. K. (2018). Automatically measuring question authenticity in real-world classrooms. *Educational Researcher, 47*, 451–464. https://doi.org/10.3102/0013189X18785613
Kucan, L. (2007). Insights from teachers who analyzed transcripts of their own classroom discussions. *The Reading Teacher, 61*(3), 228–236. https://doi.org/10.1598/RT.61.3.3
Lee, C. D. (2006). "Every good-bye ain't gone": Analyzing the cultural underpinnings of classroom talk. *International Journal of Qualitative Studies in Education, 19*(3), 305–327. https://doi.org/10.1080/09518390600696729
Lefstein, A., Snell, J., & Israeli, M. (2015). From moves to sequences: Expanding the unit of analysis in the study of classroom discourse. *British Educational Research Journal, 41*(5), 866–885. https://doi.org/10.1002/berj.3164
Lugini, L., Litman, D., Godley, A., & Olshefski, C. (2018). Annotating student talk in text- based classroom discussions. In *Proceedings of the 13th workshop on innovative use of NLP for building educational applications* (pp. 110–116). New Orleans, LA, 2018.
Matsumura, L., Bickel, D., Zook-Howell, D., Correnti, R., & Walsh, M. (2016). Cloud coaching: Web-based learning holds promise, especially for districts with limited resources. *Journal of Staff Development, 37*(4), 30–39. Retrieved from: https://ifl.pitt.edu/documents/Cloud-Coaching.pdf.
Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2012). The effect of content-focused coaching on the quality of classroom text discussions. *Journal of Teacher Education, 63*(3), 214–228. https://doi.org/10.1177/0022487111434985
Molinari, L., & Mameli, C. (2013). Process quality of classroom discourse: Pupil participation and learning opportunities. *International Journal of Educational Research, 62*, 249–258. https://doi.org/10.1016/j.ijer.2013.05.003
Muhonen, H., Pakarinen, E., Poikkeus, A.-M., Lerkkanen, M.-K., & Rasku-Puttonen, H. (2018). Quality of educational dialogue and association with students' academic performance. *Learning and Instruction, 55*, 67–79. https://doi.org/10.1016/j.learninstruc.2017.09.007
National Governors Association & Council of Chief State School Officers. (2010). *Common core state standards for English language arts & literacy in social studies, science, and technical subjects*. National Governors Association Center for Best Practices, Council of Chief State School Officers.
Newmann, F. M., Whelage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. M. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11–39). Teachers College Press.
Nystrand, M. (2006). Research on the role of classroom discourse as it affects reading comprehension. *Research in the Teaching of English, 40*(4), 392–412.
Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English, 25*(3), 261–290.
Nystrand, M., & Gamoran, A. (1997). The big picture: Language and learning in hundreds of English lessons. In M. Nystrand (Ed.), *Opening dialogue: Understanding the dynamics of language and learning in the English classroom* (pp. 30–74). Teachers College Press.
Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes, 35*(2), 135–198. https://doi.org/10.1207/S15326950DP3502_3
van de Pol, J., Brindley, S., & Higham, R. J. (2017). Two secondary teachers' understanding and classroom practice of dialogic teaching: A case study. *Educational Studies, 43*(5), 497–515. https://doi.org/10.1080/03055698.2017.1293508
Rahimi, Z., Litman, D., Wang, E., & Correnti, R. (2015). Incorporating coherence of topics as a criterion in automatic response-to-text assessment of the organization of writing. In *Proceedings of the 10th workshop on innovative use of NLP for building educational applications* (pp. 20–30).
Resnick, L. B., & Schantz, F. (2015). Talking to learn: The promise and challenge of dialogic teaching. In L. B. Resenick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 441–450). American Educational Research Association.
Resnitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist, 48*(2), 114–133. https://doi.org/10.1080/00461520.2013.775898
Rosé, C. P., & Tovares, A. (2015). What sociolinguistics and machine learning have to say to each other about interaction analysis. In L. B. Resenick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 289–300). American Educational Research Association.
Sadler, D. R. (2013). Opening up feedback. In S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualizing feedback in higher education: Developing dialogue with students* (pp. 54–63). Routledge.
Schleppegrell, M. J. (2004). *the language of schooling: A functional linguistics perspective*. Lawrence Erlbaum Associates, Inc.
Sedova, K., Sedlacek, M., & Svaricek, R. (2016). Teacher professional development as a means of transforming student classroom talk. *Teaching and Teacher Education, 57*, 14–25. https://doi.org/10.1016/j.tate.2016.03.005
Shakespeare, W. (2009). *Romeo and Juliet*. Dover Publications, Inc.
Shernoff, D. J., Kelly, S., Tonks, S. M., Anderson, B., Cavanagh, R. F., Sinha, S., & Abdi, B. (2016). Student engagement as a funciton of environmental complexity in high

school classrooms. *Learning and Instruction, 43*, 52–60. https://doi.org/10.1016/j.learninstruc.2015.12.003

Smith, F. R., & Feathers, K. M. (1983). Teacher and student perceptions of content area reading. *Journal of Reading, 26*(4), 348–354.

Song, Y., Hao, T., Liu, Z., & Lan, Z. (2020). A systematic review of frameworks for coding towards classroom dialogue. In E. Popescu, T. Hao, T. C. Hsu, H. Xie, M. Temperini, & W. Chen (Eds.), *Lecture notes in computer science*. https://doi.org/10.1007/978-3-030-38778-5_25. Vol 11984.

Soter, A. O., Wilkinson, I. A., Murphy, P. K., Rudge, L., Rninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal of Educational Research, 47*, 372–391. https://doi.org/10.1016/j.ijer.2009.01.001

Stein, M. K., & Matsumura, L. C. (2009). Measuring instruction for teacher learning. In D. H. Gitomer (Ed.), *Measurement issues and assessment for teacher quality* (pp. 179–205). Sage Publications.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal, 104*(1), 3–28.

Townsend, D. (2015). Who's using the language? Supporting middle school students with content area academic language. *Journal of Adolescent Literacy, 58*(5), 376–387. https://doi.org/10.1002/jaal.374

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, Z., Pan, X., Miller, K., & Cortina, K. (2014). Automatic classification of activities in classroom discourse. *Computers & Education, 78*, 115–123. https://doi.org/10.1016/j.compedu.2014.05.010

Webb, N. L. (2006). *Web alignment tool*. Wisconsin Center of Educational Research. University of Wisconsin-Madison. Retrieved from: https://static.pdesas.org/content/documents/M1-Slide_19_DOK_Wheel_Slide.pdf.

Wells, G., & Meija-Arauz, R. (2006). Dialogue in the classroom. *The Journal of the Learning Sciences, 15*(3), 379–428. https://doi.org/10.1207/s15327809jls1503_3

Wertsch, J. V. (1991). *Voices of the mind*. Harvard University Press.

Wilson, S. M. (2013). Professional development for science teachers. *Science, 340*, 310–313.

Witherspoon, M., Sykes, G., & Bell, C. (2016). *Leading a classroom discussion: Definition, supportive evidence, and measurement of the ETS® national observational teaching examination (NOTE) assessment series (research Memorandum No. RM-16-09)*. Educational Testing Service.