

Quality Talk: Developing Students' Discourse to Promote High-level Comprehension

P. Karen Murphy 

The Pennsylvania State University

Jeffrey A. Greene

The University of North Carolina–Chapel Hill

Carla M. Firetto

Arizona State University

Brendan D. Hendrick

The University of North Carolina–Chapel Hill

Mengyi Li

American Institutes for Research

Cristin Montalbano

The University of North Carolina–Chapel Hill

Liwei Wei

The Pennsylvania State University

Students often struggle to comprehend complex text. In response, we conducted an initial, year-long study of Quality Talk, a teacher-facilitated, small-group discussion approach designed to enhance students' basic and high-level comprehension, in two fourth-grade classrooms. Specifically, teachers delivered instructional mini-lessons on discourse elements (e.g., questioning or argumentation) and conducted weekly text-based discussions in their language arts classes. Analysis of the videorecorded discussions showed decreases in teacher-initiated discourse elements, indicating a release of responsibility to students, whereas students' discourse reflected increased critical-analytic thinking (e.g., elaborated explanations or exploratory talk). Importantly, statistically and practically significant increases were evidenced on written measures of students' basic and high-level comprehension, indicating the promise of small-group discourse as a way to foster individual student learning outcomes.

KEYWORDS: argumentation, classroom discussion, critical thinking, Quality Talk, reading comprehension

Critically analyzing and comprehending content-rich, complex text is essential in light of the rapid proliferation of print and digital media in

the 21st century. Unfortunately, students often struggle when using texts from print and digital media to complete fundamental tasks such as answering inferential questions, finding quality information, vetting sources, evaluating arguments, or comprehending complex ideas (e.g., Bråten, Britt, Strømsø, & Rouet, 2011). Recent results provided by the National Assessment of Educational Progress (NAEP) revealed a substantive proportion of tested students failed to achieve basic comprehension, much less high-level comprehension as a result of engaging with text (U.S. Department of Education [USDE], Institute of Education Sciences [IES], National Center for Education Statistics [NCES], 2015). The term *high-level comprehension* refers to the outcome produced when students critically and reflectively engage with text (i.e., critical-analytic thinking) and meaningfully consider the nature and quality of the content or arguments within text (i.e., Iordanou, Kendeou, & Beker, 2016). Educators need an instructionally supported intervention designed to

P. KAREN MURPHY is a distinguished professor of education at The Pennsylvania State University, 102 CEDAR Building, University Park, PA 16802; e-mail: pkm15@psu.edu, where she holds a joint appointment in the Educational Psychology program and the Children, Youth, and Families Consortium. She is also serving as the editor-in-chief of the Review of Educational Research. Her current research focuses on the role of classroom discussion on students' high-level comprehension (see www.qualitytalk.psu.edu).

JEFFREY A. GREENE is an associate professor of educational psychology and learning sciences in the School of Education at the University of North Carolina at Chapel Hill. His research interests include effective means of promoting digital literacy, particularly via helping students enact self-regulated learning and epistemic cognition.

CARLA M. FIRETTO is an assistant professor of educational psychology in Mary Lou Fulton Teachers College at Arizona State University. Her research aims to facilitate students' integration across multiple, complex texts as well as their high-level comprehension of those texts, particularly through the use of classroom discourse.

BRENDAN D. HENDRICK is a school psychologist in Guilford County Schools and former graduate research assistant at the University of North Carolina at Chapel Hill. He works to apply educational research in the support of students with disabilities via instructional intervention planning, mental health support/coordination, and program development/evaluation.

MENGYI LI is a researcher at the American Institutes for Research and former graduate research assistant at The Pennsylvania State University. Her research interests include reading comprehension, classroom discussion, and educational measurement, especially as applied to large-scale national survey and assessment programs.

CRISTIN MONTALBANO is a doctoral candidate in the school psychology program at the University of North Carolina at Chapel Hill. Her research focuses on various school-based interventions geared toward promoting critical-analytic thinking, relational reasoning, and self-regulation.

LIWEI WEI is a doctoral candidate in educational psychology in the College of Education at The Pennsylvania State University. Her research focuses on small-group discussion interventions that promote high-level comprehension and critical-analytic thinking for both English-speaking students and English language learners.

directly affect students' ability to complete these critical tasks and promote their subsequent high-level comprehension. Dialogically driven pedagogical approaches may prove a useful tool for augmenting student comprehension. Indeed, research has shown that classroom discussions can enhance students' basic comprehension of text (McKeown, Beck, & Blake, 2009) and that some discussions may also effectively promote critical-analytic thinking about text (Reznitskaya et al., 2008).

The challenge is that few, if any, text-based discussion models are equally effective at promoting basic and high-level comprehension, and existing approaches rarely emphasize the role of critical-analytic thinking in the comprehension of complex texts (Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009). These issues are further compounded by the fact that in existing approaches students are provided almost no explicit instruction regarding the nature of discourse necessary to promote high-level comprehension (Murphy, Wilkinson, & Soter, 2010). As such, students must rely almost exclusively on their own inferences from teacher moves or modeling to glean understandings of what should take place within the discussion. The overarching purpose of the present investigation was to examine changes in fourth-grade teachers' and students' discourse, as well as students' basic and high-level comprehension, while participating in a year-long implementation of Quality Talk (QT), a teacher-facilitated, text-based discussion approach.

Discourse Patterns in Literacy Classrooms

It is widely accepted that teachers can facilitate student learning with discourse-intensive pedagogies (Wilkinson, Soter, & Murphy, 2010). The challenge, of course, is not so much how much discourse occurs in classrooms but rather the nature of discourse in classrooms. Further, we also know talk patterns are persistent even with substantial support, as it is difficult for teachers to release control of discussions and interpretative authority to students (Alvermann, O'Brien, & Dillon, 1990; Billings & Fitzgerald, 2002; Kucan, 2009; Mehan, 1979). As a case in point, Alvermann and Hayes (1989) employed a series of 10 intervention-coaching cycles with five 7th- through 12th-grade teachers over the course of an academic year. The cycle included lesson planning, coaching with video of the teacher's instruction, and additional planning for improvement and goal setting. The goal of the research, in many ways, was to affect the discourse culture of these classrooms. Despite the intervention-coaching cycles, an Initiate-Response-Evaluate (IRE; Mehan, 1979) style of classroom discussion about text persisted at the end of the study. The authors surmised that teachers' reflection on their entrenched patterns of discourse behaviors was not robust enough to modify the culture of discourse in the classroom. Similarly, Billings and Fitzgerald (2002) found that even when interventions like Paideia seminars lead to

changes in the overall number of turns, teachers still held the floor for relatively longer periods of time and expressed difficulty altering their instructional style.

Additional challenges are introduced when the emphasis moves beyond turn taking to the nature of the discourse. Indeed, discussions intended to promote high-level comprehension, critical-analytic thinking, and reasoning require further changes to classroom norms and often challenge teachers' prior experiences and beliefs (Alvermann & Hayes, 1989; Michaels & O'Connor, 2015). As Soter et al. (2008) revealed, discussions that evidence increases in high-level comprehension are characterized by shared control between the teacher and students, students' holding interpretative authority over textual interpretation, and extended periods of student-to-student talk focused on authentic, open-ended questions, co-construction of meaning, and elaborated explanations. The teachers' role must shift to one of teacher-as-facilitator, whose primary focus is using moves to promote particular kinds of student talk (Wei & Murphy, 2017).

Despite the documented difficulties associated with effectively modifying discourse in literacy classrooms, recent studies offer some room for optimism. Specifically, Ryu and Sandoval (2012) investigated the extent to which sustained practice of argumentation would improve students' understanding of epistemic criteria for scientific arguments. An overarching goal was to shift the culture of the classroom toward one that valued argumentation and reasoning as routine discourse practices. In doing so, both teachers and students played primary roles in facilitating the cultural shift. Similar findings were reported by Van den Bergh, Ros, and Beijaard (2014) in a study focused on improving teachers' feedback patterns during active student learning. The investigation involved professional development focused on altering teacher discourse practices and concomitant intervention with students encouraging active learning. Through the course of the study, teachers and students gradually modified their discourse patterns. Improving classroom discourse communities and students' high-level comprehension may require that both teachers and students shift their roles and talk patterns.

Predictors of Students' High-Level Comprehension

High-level comprehension requires students to think *about* the text (i.e., basic or explicit comprehension) as well as *around* and *with* the text in critical and analytic ways through argumentation and epistemic cognition (Bråten et al., 2011; Iordanou et al., 2016). Students' critical-analytic thinking includes "cognitive processing through which an individual or group of individuals comes to an examined understanding" (Murphy, Rowe, Ramani, & Silverman, 2014, p. 563), which includes both how students assert their own views (i.e., elaborated explanations) as well as how they interact with and challenge each other's claims (i.e., exploratory talk; Murphy,

Firetto, Greene, & Butler, 2017). Argumentation is the overarching process of developing, critiquing, and defending those claims with reasons and evidence, and students develop argumentation skills by engaging in discourse with others (Iordanou et al., 2016). In literary reasoning in particular but also in other disciplines, the quality of students' critical-analytic thinking and their ability to successfully engage in argumentation with others both depend upon students' epistemic cognition (Chinn, Buckland, & Samarapungavan, 2011; Kuhn, Cheney, & Weinstock, 2000; Lee, Goldman, Levine, & Magliano, 2016).

Students' epistemic cognition involves all the ways students acquire, understand, justify, change, and use knowledge (Greene, Sandoval, & Bråten, 2016). When students determine what they know versus what they think, believe, doubt, or outright discount, they are engaged in epistemic cognition. Epistemic cognition exerts a "quiet but powerful" (Alexander, Murphy, Guan, & Murphy, 1998, p. 97) influence on how students approach and use text and engage in argumentation in academic disciplines (Chinn et al., 2011; Chinn, Rinehart, & Buckland, 2014; Iordanou et al., 2016). To engage in the argumentative processes necessary to achieve high-level comprehension of text in literacy classrooms, students must adopt effective epistemic beliefs, such as viewing texts as constructed and thus requiring critical examination of the arguments in that text (Bråten, Anmarkrud, Brandmo, & Strømsø, 2014; Greene, Azevedo, & Torney-Purta, 2008; Mason & Scirca, 2006; Sandoval, 2005). This critical approach to texts requires students to adopt normative epistemic practices (Lee et al., 2016), such as ways of determining the authority and veracity of the source, the coherence of the claims with their own prior knowledge, and the degree to which the claims meet the standards and criteria of the discipline (e.g., replicability in science or breadth and specificity in history; Murphy, Alexander, & Muis, 2012). Students who do not adopt these effective epistemic beliefs and practices (e.g., instead passively committing information to memory without the kind of inquiry necessary to sort knowledge from speculation or to construct coherent, accurate models of the world) are highly unlikely to have the textual knowledge necessary to engage in critical-analytic thinking or argumentation, nor is it likely they will internalize what they learn during discourse to achieve high-level comprehension (Bråten et al., 2011; Greene et al., 2008; Kuhn et al., 2000; Mason & Boscolo, 2004; Weinstock, Neuman, & Glassner, 2006).

The connections between epistemic cognition and argumentation, both oral and written, are clear, with interventions leading to changes in student learning across numerous academic disciplinary contexts such as science (Bråten, Ferguson, Strømsø, & Anmarkrud, 2014) and literacy (Reznitskaya et al., 2012). Indeed, discourse with peers, when structured intentionally and practiced over long periods of time, can lead to increased argumentation reasoning skills in a variety of settings from literacy to science classrooms (Berland & Reiser, 2011; Iordanou & Constantinou, 2015; Kuhn, Zillmer,

Crowell, & Zavala, 2013; Schwarz & De Groot, 2007). For example, as mentioned previously, Ryu and Sandoval (2012) found with elementary school students that explicit instruction in argumentation and epistemic criteria for evaluating and making knowledge claims (e.g., citing evidence) led to better argumentation performance. Such changes in oral argumentation skill has also been shown to transfer to written performance, particularly in terms of normative ways of making arguments (Kuhn et al., 2013).

In addition to epistemic cognition and argumentation skills, a number of student factors appear to account for variance in students' comprehension of text including gender and reading fluency. Gender differences in reading achievement have been widely documented in the extant literature. The recent report released by NAEP (USDE, IES, & NCES, 2015) showed that female students generally outperformed their male counterparts in the domain of reading. This trend has also been evidenced in a number of large-scale studies of students' performance across various ages and academic levels (e.g., Elley, 1994; Wagemaker, 1996). Such gender differences may be due in part to students' motivation, attitudes, and engagement. Indeed, students' motivation in reading was found to significantly mediate such gender differences (Chiu & McBride-Chang, 2006). Further, female students reported more positive attitudes toward reading tasks (Logan & Johnston, 2009), and in classroom discussions, female students also exhibited greater motivation and engagement than male students (Wu, Anderson, Nguyen-Jahiel, & Miller, 2013).

Lastly, extensive research has demonstrated that oral reading fluency (ORF) serves as a strong predictor of students' overall reading competency, especially in the early elementary grades (e.g., Adams, 1990; Fuchs, Fuchs, Hosp, & Jenkins, 2001). ORF scores have been identified in numerous empirical investigations as the most valid predictor of student reading comprehension ability (Fuchs et al., 2001; Goffreda & DiPerna, 2010; Johnson, Jenkins, Petscher, & Catts, 2009). For example, among the measures assessed in DIBELS (i.e., Dynamic Indicators of Basic Early Literacy Skills; a widely used screening and placement tool in the primary grades), Johnson and colleagues (2009) singled out ORF as having the highest classification accuracy. The compelling and consistent empirical findings on the roles of epistemic cognition, gender, and oral fluency in reading achievement begs the question of how can educators intervene to stimulate students' high-level comprehension of text.

Effects of Discussion on High-Level Comprehension in Literacy

The use of discourse to improve comprehension is undergirded by various theoretical foundations including cognitive, sociocognitive, sociocultural, and dialogic perspectives. First, engaging in a discussion allows students to cognitively participate in meaning-making (McKeown et al.,

2009) while also evaluating textual claims and evidence (Greene et al., 2008). From a sociocognitive perspective, as students participate in discussions, they state their thoughts and opinions, while also considering the thoughts and opinions of their peers. When students are faced with thoughts and opinions that differ from their own, they must decide how to reconcile such conflicts (Almasi, 1995; Michaels, O'Connor, & Resnick, 2008). Socioculturally, as students participate in discussions, they co-construct knowledge by using language as a tool to build on others' ideas and construct knowledge together, collaboratively (Vygotsky, 1962). Through the discourse, students internalize what initially is derived in conjunction with their peers. Then, students may transfer this knowledge and understanding to other texts, after participating in the discussion (Wells, 2007). Finally, the dialogic perspective suggests that students' comprehension is influenced by the conflicting voices within a discussion (Nystrand, 2006).

There has been considerable research on literacy classroom approaches to conducting discussions about text and their effects on comprehension. Murphy et al. (2009) conducted a meta-analytic investigation of 42 empirical studies examining the effects of these prominent discussion approaches on both teacher and student talk as well as students' outcomes. The various approaches focused primarily on the comprehension of literary text and served an array of purposes based on the goals that teachers set for their students (e.g., responding to literature on an aesthetic level, adopting a critical-analytic stance, or acquiring information on an efferent level). Results of the meta-analysis indicated that the approaches differentially promoted high-level comprehension of text. Most approaches supported students' literal and inferential comprehension, particularly those that were more efferent in nature (i.e., focused on knowledge seeking) such as *Instructional Conversations* (Goldenberg, 1993), *Junior Great Books Shared Inquiry* (Great Books Foundation, 1987), and *Questioning the Author* (Beck & McKeown, 2006), whereas other approaches were exceptionally effective at enhancing students' critical-thinking and reasoning with text (e.g., *Collaborative Reasoning* [CR]; Anderson, Chinn, Waggoner, & Nguyen, 1998). This latter type of reasoned thinking and consideration aligns well with our notion of high-level comprehension. *Philosophy for Children* (P4C; Billings & Fitzgerald, 2002) along with CR were the only two approaches that examined the effect of discourse on argumentation outcomes (P4C, effect size [ES] = 0.214; CR, ES = 0.260). The meta-analysis also revealed that students' comprehension gains were not directly related to overall increases in student talk. Instead, the meta-analytic results revealed that a particular kind of critical-analytic talk was necessary to enrich students' comprehension. Taken together, these findings suggest that the more critical-analytic approaches seemed to encourage high-level comprehension.

To further investigate the nature of student talk, Soter et al. (2008) conducted a comprehensive discourse analysis of transcripts from the aforementioned nine discussion approaches and found that the approaches that were more effective at promoting students' high-level comprehension showed high incidences of certain discourse features or elements (i.e., proximal indices of students' learning and comprehension). Among the discourse elements present in approaches that stimulated students' high-level comprehension were (a) knowledge construction through frequent incidences of authentic questions and uptake (see Nystrand, 1997), (b) high rates of questions that elicited high-level thinking, and (c) high incidences of elaborated explanations (i.e., an elaborated explanation is a statement of a claim that is based on at least two independent, conjunctive, or causally connected forms of support; Webb, 1991), collective reasoning, and/or exploratory talk (i.e., an instance of exploratory talk is where students co-construct knowledge together; Mercer, 2000). Importantly, the approaches labeled as critical-analytic by Murphy et al. (2009) showed high frequencies of both elaborated explanations as well as exploratory talk (e.g., P4C and CR). By contrast, the more expressive approaches (e.g., *Book Club*) elicited high occurrences of exploratory talk with somewhat fewer occurrences of elaborated explanations. It seems that the shared control between teachers and students evident in the more critical-analytic discussions gave way to richer reasoning (i.e., Murphy et al., 2009; Soter et al., 2008). Shared control provided opportunities for students to engage in extended episodes of collective reasoning but, at the same time, afforded opportunities for modeling and scaffolding of students' individual reasoning.

Based on this work, Wilkinson et al. (2010) developed an initial QT discussion model for fostering high-level comprehension of literary text. This model combined the best features of the nine approaches, while giving prominence to those approaches that emphasized a critical-analytic stance (e.g., CR). Perhaps most importantly, a strong emphasis was placed on the teacher and students sharing control of discussion with a moderate to high degree of emphasis placed on the generation of expressive and efferent connections to the texts. Arguably, such an approach allows for a gradual release of responsibility by teachers as student began to take on interpretive authority of the text.

Quality Talk

Quality Talk (QT) is a multifaceted approach toward classroom discussions designed to increase students' high-level comprehension by encouraging students to think and talk about, around, and with the text. In QT, high-level comprehension is achieved through critical-analytic thinking in discourse, which fosters students' basic comprehension, epistemic cognition, and ability to engage in oral and written argumentation. The approach

consists of four interrelated components: an ideal instructional frame, discourse elements, teacher discourse moves, and pedagogical principles (Murphy & Firetto, 2017; Wilkinson et al., 2010). The present application of the approach, reflecting our iterative refinements of the Wilkinson et al. (2010) model, also involves initial and ongoing professional development and coaching, student journals, and a series of sequential, explicit mini-lessons for students pertaining to the elements of productive discourse (e.g., authentic questioning or argumentation; Murphy & Firetto, 2017).

Ideal Instructional Frame

The *ideal instructional frame* embodies a set of conditions we deem fundamental for promoting productive talk about text. QT discussions take place in small groups of four to six students with shared control between teacher and students. As facilitator, the teacher has the authority to choose both the text and topic, but students have control over turns via an open participation structure as well as interpretive authority (Anderson et al., 2001; Nystrand, 1997). Prior to the discussion, students receive explicit instruction through a series of mini-lessons on how to ask and respond to meaningful, authentic questions in critical-analytic ways. Students must read the text and complete a prediscussion activity in their QT journals that ensures basic comprehension of the text by identifying relevant text structures or features (e.g., main idea) and crafting authentic questions. During the discussion, the teacher plays the role of facilitator by fostering a moderate degree of affective and knowledge-driven engagement as well as encouraging students to interrogate or query the text in search of its underlying arguments, assumptions, or beliefs (i.e., epistemic competence; Murphy & Alexander, 2016). Students encourage each other to talk about personal connections to the text (i.e., an expressive response) as well as to retrieve information (i.e., an efferent stance) during the discussions. When students have a basic understanding of the text and an opportunity to generate connections to it, they are better positioned to take on a critical-analytic stance. Finally, in alignment with the Vygotskian (Vygotsky, 1978) notion of internalization, students must take part in a postdiscussion activity in their journal where they individually commit to their text-based perspectives in writing (Graham & Harris, 2014).

Discourse Elements

The *discourse elements* comprise the second component of the model (see Table 1) and serve as tools for facilitating critical-analytic thinking. For example, authentic questions are among the discourse elements essential to QT. Teachers and students may ask a variety of open-ended, authentic questions (e.g., “What if Jim’s brother lived from the disease?”) that include follow-up questions that build on others’ contributions (i.e., uptake

Table 1
Description of Discourse Elements

Discourse Element	Initiator (Teacher/Student)	Description
Authentic question	Teacher/student	A question in which the person asking does not know the answer or genuinely wants to know how others will answer
Test question	Teacher/student	An inauthentic question that presupposes an answer
Uptake question	Teacher/student	A question that occurs when a person asks a question about something that someone else previously said
High-level thinking question	Teacher/student	A question that elicits generalization or analysis
Speculation question	Teacher/student	A question that requires students to consider and/or weigh alternative possibilities
Affective question	Teacher/student	A question that elicits information about students' feelings or about their personal experiences in relation to the content they are discussing
Intertextual question	Teacher/student	A question that elicits a reference to other literary or nonliterary works
Shared knowledge question	Teacher/student	A question that elicits reference to information that may be assumed to be common knowledge among the students in a given discussion
Elaborated explanation	Student	A statement of a claim (e.g., position, opinion, or belief) that is based on at least two independent, conjunctive, or causally connected forms of support (e.g., reasons or evidence)
Exploratory talk	Student	An instance or episode of student talk in which students co-construct knowledge together
Teacher move	Teacher	A question or response intended to support or scaffold students' discourse (e.g., modeling, summarizing, marking, prompting, challenging)

Note. Adapted from the coding manual used in Soter et al. (2008).

questions) as well as questions that elicit critical-analytic thinking (i.e., generalization, analysis, and speculation; Nystrand, 1997; Nystrand, Wu, Gamoran, Zeiser, & Long, 2003) or textual connections (i.e., affective,

intertextual, and shared knowledge connections; Applebee, Langer, Nystrand, & Gamoran, 2003; Edwards & Mercer, 1987; Taylor, Pearson, Peterson, & Rodriguez, 2003). Further, as students respond to these questions, they may generate elaborated explanations (e.g., “I think that would change a lot because then he wouldn’t ... have gone back home, and then he wouldn’t have gone to a whole bunch of different schools, he wouldn’t have made it to Carlisle”) and instances of exploratory talk where they consider alternative perspectives and challenge each other (Chinn, O’Donnell, & Jinks, 2000; Mercer, 1995, 2000; Webb, 1989). Students’ epistemic cognition, including how they scrutinize sources as well as construct and critique justifications for claims, develops as students receive direct instruction in working with reasons, evidence, and counterarguments (Bråten et al., 2011; Greene et al., 2016).

Teacher Discourse Moves

In order to implement the instructional frame, the way teachers engage in and lead the discussion changes over time as they implement QT. We delineate certain kinds of talk and support that teachers can provide to promote productive discussions that we refer to as *teacher discourse moves* (Wei, Murphy, & Firetto, in press). In the beginning, teachers may need to provide more support and guidance—they talk more frequently and use more teacher moves. They may model the talk they expect students to generate (e.g., “I’m going to start by asking an authentic question...”), or they may reinforce instances where a student excels (e.g., “That was a great elaborated explanation, Sienna”). Gradually, as students learn what is expected of them and how to engage in QT, teachers should release control and allow students to talk more, thus decreasing the number of teacher moves. However, teachers still remain present to provide occasional, necessary support or scaffolding. Notably, while both teacher discourse moves and the discourse elements are present within the QT discussions, teacher discourse moves are uniquely differentiated from discourse elements as they are employed by the teacher specifically as a way to scaffold specific elements of critical-analytic thinking.

Pedagogical Principles

The final component pertains to five pedagogical principles, with each encompassing a core idea about teaching that is requisite for fostering a culture of dialogically enhanced, text-based learning in the classroom. First, teachers must embrace the notion that language, or talk, is conceptualized as a *tool for thinking* (Mercer, 1995, 2000) and scrutinizing knowledge (Murphy et al., 2012) and, more generally, recognize the importance of discourse in learning. Second, the discussions must be grounded through a set of *normative discourse expectations* (i.e., ground rules) and *dialogic*

responsiveness. For example, the normative discourse expectations are set through a series of explicit rules for the QT discussions, such as “We don’t need to raise our hands” and “We respect others’ opinions” (Murphy & Firetto, 2017). Then, as students become familiar with, and engage in, discourse aligned with the normative expectations, teachers are able to gradually release responsibility and students take on interpretive authority, showing evidence of dialogic responsiveness (i.e., teachers’ receptivity to allowing their students to lead the discourse; cf. Pearson & Gallagher, 1983). Third, as teachers facilitate the discussions, they balance *structure and responsiveness*, using teacher moves to guide or reframe the conversation when necessary while allowing students the freedom to contribute in ways that are meaningful to them (cf. Cohen, 1994; King, 1999). Fourth, teachers must have *clarity of the content* being discussed, including a strong grasp of the story, and be prepared with potential questions to ask if necessary. Finally, teachers must *embrace space and diversity* within the discourse by allowing students the freedom to discuss their own unique individual experiences and backgrounds, resulting in discourse with broader and richer perspectives.

QT differs from other prominent discussion approaches with a critical-analytic focus toward text and content in several ways. QT is premised on the belief that talk can be used as a tool for promoting thinking and interthinking. Further, the QT intervention focuses on both teachers and students as change agents in discursive cultural shifts. To our knowledge, it is the only approach that incorporates researcher-designed, teacher-delivered instructional mini-lessons to equip students with discourse elements (e.g., questioning or argumentation) so that they can actively contribute and co-construct meaning and knowledge in text-based discussions. Further, rather than focusing on researcher-selected texts or series, QT is situated in authentic classrooms where the discussions are conducted with the school’s existing language arts curriculum. As a result, studies with QT have high ecological validity. Finally, the QT intervention also consists of a series of initial teacher professional development sessions and ongoing discourse coaching, which ensures teacher’s full understanding of the QT model, the high fidelity of QT implementation, and the gradual release of teacher responsibility to students in classroom discussions. These features set QT apart from other prominent discourse approaches.

Two preliminary studies have examined the effects of QT on students’ comprehension of text. In Wilkinson, Soter, Murphy, and Li (2008), 14 language arts teachers in Grades 4, 5, and 6 along with 272 of their students volunteered to participate in a year-long study using researcher-selected texts at all assessment time points. All teachers were instructed in the use of the initial QT model (Wilkinson et al., 2010) through a series of professional development sessions at the beginning of the school year, but they were not provided with any explicit instructional materials for their students.

Thereafter, seven teachers were given three follow-up professional development sessions and in-class coaching throughout the year (i.e., experimental group). The remaining teachers, matched on grade, school sector, and socioeconomic status (SES), received only the initial professional development (i.e., comparison group). Results showed that all teachers varied in their implementation of QT. Some teachers, especially those who already had a classroom culture of discourse, were able to change their practices on the basis of the initial professional development, some teachers seemed to benefit from the follow-up professional development, and some teachers had difficulty making the change irrespective of the professional development. Nevertheless, statistically significant effects in favor of the extended professional development group were obtained on a transfer test assessing students' persuasive essay writing, indicating that students in the experimental group more often articulated their positions and repeated their positions when writing arguments.

In a second study, Reninger and Wilkinson (2010) explored the involvement and discourse of two fourth- and fifth-grade teachers and selected "striving students" (i.e., students whose reading comprehension scores were below grade level) during discussions using the initial QT model. Students were placed in heterogeneous reading ability groups. Teachers were given latitude to modify the instructional approach to the needs of their classroom, and both teachers found it necessary to create impromptu mini-lessons to explicitly instruct, reinforce, or scaffold select discourse elements (e.g., exploratory talk). Given that the teachers created the brief lessons themselves, they necessarily varied across the two classrooms. Documentation of over 30 discussions across the school year revealed that participation in the model increased students' critical-analytic thinking, as evidenced in their talk. Moreover, informal assessments showed increases in students' comprehension of texts that were read and discussed in class. However, the texts varied between the two classrooms, and the results may have been influenced by the specific text that was discussed and assessed.

The current investigation extends what is known about the effects of QT on high-level comprehension in several ways. Specifically, we have used prior research (Li et al., 2016; Reninger & Wilkinson, 2010; Wilkinson et al., 2010) to inform a revision and expansion of QT. This investigation provides the first test of QT with researcher-developed explicit instruction on the discourse elements, documenting changes in discourse over a year-long implementation. Second, we explored the influence of QT using both group discourse elements and individual student outcomes (i.e., basic and high-level comprehension), while accounting for nested effects. Third, this was the first examination of QT under ecologically valid conditions using a variety of text genres from teachers' regular curriculum. Finally, given the tremendous challenges and potential rewards of authentic classroom research (Murphy, 2015), we enacted a number of procedures to ensure

high fidelity of implementation of QT. We captured data on those procedures to better understand the effects of our work and how to improve QT in the future (Greene, 2015). A number of specific research questions guided this investigation, including:

1. To what extent do *teachers* release control to students after participating in QT professional development and coaching, as evidenced by decreases in frequency of various teacher-initiated discourse elements in their discussions?
2. To what extent does *students'* critical-analytic thinking change, as evidenced by student-initiated discourse elements, when participating in QT after mini-lessons targeting those key elements, as well as from baseline to postintervention?
3. How does student performance on written *basic comprehension measures* change over the course of the QT intervention?
4. How does student performance on written *high-level comprehension measures* change over the course of the QT intervention?

Method

Participants and Design

Student participants ($n = 35$, female = 19) were recruited from two fourth-grade classrooms in one elementary school at the beginning of the school year. Across all of the students in both classrooms, 100% of parents consented and 100% of students assented to participate in the research. The elementary school was located in a small city in a Midwestern state and served approximately 300 students (30% free or reduced lunch) from kindergarten through fifth grade. The students at the school were predominantly Caucasian (86%); however, a few students were American Indian/Alaska Native (2%), Asian (2%), Black (2%), Hispanic (2%), and a small percentage identified with more than one racial group (5%). This study employed a single-group, time-series design, where two fourth-grade teachers implemented the QT intervention as part of their language arts curriculum over the course of one academic year.

Intervention

Through a series of professional development workshops, teachers learned about QT as well as how to use the intervention materials and conduct QT discussions. The researcher-developed intervention materials, which were implemented by teachers in their classrooms, included (a) a set of mini-lessons that specifically taught students how to generate authentic questions, (b) a set of mini-lessons that explicitly taught students how to respond to authentic questions using elements of argumentation, and (c) literacy journals that facilitated students' acquisition of the QT model

through prediscussion and postdiscussion activities in alignment with the existing reading curriculum (i.e., *Reading Street*).

Professional Development

An initial professional development workshop was conducted in September over 2 full days where teachers were provided with an overview of the QT instructional frame, discourse elements, discourse moves, and pedagogical principles. Ongoing professional development was provided through four additional half-day workshops and an additional full-day workshop in January. Throughout the initial and ongoing professional development, teachers were presented with the mini-lessons and taught how to incorporate the literacy journals into their classrooms. Teachers were also mentored in the coding of QT discourse elements in their discussions. Logistical implementation and scheduling questions were also discussed as needed.

Ongoing professional development was also provided through a series of nine discourse coaching sessions distributed approximately monthly during the intervention to provide support and training. For each coaching session, teachers reviewed a videorecording of one of their previously conducted discussions and completed the Discourse Reflection Inventory for Teachers (DRIFT), a semistructured tool designed to assist teachers to code and reflect on their discussions while also supporting fidelity of implementation. As teachers completed the DRIFT, they recorded the turn-taking pattern of the discussion, identified the discourse elements present in their students' talk, and assessed their progress toward pre-established goals. Having completed the DRIFT, teachers met individually with a discourse coach and established new goals and methods for continued success.

Questioning Mini-Lessons

Each of the four mini-lessons on questioning included a lesson plan, a set of presentation slides, and practice activities. The mini-lessons were developed to foster students' critical-analytic thinking in the QT discussions and specifically addressed subtypes of authentic questions including (a) uptake questions, (b) high-level thinking questions, (c) affective questions, and (d) intertextual questions. These mini-lessons were targeted toward teaching students how to ask various subtypes of authentic questions in their discussions and contrasted them with test questions (i.e., questions that typically have a particular *correct* answer). Specifically, students were taught how to ask questions that build on others' questions and elicit high-level thinking as well as connections to personal experiences, shared experiences, and other texts. Students were encouraged to use authentic questions to guide their discussions of the texts and to sparingly use test questions. Teachers delivered each mini-lesson over approximately 4 weeks as part

Murphy et al.

of their language arts instruction (e.g., two 15-minute lessons using the presentation slides and two 15-minute practice activities). Each portion of the mini-lesson delivery was videorecorded to assess fidelity.

Response Mini-Lessons

Students also received two mini-lessons focused on argumentation. Each of these mini-lessons included a lesson plan, a set of presentation slides, and a practice activity. The presentation slides for the argumentation lessons also included animated videos that provided exemplars and models of students engaging in argumentation. These mini-lessons were developed to foster students' epistemic cognition in the QT discussions and to enhance their use of argumentation in response to authentic questions. The argumentation mini-lessons addressed (a) how to generate an argument by stating a claim and supporting their claim using reasons and evidence and (b) how to generate counterarguments as well as rebuttals in order to ensure both sides of an argument were considered. These lessons helped students respond to the authentic questions in the discussions by elaborating their explanations and providing reasons and evidence to support their claims as well as challenging each other by evaluating others' sources of evidence and posing counterarguments. Each of these mini-lessons were unfolded over approximately 4 weeks as part of their language arts instruction. Each portion of the mini-lesson delivery was videorecorded to assess fidelity.

Literacy Journals

In alignment with the extant language arts curriculum, a literacy journal was produced with accompanying prediscussion and postdiscussion activities for each of the main selection texts that were part of their existing reading curriculum. Students completed corresponding portions of the literacy journal for each selection, before and after each discussion. Prior to discussions, students identified the main idea and supporting details of the text in response to prompts. They also generated four questions based on content learned in the QT mini-lessons that they could ask during their discussions.

QT Discussions With Text

Teachers facilitated small-group discussions in their classrooms approximately weekly throughout the school year from mid-September through mid-May. Discussions were conducted in the class period designated for language arts instruction. All of the discussions were based on the main selections for that week drawn from the *Reading Street* basal series, which was the adopted reading curriculum in the school, and both teachers conducted their discussions on the same texts. Throughout the year, students read and discussed texts that varied in genre. Thus, some discussions were based on

expository texts (e.g., Time 8, “Encantado: Pink Dolphin of the Amazon,” an informational text from a tourist guide’s perspective, including descriptions and details about the unique pink dolphins of the Amazon; 1,882 words; Flesch-Kincaid Grade Level 4.6), some were on narrative texts (e.g., Time 2, “Coyote School News,” a story about the various school and family adventures of a boy growing up in rural Arizona with an emphasis on the influence of the cultural traditions of his Mexican-American family; 2,587 words; Flesch-Kincaid Grade Level 3.5), and others pertained to mixed-genre texts (e.g., Time 13, “Jim Thorpe’s Bright Path,” a biographical text about the life of Jim Thorpe, from his many childhood challenges through the achievement of his athletic greatness; 2,413 words; Flesch-Kincaid Grade Level 4.7). Across all texts, Dale-Chall readability ranged from 3.5 to 5.1, and as would be expected, the texts generally became more difficult over the course of the school year.

Discussions began with a question about the text for that week; this was often a question that one of the students had written in their literacy journal. The remainder of the 15- to 20-minute discussions centered on students asking and answering authentic questions about that text. Discussions typically occurred on the day following a mini-lesson, which allowed teachers to encourage the use of the specific discourse elements that were taught in the most recent mini-lesson (e.g., high-level thinking questions or generating counterarguments). Twenty-five discussions were conducted in each classroom after the baseline data collection, but video-recordings and data were only collected for 14 of these discussions. During the discussions, teachers used discourse moves, when necessary, to facilitate students’ engagement in productive discourse. See Figures 1 and 2 for excerpts of typical QT discussions at midyear and at the end of the year.

Measures

ORF

Students’ ORF was assessed at baseline using the AIMSweb Reading Curriculum-Based Measure, which is a standardized assessment that examines the number of words read correctly per minute (Shinn & Shinn, 2002). Trained research assistants individually assessed students’ ORF. Students individually read aloud an unpracticed passage for 1 minute. On a score sheet, the researcher marked how many words the student read as well as any errors (e.g., pronouncing a word incorrectly). The process was repeated for two additional passages.

The score for each passage was calculated by subtracting the number of errors from the total number of words read by the student. The final score was calculated by taking the median of the scores across the three passages. The validity and reliability of the ORF assessment have been established, and ORF is used as an indicator for overall reading proficiency (Shinn, Good,

Week 5: <i>The Man Who Named the Clouds</i> Total Codes: 8 AQs, 1 TQ, 9 EEs, and 2 ETs	Week 14: <i>Gift from the Heart</i> Total Codes: 9 AQs, 1 TQ, 10 EEs, and 4 ETs
35: Why do you think Luke's dad wanted Luke to work in a chemistry shop?	36: [37], would you, um, throw your doll in the fire, or any of your other prized possessions?
33: Um, maybe, like, 'cause, like, maybe he was thinking his son was being lazy of watching clouds, and wanted him to have a job.	37: Um, yes, because (inaudible) doll. Um, and your parents died, so I feel that, if they threw it in the fire, I -- it would feel more special-- even though I wouldn't have it anymore, I would feel it would have been more special, because not only was it the last thing, um, that reminds me of my family, but also I used it to save, like, the entire tribe.
34: Yeah, I agree.	38: But she didn't know -- like, at the time, she didn't know what -- that, oh, this doll is going to stop the drought. So --
38: 'Cause, like, yeah, um, like (inaudible) said before --	35: But --
34: It's basically, like, the same thing, just (inaudible).	38: -- like, if I were her, and I didn't know, I would definitely not. And just to -- that meaning -- that how much meaning in that doll, that -- like, how much that doll means to me, I wouldn't. Like, because I -- you, you can't be completely sure.
38: -- if you spend the whole day watching clouds you don't get money, and yeah, you could become famous for naming some, but it's not likely to happen, really.	35: Yeah, I would -- I would do it, even if it -- maybe if it didn't work. But at least I would know that I -- I mean, that was a not -- it wasn't the best choice to make, but maybe -- I mean, if it -- if it did -- even if you weren't sure that you know that it would happen, it would be the best. I mean, if it did work, you would be really happy. So this kind of -- you should -- risks can take -- can cause stuff. I mean, can -- sometimes the outcome can be better than the risk. So I would -- do it, probably.
37: Yeah.	33: Yeah, um, if I -- um, if, like, no one else, um, threw away their most -- a favorite thing, I would do it, but with something else I really love.
38: And you -- and you wouldn't get money to...	35: She probably didn't have really anything else that she -- I mean, she -- yeah, she probably -- I don't -- maybe she did, but it seemed like that was the only thing, because she carried that around everywhere, and it said that the other people were taking care of her. And so probably didn't live in the same house that her parents did, and so (inaudible) she probably didn't own stuff other than that.
34: Well, this man got famous. I can say that.	
38: Well, yeah, but --	
36: Yeah, but his dad probably didn't know that he was going to get famous.	
33: Yeah.	
37: One out of a thousand --	
38: 'Cause then that was just like --	
37: One out of thousands of people got famous from watching clouds.	
38: Yeah, 'cause you could watch clouds all day --	
37: All day.	
38: -- but it, like... He probably didn't think he was going to be famous, because you're not really -- most people aren't likely to be really famous.	
35: Mm-hmm. And if he wasn't interested in clouds --	
34: Except if you're like a singer or something.	
35: -- probably there wouldn't be any cloud naming thing (inaudible), because probably -- probably it would still have, like, (inaudible) the other guy (inaudible) that lived in England, and English -- or what -- French, I think, wasn't it?	
T1: He was French, the other guy.	

Figure 1. Sample of QT discourse from midyear (i.e., Week 5) and again at the end of year (i.e., Week 14).

Knutson, Tilly, & Collins, 1992). Further, previous studies reported high alternative-form reliability (i.e., above .85 for a single passage and above .94 for three probes; Daniel, 2010).

Discourse Elements

Fifteen discussion time points were videorecorded and coded according to a detailed coding manual (adapted from the coding manual used in Soter et al., 2008; see also Murphy et al., 2017), which included definitions of the discourse elements indicative of critical-analytic thinking, exemplars, and coding rules (for a list of discourse elements and corresponding definitions, see Table 1). At baseline, teachers were asked to conduct a business-as-usual discussion in their classroom, which was coded as their baseline discussion (see Figure 3 for an excerpt of coded baseline discourse). A 30-minute segment of typical classroom discussion was coded for each teacher. Fourteen QT discussions, conducted approximately every other week, were also recorded and coded (see Figure 4 for an excerpt of coded discourse from Time 13). Given slight variations in the duration of the discussions across the 14 time points (i.e., between 15 and 20 minutes long), the middle 10-minute segment of each of the discussions were coded (i.e., 30 minutes of total discussion time was coded per teacher per time point).

Turn	Speaker	Discourse
1.	Mrs. Glen	So Maria is going to start us off with a question today. And remember to work on uptake. And also, think about the arguments, um, lessons that we had. And if you have a claim, make sure you have what? Have it reason, and --
2.	Class	[multiple students respond] And evidence.
3.	Mrs. Glen	Good. OK, Maria.
4.	Maria	Why do the pink dolphins not have dorsal fins?
5.	Laila	I don't know.
6.	Maria	You know, the fins on the top of the back.
7.	Alex	Yeah, yeah, yeah.
8.	Jordan	Maybe it helps them swim better in the Amazon.
9.	Alex	Yeah, yeah, that's probably it.
10.	Laila	Maybe -- yeah, maybe if they had a long fin they would get caught in reeds.
11.	Maria	Yeah.
12.	Laila	And then they would be like [makes face/gesture].
13.	Alex	Yeah, because their top could get caught.
14.	Jordan	And it's really [gesture] like that --
15.	Alex	He makes it taller--oh from the height, from the stomach -- from the bottom of the stomach to the back.
16.	Jordan	And maybe to keep them hidden.
17.	Trevor	Yeah, yeah.
18.	Laila	Yeah, and also, if like, it hits a tree, if the dolphin hits a tree, like it said, those ants in the tree would come out to defend their house.
19.	Alex	Yeah.
20.	Jordan	It would be quick to get away, because they couldn't see the fin, and they'd be --
21.	Alex	Yeah, and people might think it's a pink shark. Like, just like, if it was not in the Amazon.
22.	Laila	How about the ants? Attacking the thing -- what if the dolphin hit a tree?
23.	Skaki	I don't think the ants would go into the water.
24.	Trevor	Yeah, I don't think that happens.
25.	Jordan	But it said they would do anything to defend their tree.
26.	Maria	Yeah, and they would even go in their canoes. So they would probably go in the water.
27.	Laila	Yeah, so they have pretty good jumping --
28.	Skaki	They'd drown.
29.	Alex	They, they -- they jump.
30.	Laila	They're not afraid. We're not talking about the hazard that the ant face, we're talking about the ant will face, like, what they won't do.
31.	Alex	Like you say, they -- because if a tree's piece of bark fell, they could jump on that, maybe. They have their own inventions, like boats, maybe.
32.	Trevor	I doubt that, Alex. Ants do not have any brains.
33.	Jordan	Yes, they do, small ones.
34.	Alex	Yeah, but it's still possible that they could, um -- they could still think of ways to go across, like canoes. If they think -- if -- even if they don't have brains, they can still say that -- that is a good way to travel through the water.
35.	Trevor	I don't think -- this is just what I say -- I don't think ants have thoughts.
36.	Laila	I do think ants have thoughts, because they build and work together. First, here's my -- here's my -- here's my, uh, claim. Ants do have minds. My reasons are they move around --
37.	Alex	Yeah, they have to think, like, do they want to go left or right, forward or backwards?
38.	Laila	My evidence is that -- that, like, lots of books say how ants have to work together to fight off animals, because they have to have brains to work together.
39.	Alex	Yeah, they have to have brains to defend their home.
40.	Trevor	Well, still, they have very little brains.
41.	Alex	But that doesn't mean they don't have their own defense systems.
42.	Trevor	I still don't think the ants would build a boat.
43.	Alex	Well, they wouldn't make one. They might, um -- like, they might see a piece of bark floating, so they jump on it.
44.	Laila	They will use something as a flotation device. A stick.
45.	Alex	Yeah.
46.	Jordan	Maybe like a leaf.
47.	Laila	Yeah, a leaf, maybe.
48.	Alex	Cause remember, they can jump. Because they jump out from the top, too.
49.	Trevor	Anyway, what would the leaf do? What do you think?
50.	Trevor	It's a lot of ants -- if one inch ant, and a lot of them kept jumping on the leaves -- it will eventually sink.
51.	Jordan	Only one or two could.
52.	Laila	Who said the ants couldn't swim?
53.	Alex	Yeah, yeah, yeah. Yeah, that's the answer.
54.	Laila	Some ants can swim.
55.	Alex	Yeah, there are some ants, like water ants.
56.	Laila	Also, I've seen ants fall into, like, pools, and they float.
57.	Alex	Yeah, they do float. Because they --
58.	Mrs. Glen	[laughs] They float because they're tiny. And then they die.
59.	Alex	And then, when these two dolphins bumped into the tree, the dolphin would be pretty close. So then they have their chance to jump onto the dolphin and start attacking.
60.	Skaki	Yeah, they float, and then they die.
61.	Laila	It's -- or, like its thing -- its fin stuck in a tree, if they had a fin.
62.	Mrs. Glen	You might want to go back real quick to something Jordan said. Jordan said, maybe they don't have a fin so they can hide, right?
63.	Jordan	Yeah, from like, predators or something.
64.	Laila	Yeah, ants.
65.	Trevor	Yeah, cause with, um, like, dolphins, you can always see where they're going, because their fin is sticking up.
66.	Alex	And maybe if there's some, like other things sunlight can be shining on it, but that one, the sunlight would go through that spot where it, it should have been. But it would miss the thing, so it might have -- so it's going to be hard to see.
67.	Jordan	Maybe it doesn't like sunlight, and it doesn't like it -- and if it had a fin, it would get hot.

Figure 2. Extended discourse excerpt from Time 8 discussion.

Using the coding manual and Studiocode software, coders watched and listened to the videos in order to identify *teacher-initiated discourse elements* (i.e., authentic question, test question, uptake question, high-level thinking question, speculation question, affective question, intertextual question, shared knowledge question, and teacher discourse moves) in order to detect teachers' release of control to students over time. In addition, coders identified *student-initiated discourse elements* (i.e., authentic question, test question, uptake question, high-level thinking question, speculation question, affective question, intertextual question, shared knowledge question, elaborated explanation, and exploratory talk) in order to identify changes in students' critical-analytic thinking over time.

Specifically, based on the coding manual, four rounds of discourse element coding were conducted on the discourse data. First, question events (i.e., a question and all responses to that question) were identified and given a primary question code of either test or authentic (i.e., the primary question codes were mutually exclusive), and the question event was identified as being either teacher-initiated or student-initiated. Next, secondary codes were applied to question events as applicable: (a) Test questions only could have a secondary code if the event elicited uptake, and (b) authentic questions could have multiple secondary codes based on the responses that they elicited. Secondary codes were applied to the question event if the question

Turn	Speaker	Discourse	QT Codes
1.	Mrs. Glen	How is her father different from Reba Jo, Gabriel?	TQ
2.	Gabriel	He, um, Reba Jo doesn't like the toad that's in the bet, but her dad, her father, says, "A bet is a bet."	
3.	Mrs. Glen	Very good, so Reba Jo is not a huge fan of this toad. But her father says, "A bet's a bet," and what else does the father do? Alex?	TM
4.	Alex	He actually opens the door and lets him in.	
5.	Mrs. Glen	Very good. He opens the door and lets him in, respects the deal that was made. Tomika?	TM
6.	Tomika	Um, Reba Jo doesn't like, um, the deal and maybe, um, her father says, "Well, some deals, maybe I had some deals when I was little, but a deal is a deal and we have to do it."	
7.	Mrs. Glen	Very good. I like how you used the words that he used. Good. A deal is a deal; you have to respect a deal that is made. Good. So, why do you think the author chose to use, uh, like the words that people say? Like the dialect? Uh, so why do you think the author chose to use, like, "Hello, Señor?" Why do you think the author chose to do that? What do you think? Laila?	AQ/SQ TM
8.	Laila	Because now he lives really far out, and then he could be, like, somewhere, in --	
9.	Mrs. Glen	So like the location?	
10.	Laila	Yeah. It's in Spanish.	
11.	Mrs. Glen	It's in Spanish, okay. And --	
12.	Seth	Uh, they could be in Mexico or the West --	
13.	Mrs. Glen	It could be because of their location.	
14.	Seth	And that's like, um, that animal is probably from Mexico and he probably like, um, walked there or something.	
15.	Mrs. Glen	So it has -- go ahead, Sofia.	
16.	Sofia	Um, maybe the author thought it would be entertaining -- er, interesting -- to put the toad speaking in Spanish.	
17.	Mrs. Glen	Good! Now, do -- toads -- if I -- you know, there is a toad over at the preschool if anyone remembers. Would he speak to me if I started talking to him?	TQ
18.	Class	[mixed responses of "no" and yes"]	
19.	Mrs. Glen	[laughs] No, so the toads don't speak to us, right?	
20.	Evan	Yes, they do.	
21.	Mrs. Glen	So, what does this make this story when the toad is speaking? Remember when we talked about our genres? What type of genre does that make this story? Andre?	TQ
22.	Andre	A realistic fiction?	
23.	Mrs. Glen	Is it realistic? Could we actually talk to a toad?	TM
24.	Seth	No.	
25.	Andre	It's fiction.	
26.	Mrs. Glen	So it's just fiction. It's unrealistic. It couldn't actually happen.	

Figure 3. Sample of coded discourse from baseline, business-as-usual discussion. AQ = Authentic Question; EE = Elaborated Explanation; SQ = Speculation Question; TM = Teacher Move; TQ = Test Question.

elicited responses that indicated students' uptake, high-level thinking, speculation, affective, intertextual, and/or shared knowledge. Third, within question events, students' responses were examined for evidence of individuals' elaborated explanations or co-constructed exploratory talk instances.

Turn	Speaker	Discourse	QT Codes
1.	Nakia	Um, yeah, um, what if, what if Jim's brother, um, lived from the disease?	AQ/SQ
2.	Felipe	I think that would change a lot, because then he wouldn't be able to — he wouldn't, um, have gone back home, and then he wouldn't have, like, you know how he went to a whole bunch of different schools, he wouldn't have made it to Carlisle. And then he wouldn't —	ET EE
3.	Nakia	Like, played football.	
4.	Felipe	— he may not have been a athlete, like a famous athlete.	
5.	Louis	Well, he still could have made it to Carlisle.	
6.	Felipe	Yeah.	
7.	Louis	There's still possibilities that he joined the Haskell — or, no, no. They — or, that he's just —	
8.	Felipe	Yeah, I'm not saying that he — I'm just saying that there could have been a chance that he might have not made it, so, because he didn't go to the other, all the other schools, he might have not have made it to Carlisle because —	
9.	Louis	Yeah. Yeah.	
10.	Felipe	But I'm guessing he would be — I mean, even if he didn't make it to the famous athlete because of that, he would still be a lot more happy with that his brother lives. And he didn't —	
11.	Louis	Yeah.	
12.	Felipe	Besides his brother would probably encourage him to do, like, what he was always encouraging him in the story. So, he probably was encouraging him to do, um, to, like, maybe become — try out for, like, football, or something like that.	
13.	Tanya	Yeah, but I do agree with you, but also that because he went home, then he moved to a different school, and at the different school, that's when he saw a football game and decided that he thought he liked it. So, he might not have realized that he liked sports, if he, if his brother didn't die, and he didn't move to there.	EE
14.	Felipe	Well, if he eventually went to a different school, like, a, a different college school, he would probably have seen another football game anyways, but he would be older and different. And so, I think he would have heard of football sooner or later.	
15.	Tanya	Mm-hmm. Yeah. And, it said that his brother was always encouraging him to, um, to keep going in school, because he always had a hard time in school —	
16.	Felipe	Yeah.	
17.	Tanya	— and wanted to quit, but his brother —	
18.	Felipe	But if he was encouraging him in school, don't you think he would encourage him into other, um — at other stuff too.	
19.	Louis	But, yeah, if he was encouraging him to be in school, and doing good, to do good in school, then, then, like, a bigger, um, college would want him, maybe want him to — want him on, in their college. And he would, he might, for a different reason, but he still might be able to play on their football team, too.	

Figure 4. Sample of coded discourse from Time 13 discussion. AQ = Authentic Questions; EE = Elaborated Explanation; ET = Exploratory Talk; SQ = Speculation Questions.

Fourth, teachers' responses were coded for their use of discourse moves (e.g., prompting, marking, or summarizing). As a case in point, an authentic

question asked by a student could elicit over a dozen different responses by students, some offering intertextual connections and others revealing evidence of high-level thinking. Thus, this example question event would have the primary question code of authentic, noted as student-initiated, and also have secondary codes of intertextual and high-level thinking. Further, within the question event, one or more of the students' responses could be coded as an elaborated explanation if they responded in a manner that included at least two independent pieces of reasons or evidence in support of a claim; teachers' responses could also be coded if they were supporting or scaffolding students' discourse. Thus, while there may be overlapping portions of codes in the discussion, each coding category (e.g., questions or responses) was conceptually independent as designated by the rules of the coding manual (Murphy et al., 2017).

Output from the software enabled frequency counts of each of the coded discourse elements (e.g., authentic questions) for both teacher-initiated and student-initiated elements. Two coders, trained by the first author, initially double-coded all question events (i.e., test and authentic questions along with any applicable secondary codes), responses (i.e., elaborated explanations and exploratory talk), and teacher discourse moves in the discussions. They then met to discuss and reconcile any discrepancies between their codes. During this reconciling period, the coders derived a new set of fully agreed upon codes (i.e., reconciled codes) for all discourse elements. Individual coders' discourse codes were then compared to this set of reconciled codes, and interrater agreement with the reconciled codes was calculated for each rater. Further, during interrater calculation, agreement for each discourse element was hand checked to ensure no discourse element had markedly low agreement. After both coders exceeded 80% agreement with the reconciled codes, the coders independently coded the majority of the remaining discussions. Interrater reliability was periodically checked to protect against drift and to ensure continued coder consistency; both coders' overall average agreement exceeded 90%. In this way, interrater calculation took into consideration both the consistency between coders as well as that they both independently adhered to the coding manual for all discourse elements.

Basic Comprehension

Research-designed assessments were created to evaluate students' basic comprehension of text. The development of the basic comprehension measure followed a protocol that integrated the guidelines delineated in Popham (2006). Test items were developed to reflect cognitive targets specified in the national standards (i.e., locate/recall and integrate/interpret; National Assessment Governing Board [NAGB], 2013). A table of specifications was then constructed to facilitate the alignment between the items designed for the basic comprehension measure and the cognitive targets stressed in

Week 5: The Man Who Named the Clouds
Score: 2

Week 14: Gift from the Heart
Score: 6

#37

Part One: Multiple Choice
Directions: Circle the best answer.

- Luke was happy after becoming the manager of a large chemical factory because:
 - a. he had a great paying job.
 - b. he was a rich and famous scientist.
 - c. he could teach students about cloud-naming.
 - d. he was able to return to studying clouds.
- How did Luke's knowledge of Latin help him study clouds?
 - a. He was able to communicate with all the other scientists studying clouds.
 - b. Latin enabled him to come up with his famous cloud-naming system.
 - c. He was able to translate other people's systems into English.
 - d. Latin enabled him to learn Lennack's French cloud-naming system.

Part Two: Short Answer Questions
Directions: Answer the questions and support your answers with details from the text.

- How did Luke's weather journal help him learn about the clouds?
 he kept track of what they were like
- What influence did Luke's father have on Luke becoming a chemist?
 Luke wanted to be like his father
 Luke having a good job.
- How did the Aztecian Society help Luke to write the essay on the classification of clouds?
 they translated for him
 to [translation]

Part One: Multiple Choice
Directions: Circle the best answer.

- Except for Little One, why did the rest of the Comanche People keep their prized possessions?
 - a. They were afraid of the Great Spirit.
 - b. They wanted to make music for the Great Spirit.
 - c. They wanted to show Wise Eagle they were strong.
 - d. They were too selfish to sacrifice for the People.
- What made Little One's doll so special to her?
 - a. The Great Spirit gave her the doll.
 - b. The doll was worth a lot of money.
 - c. The doll reminded her of her parents.
 - d. Her grandmother made the doll.

Part Two: Short Answer Questions
Directions: Answer the questions and support your answers with details from the text.

- Why did the Great Spirit want a special gift from the People?
 to show they praise the greater spirit
- How did the lives of the Comanche People depend on rain?
 for water crops and animals
- Why did Little One have a hard time sleeping before she sacrificed her doll?
 she was worried

Score: 5

Score: 5

#38

Part One: Multiple Choice
Directions: Circle the best answer.

- Luke was happy after becoming the manager of a large chemical factory because:
 - a. he had a great paying job.
 - b. he was a rich and famous scientist.
 - c. he could teach students about cloud naming.
 - d. he was able to return to studying clouds.
- How did Luke's knowledge of Latin help him study clouds?
 - a. He was able to communicate with all the other scientists studying clouds.
 - b. Latin enabled him to come up with his famous cloud-naming system.
 - c. He was able to translate other people's systems into English.
 - d. Latin enabled him to learn Lennack's French cloud-naming system.

Part Two: Short Answer Questions
Directions: Answer the questions and support your answers with details from the text.

- How did Luke's weather journal help him learn about the clouds?
 He described what he saw in the clouds.
 He described what he saw in the clouds.
- What influence did Luke's father have on Luke becoming a chemist?
 His father liked Luke becoming a chemist.
 His father liked Luke becoming a chemist.
- How did the Aztecian Society help Luke to write the essay on the classification of clouds?
 they all shared their ideas and thoughts about
 They all shared their ideas and thoughts about the clouds.

Part One: Multiple Choice
Directions: Circle the best answer.

- Except for Little One, why did the rest of the Comanche People keep their prized possessions?
 - a. They were afraid of the Great Spirit.
 - b. They wanted to make music for the Great Spirit.
 - c. They wanted to show Wise Eagle they were strong.
 - d. They were too selfish to sacrifice for the People.
- What made Little One's doll so special to her?
 - a. The Great Spirit gave her the doll.
 - b. The doll was worth a lot of money.
 - c. The doll reminded her of her parents.
 - d. Her grandmother made the doll.

Part Two: Short Answer Questions
Directions: Answer the questions and support your answers with details from the text.

- Why did the Great Spirit want a special gift from the People?
 The Great Spirit wanted a special gift from the People, because he or she wanted to know that they wanted the best for their people.
- How did the lives of the Comanche People depend on rain?
 The lives of the Comanche People depend on the rain, because they need water and plants to survive.
- Why did Little One have a hard time sleeping before she sacrificed her doll?
 Little One had a hard time sleeping before she sacrificed her doll, because she was so caught up in the Great Spirit wanting a gift.

Figure 5. Basic comprehension measure example for two students at Week 5 and Week 14.

the reading framework for NAEP (NAGB, 2013). The posttests included two selected-response questions that required simple inference and three constructed-response items that tapped into students' ability to make complex inferences (see Figure 5). Careful consideration was given to ensure that all complex inference questions could elicit multiple idea units. A basic comprehension assessment was administered after students discussed each

text, with each assessment following the aforementioned specifications and format but tailored to the text's content.

The scoring of selected-response items on pretests and posttests were scored as either correct (1 point) or incorrect (0 point). The constructed-response items were scored based on the number of correct idea units present in the response on an item-to-item basis, up to two points per question. Posttest total scores included the selected-response and constructed-response item scores for a total of eight possible points. All basic comprehension measures were independently scored by two raters (interrater reliability average per text = 80%), and all discrepancies were discussed and resolved.

High-Level Comprehension Measure

Writing prompts were created by the authors and were based on the content of the main selection text that students read as part of their existing language arts curriculum and discussed as part of QT (e.g., "Encantado: Pink Dolphin of the Amazon"). The prompts were printed in the literacy journals for students to complete after each discussion (see Figure 6). For each of the written high-level comprehension measure prompts, the authors derived a question that required students to consider and weigh at least two positions. Students were prompted to choose a position and to provide supporting reasons and evidence (e.g., "Would you like to travel to the Amazon with a guide or not? Provide reasons and evidence to support your answer."). Thus, the high-level comprehension measure aimed to assess the cognitive target of critique/evaluate (NAGB, 2013) specifically with respect to the transfer effects of QT to students' writing. Like the basic comprehension measure, the high-level comprehension measure was administered after student discussion of each text.

Students' writing was scored based on a scoring rubric, which was developed in alignment with the argumentation schema taught to participants in the QT argumentation mini-lessons (i.e., claim, reason, evidence, counterargument, and rebuttal). Students earned points based upon the quality of their argument (e.g., arguments with a claim, reason, and evidence earned more points than arguments that were missing one or more of those components), counterargument, and rebuttal. All responses were independently scored by two raters, trained by the second author, and all discrepancies were discussed and resolved (interrater reliability average per text = 76%).

Procedures

Teachers, parents, and students were consented/assented at the beginning of the school year. Baseline data on teachers' typical use of discourse elements was collected by coding a business-as-usual discussion in each

Week 5: *The Man Who Named the Clouds***Week 14: *Gift from the Heart***

What personality trait was most important in helping Luke Howard become successful? Provide reasons and evidence to support your response.

Should Little One have sacrificed her doll? Provide reasons and evidence to support your response.

Score: 2

Score: 4

#37

Luke loved sitting
clouds. He was very
sittic. He was
a very smart man

No, she should not
sacrifice her doll cause,
it was her only thing
get from her family also
other people could make her

Score: 4

Score: 6

#38

The most important characteristic trait
that Luke had was being curious. Without
being curious he would have never wondered
about the clouds. And he would definitely
not name the clouds.

Yes, I think she was right to
do what she did. Without her sacrifice
her people might die. Wise Eagle might be
sacrificed his ceremonial stick, but that
might not even work. I think her sacrifice
was one of a kind.

Figure 6. High-level comprehension measure example for two students at Week 5 and Week 14.

classroom. Additionally, baseline data included measures of ORF, basic comprehension, and high-level comprehension. Teachers began implementing the QT model as part of their language arts curriculum immediately after baseline data collection and continued throughout the remainder of the school year. As part of the intervention, each month, teachers delivered one mini-lesson, over four 15-minute lessons or activities approximately once per week, and conducted weekly small-group discussions. Teachers each facilitated three heterogeneous ability discussion groups with five to six students each. In order for the teacher to facilitate all three groups within the approximately 50-minute period allotted to language arts, teachers rotated through the groups, spending approximately 15 minutes per group. While teachers were facilitating one group, the remaining students were quietly engaging in seatwork. Groups were established immediately after baseline based on students' baseline ORF scores. The composition of the groups was honed in the first 2 weeks, after which time they remained largely consistent throughout the year.¹ All mini-lesson instruction and discussions were videorecorded with supplementary audiorecordings serving as backup. Videorecordings of all groups' discussions were coded for discourse elements at approximately equal intervals (i.e., approximately twice a month) across a total of 15 coded discussions or time points. Following each of

the coded discussions, students completed the basic and high-level comprehension measures on the discussed text.

Fidelity of Implementation

We instituted numerous procedures to ensure high fidelity of implementation (Greene, 2015; O'Donnell, 2008). Each professional development workshop was videorecorded and reviewed to ensure all materials were addressed and teacher questions answered. Researchers provided teachers with lesson plans and powerpoint slides for each QT mini-lesson, and reviewed those materials with them, including examples of how to deliver lessons with fidelity. All instructional materials were provided to teachers in .pdf format in order to decrease the likelihood of the materials being altered. Researchers video- and audiorecorded each teacher's mini-lesson instruction and reviewed them to assess fidelity. As stated previously, teachers' discussions with students were also videorecorded and reviewed by researchers. Coaching sessions with the teachers, designed to provide feedback and support for QT based upon the videorecordings of discussions, were also videorecorded and reviewed to ensure that coaching aligned with the QT model.

Our analyses indicated a high degree of implementation fidelity. We analyzed each of the six QT mini-lessons for main components, identifying 74 in total per teacher. Then, we watched all recordings that corresponded to each teacher's delivery of the QT mini-lessons, coding for the presence or absence of each component. Across the video and audiorecordings, there were only three times where researchers felt the teachers did not cover the material as expected. Variance in implementation across the teachers was limited to relatively minor differences in time spent on certain parts of each lesson, the degree to which the teacher asked students questions during QT mini-lesson instruction, and some of the teacher-generated additional examples used to illustrate the main components or answer student questions. In sum, we felt the teachers and researchers achieved high fidelity of implementation for this study.

Results

Changes in Teacher Discourse

Our first research question concerned the ways in which teachers' involvement in QT discussions changed over the course of the intervention. Across the 15 time points, from the baseline to the final discussion session, trained graduate students coded all six discussion groups' videos for both the quantity and types of discourse elements exhibited by the two teachers. Aggregating across teachers, we found a sharp decrease in the teachers' use of test questions, as predicted (see Table 2). Likewise, the predicted

Table 2
Frequencies of Teacher Quality Talk Discourse Elements by Time Point

	AQ	TQ	UT	HLT	SQ	AFQ	IQ	SKQ	TM
Baseline	19	49	4	1	3	0	0	0	82
Time 1	33	19	8	5	3	0	0	0	48
Time 2	40	15	9	6	6	0	0	0	44
Time 3	35	17	13	4	4	3	0	1	47
Time 4	25	26	13	4	0	0	0	0	27
Time 5	36	22	8	6	3	1	0	0	33
Time 6	24	6	10	1	2	5	0	0	33
Time 7	15	5	8	1	1	0	0	0	41
Time 8	22	16	11	0	2	0	3	0	46
Time 9	27	14	10	0	2	1	0	0	50
Time 10	15	4	8	3	2	7	0	0	51
Time 11	4	4	8	0	1	0	0	0	38
Time 12	13	6	13	0	0	1	0	0	60
Time 13	21	8	13	1	0	0	4	1	39
Time 14	16	2	10	0	0	1	1	1	32

Note. AFQ = Affective Questions; AQ = Authentic Questions; HLT = High-Level Thinking Questions; IQ = Intertextual Questions; SKQ = Shared Knowledge Questions; SQ = Speculation Questions; TM = Teacher Moves; TQ = Test Questions; UT = Uptake.

profound increase in the use of authentic questions, from baseline to Time 1, was followed by a gradual decrease in the frequency of teacher questioning through the end of the intervention (see Figure 7). As expected, over the course of the intervention, the teachers asked fewer test questions and modeled the use of authentic questions, gradually fading their engagement over time to allow students to gain control of the discourse.

Teachers' use of other types of questions was, for the most part, sporadic over the course of the intervention. This was to be expected given their role as facilitator of the discourse, rather than a leader of that discourse. Overall, total coded teacher moves decreased dramatically between baseline and Time 1 and then showed a general downward trend through the end of the intervention, as expected (see Table 2 and Figure 8). The increase in the frequency of teacher moves at Time 12 coincided with the mini-lesson instruction of more complex argumentation discourse elements (e.g., rebuttals), suggesting that teachers may have felt the need to engage more often at this point to guide student use of these elements.

Changes in Student Discourse

Our second research question addressed predicted changes in the type and frequency of students' critical-analytic thinking as evidenced in discourse. Aggregating across the six discourse groups (i.e., three discussion

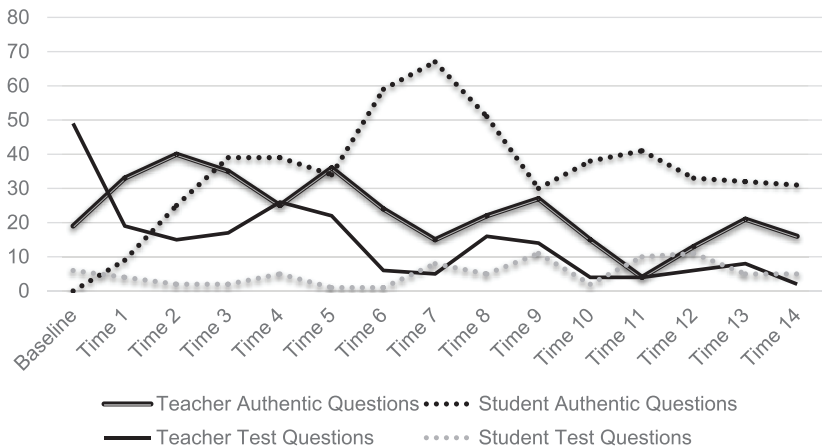


Figure 7. Line graph showing the frequencies of teacher and student authentic and test questions by time point.

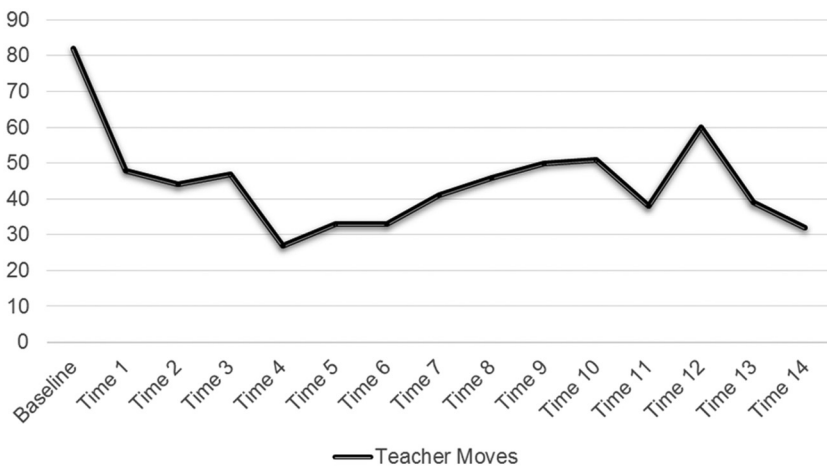


Figure 8. Line graph showing the frequencies of teacher moves by time point.

groups per each of the two teachers) and examining the trends over the 15 time points, we found an expected increase in the use of authentic questions, from none at baseline to a high of 67 coded instances at Time 7, and then a tapering off to an average of 37 per time point over the remainder

Table 3
Frequencies of Student Quality Talk Discourse Elements by Time Point

	AQ	TQ	UT	HLT	SQ	AFQ	IQ	SKQ	EE	ET
Baseline	0	6	0	0	0	0	0	0	6	0
Time 1	9	4	4	0	3	0	0	0	26	4
Time 2	25	2	4	6	1	0	0	0	19	4
Time 3	39	2	11	1	6	1	0	0	35	6
Time 4	39	5	12	3	7	1	0	0	21	3
Time 5	34	1	13	4	8	3	0	0	30	8
Time 6	59	1	19	2	8	12	0	0	24	10
Time 7	67	8	22	1	8	8	0	0	36	4
Time 8	51	5	21	2	11	6	0	0	17	8
Time 9	30	11	9	1	5	5	0	0	22	9
Time 10	38	2	13	0	2	7	0	0	28	14
Time 11	41	10	19	0	4	7	0	0	23	15
Time 12	33	11	17	1	2	1	0	0	28	13
Time 13	32	5	10	2	2	2	5	2	31	10
Time 14	31	5	8	0	1	3	5	7	22	13

Note. AFQ = Affective Questions; AQ = Authentic Questions; EE = Elaborated Explanations; ET = Exploratory Talk; HLT = High-Level Thinking Questions; IQ = Intertextual Questions; SKQ = Shared Knowledge Questions; SQ = Speculation Questions; TQ = Test Questions; UT = Uptake.

of the intervention (see Table 3 and Figure 7). We predicted such a trend in the use of authentic questions from baseline to Time 7, because the first half of the QT intervention focused on the questioning mini-lessons. The subsequent reduction in the frequency of authentic questions after Time 7 coincided with the switch from asking effective questions to providing comprehensive and productive responses through argumentation, which in our case were categorized as elaborated explanations. As shown in Figure 9, the frequency of elaborated explanations increased from an average of one coded instance per group at baseline to a high of six per group at Time 7. As students switched their focus towards elaborating upon responses to authentic questions, the frequency of authentic questions asked decreased accordingly.

Parallel to the rise in the frequency of coded elaborated explanations over time, we found an expected increase in coded instances of exploratory talk. The positive trend of elaborated explanations plateaued around Time 10, commensurate with an increase in exploratory talk. Again, this was predicted, as more frequent instances of exploratory talk (i.e., longer and more complex discussions about an authentic question involving challenges and critiques of elaborated explanations) limits floortime for additional

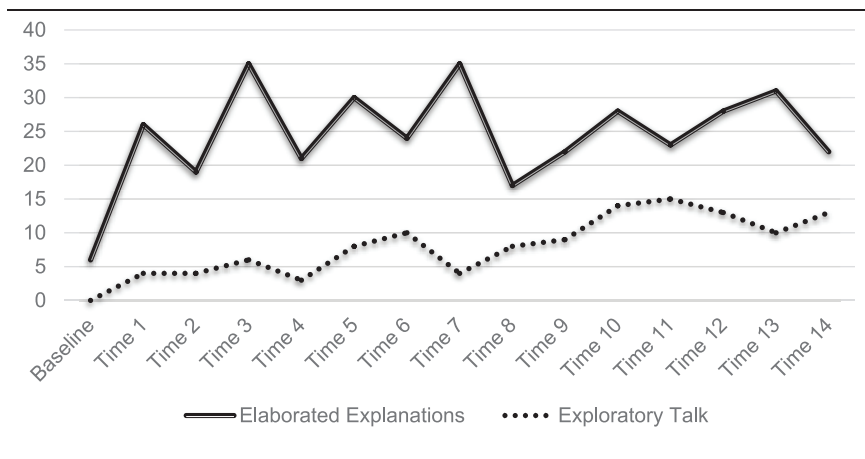


Figure 9. Line graph showing the frequencies of student elaborated explanations and exploratory talk by time point.

questions. In essence, over time, and particularly after Time 8, students talked in more depth about fewer questions.

Student Basic Comprehension Performance

We predicted that positive changes in the frequency and quality of student discourse would parallel increases in student performance on our written comprehension measures, from Time 5 (i.e., after the instruction on questioning) through the end of the QT intervention. Student mean basic comprehension scores showed a general upward trend (see Table 4). We utilized multilevel modeling to examine changes in basic comprehension scores over time, given that scores were nested in students. We had 35 students (i.e., Level 2 units) with no missing data on Level 2 predictors (i.e., gender and AIMSweb score). We had 341 Level 1 scores, meaning that there were only 9 units of missing data, which appeared to be missing completely at random (i.e., no clear missingness mechanism, particularly given the low percentage of missing data; Graham, 2009). All multilevel modeling was conducted using the program HLM version 7.01 (Raudenbush, Bryk, & Congdon, 2010) with restricted maximum likelihood estimation. Correlations among Level 1 and Level 2 predictors were moderate to small (see Table 5).

The intraclass correlation coefficient (ICC) for the basic comprehension outcome variable was .18, indicating that 18% of the variance in scores was due to student differences and therefore warranted a multilevel modeling approach to these data. We lacked sufficient numbers of groups or classes

Table 4
**Descriptive Statistics for Comprehension and Written
 Argumentation Outcome Variables**

Time point	Comprehension		Written Argumentation	
	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>
Baseline	35	6.89 (0.83)	35	3.74 (1.04)
1 ^a	35	5.97 (1.38)		
2 ^a	34	6.38 (1.21)		
3 ^a	35	4.74 (1.20)		
4 ^{a,b}				
5	35	4.00 (1.24)	34	2.82 (1.59)
6	32	4.50 (1.05)	35	4.03 (1.18)
7	35	5.20 (1.18)	35	3.63 (0.97)
8	34	4.88 (1.34)	33	3.61 (1.80)
9	33	4.61 (1.37)	33	3.33 (0.99)
10	34	5.59 (0.99)	34	3.79 (1.65)
11	34	5.68 (1.12)	34	4.12 (1.47)
12	35	5.46 (1.09)	34	4.06 (1.67)
13	35	6.17 (1.20)	34	4.12 (1.92)
14	35	6.09 (1.12)	34	4.03 (0.94)

^aWritten argumentation responses were unavailable at these time points.

^bComprehension measure was not administered at this time point.

Table 5
Correlation Matrices

Level 1 Correlations	Time	Comprehension
Comprehension	.456*	-
Argumentation	.181*	.098
Level 2 Correlation	Gender	
AIMS	.104	

* $p < .01$.

(i.e., six and two, respectively) to include additional levels of analysis, but we did include these variables as Level 2 predictors and found them to be statistically nonsignificant.² Therefore, we did not investigate these variables further.

We scaled the Time variable such that the first time point was coded as 0 and the last as 9. This allowed us to interpret the intercept as the average score at the first time point, after accounting for student-level variables, if applicable. We did not center the Time variable, nor did we center any

Level 2 predictor variables. We posited a linear growth trajectory for the Time variable, and initially we expected both the Level 1 intercept and Time variable to have random effects. The results from this model, however, revealed that the Time variable's variance component was statistically non-significant and quite small. Therefore, we decided to treat Time as a fixed effect. On the other hand, the intercept's variance component was sufficiently large, and statistically significant, to warrant treating the intercept as a random effect.

We utilized a model-building approach (Raudenbush & Bryk, 2002) to investigate Level 1 and Level 2 predictors (see Table 6). A Time-only model revealed a positive growth rate for basic comprehension scores over the course of the QT intervention, and this finding persisted through the investigation of Level 2 predictors. We entered each Level 2 predictor individually and then together. The best-fitting model, based upon deviance values and statistical significance, was the full model with both gender and AIMSweb scores entered as Level 2 predictors of intercept variance. The full model (i.e., Model 4 in Table 6) indicated that, on average, females scored higher than males at the first time point and that higher AIMSweb scores were associated with higher basic comprehension scores at the first time point. The most critical finding in terms of our research question was that, on average, participants' basic comprehension scores increased by .21 over each of the 10 time points. This translated into an average gain of 2.1 points on our basic comprehension measures, which had a total possible score of 8 points. The *R*-squared estimate for this model, using a formula that divides the change in within-student variance between the null and final model by the within-student variance of the null model, is .28, which converts into a Cohen's *d* value of 1.25, a large effect. These findings parallel the predicted growth in positive student critical-analytic thinking, as evidenced in discourse, over the entirety of the QT intervention.

Student High-Level Comprehension Performance

Our analysis of students' high-level comprehension performance was conducted in a manner parallel to our analysis of students' comprehension performance. As shown in Table 4, students' average written high-level comprehension scores increased over time after Time 5. For this analysis, we had missing data at Level 1 in the form of 10 missing data points. Again, these data were treated as missing completely at random given their low percentage of the total data and dispersed nature across participants. All multilevel modeling was conducted using the program HLM version 7.01 (Raudenbush et al., 2010) with restricted maximum likelihood estimation.

The ICC for the written high-level comprehension outcome variable was .07, indicating that 7% of the variance in scores were due to student differences and therefore warranted a multilevel modeling approach to these

Table 6
Multilevel Models for Comprehension Outcome Variable

Variable	Model ₀		Model ₁		Model ₂		Model ₃		Model ₄	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Fixed effects										
Intercept	5.22***	.12	4.26***	.15	3.94***	.18	3.10***	.39	2.94***	.44
Gender					.59**	.21			.54*	.20
AIMS							.008**	.004	.007*	.003
Growth rate			.21***	.02	.21***	.02	.21***	.02	.21***	.02
Random effects										
Intercept	.33***	.03	.37***	.03	.29***	.03	.31***	.03	.25***	.03
Within student	1.49		1.08		1.08		1.08		1.08	
Deviance	1,145.20		1,049.41		1,045.52		1,055.57		1,048.55	

Note. Model₀ is an intercept only; Model₁ includes time as a Level 1 predictor; Model₂ includes time at Level 1 and gender as a Level 2 predictor; Model₃ includes time at Level 1 and AIMS as a Level 2 predictor; and Model₄ includes time at Level 1, gender and AIMS as Level 2 predictors. All models were estimated with 341 Level 1 units and 35 Level 2 units, and each model had two estimated parameters.

* $p < .05$. ** $p < .01$. *** $p < .001$.

data. We investigated whether class and student group were statistically significant Level 2 predictors, and upon finding that they were not, we omitted them from all analyses.

Again, we scaled the Time variable so that the first time point was coded as zero and the last as nine. We did not center any Level 2 predictor variables, nor did we center the Level 1 Time variable, which we posited to have a linear trajectory. Our investigation of the variance components for the intercept and Time variable showed that there was statistically significant variance to model in the intercept but not in the Time variable. Therefore, we fixed the Time variable and used our Level 2 predictors to model variance in initial written high-level comprehension scores.

Our Time-only model showed that, on average, written high-level comprehension scores increased over the course of the QT intervention (see Table 7). We entered each Level 2 predictor individually and then together. Interestingly, while Gender was a statistically significant predictor of initial written high-level comprehension scores, AIMSweb scores were not. The best fitting model, taking into account statistical significance as well as deviance scores, was Model 2 (see Table 6) with the only Level 2 predictor being Gender. The results of this model indicated that, on average, students' written high-level comprehension scores increased by .09 points per time point or .90 points over the course of the intervention on a scale of 0 to 15. The R -squared estimate for this model, using a formula that divides the change in within-student variance between the null and final model by the within-student variance of the null model, is .03. This R -squared converts into

Table 7
Multilevel Models for Written Argumentation Outcome Variable

Variable	Model ₀		Model ₁		Model ₂		Model ₃		Model ₄	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Fixed effects										
Intercept	3.75***	.10	3.34***	.16	3.09***	.17	2.86***	.42	2.72***	.40
Gender					.45*	.18			.43*	.18
AIMS							.00	.003	.00	.002
Growth rate			.09***	.03	.09***	.03	.09***	.03	.09***	.03
Random effects										
Intercept	.15**	.02	.15**	.02	.11*	.01	.15**	.02	.12*	.01
Within student	2.08		2.01		2.01		2.01		2.01	
Deviance	1,232.15		1,227.47		1,221.90		1,234.15		1,232.63	

Note. Model₀ is an intercept only; Model₁ includes time as a Level 1 predictor; Model₂ includes time at Level 1 and gender as a Level 2 predictor; Model₃ includes time at Level 1 and AIMS as a Level 2 predictor; and Model₄ includes time at Level 1, and gender and AIMS as Level 2 predictors. All models were estimated with 350 Level 1 units and 35 Level 2 units, and each model had two estimated parameters.
 * $p < .05$. ** $p < .01$. *** $p < .001$.

a Cohen’s d value of .35, indicating a small to medium effect. This finding was in line with our expectations and provides further support of a positive trajectory in students’ high-level comprehension as evidenced in discourse, individual student basic comprehension, and written high-level comprehension performance over the course of engaging in QT.

Trends Across the Results Pertaining to Students’ Performance

In Research Questions 2, 3, and 4, we examined both the changes in the key indicators of students’ critical-analytic thinking evidenced in the small-group discussions as well as students’ individual performance outcomes on basic and high-level comprehension measures. Across all three research questions, on average, positive growth trends were evidenced as expected. However, because the discourse coding was conducted at the group level and comprehension was assessed at the student level, we were unable to statistically delineate the relationship between the discourse indicators and students’ performance on basic and high-level comprehension measures.

Despite this, we selected two cases to exemplify some of the individual changes evidenced in the discourse as part of the small-group discussions and their subsequent performance on comprehension measures. For illustrative purposes, these two cases were intentionally selected from one of the discussion groups to represent students with both higher and lower initial performance and to showcase a broad range of individual changes over time. In Figure 1, the responses in the Week 5 discussion from Student 37

were short and unelaborated. They primarily consisted of verbal affirmations and repetitions of responses from previous turns and did not contribute to any indicators of critical-analytic thinking. This type of shallow response was also evidenced in Student 37's responses for the Week 5 basic (Figure 5) and high-level (Figure 6) comprehension measures and was signified by low scores on both measures. Yet in the Week 14 discussion, the growth for this student was evidenced by an extended response to the question that included a claim and supporting reasons and evidence; concomitant increases were also evidenced in the student's scores on both basic (Figure 5) and high-level (Figure 6) comprehension. Indeed, the changes in high-level comprehension were particularly prominent, as the claim was not only explicitly stated but also well supported. Likewise, another student, Student 38, initially had more substantive responses in the Week 5 discussion than Student 37, but the counterargument posed in the Week 14 discussion is indicative of continued improvement. Further, their response to the high-level comprehension measure at Week 14 was particularly noteworthy (Figure 6). First, the student appeared to have refined his or her understanding related to the question during or after the discussion, and second, the response included complex notions weighing the sacrifices of characters as part of his or her argument in support of his or her position. In sum, these two cases illustrate patterns in line with our expectations and provide initial support for future research related to the relationships between changes evidenced in the discourse and on comprehension measures.

In addition, the nature and changes of teacher and student talk were also examined over time from the baseline, business-as-usual discussion (Figure 3), to discourse excerpts from midyear (Figures 1 and 2) and, finally, to discourse excerpts occurring at the end of the intervention (Figures 1 and 4). As can be seen in the baseline discussion, conducted prior to QT, the teacher generally posed a question and directed it to one particular student. Following the student response, the teacher evaluated and/or rephrased the answer. Both teachers' business-as-usual discussions evidenced this same IRE pattern: initiating a question, calling upon a student to respond, and evaluating the student answer. While this pattern of questioning may be effective in increasing students' factual understanding or declarative recall from a text, it does not necessarily foster students' high-level comprehension. Through these discussions students are provided with few opportunities to engage in extended episodes of collective reasoning and are left with the impression that only one answer, as affirmed by their teacher, is correct.

In contrast to the business-as-usual discussion, a striking change was evidenced in QT discourse excerpts. In the midyear excerpt (Figure 2), the QT ideal of shared responsibility for talk between teacher and students was particularly evident. After the teacher began the discussion by reviewing the ground rules of QT, she explicitly reminded students to practice engaging in the recently taught discourse elements. Unlike in the baseline

discussion, she invited a student to initiate the discussion with a question (i.e., “Why do the pink dolphins not have dorsal fins?”). The subsequent discussion was centered on students responding to this question as well as other student-initiated authentic questions that were generated organically over the discussion. As part of the discussion, students generated rich discourse as they co-constructed understandings not only *about* the text (i.e., why the dolphins lack dorsal fins) but also *around* and *with* the text (e.g., whether ants think, what evidence supported or refuted the claim that ants can think). Importantly, the teacher facilitated, rather than directed, the discourse using carefully chosen teacher moves to guide students toward engaging in productive talk. For instance, when the discussion diverted too far from the topic, the teacher redirected the focus using procedural and summarizing moves (i.e., “You might want to go back really quick to something Jordan said. Jordan said, ‘Maybe they don’t have a fin so they can hide, right?’”). With increased control over discussion, students had the freedom to make public their thoughts while encountering different points of view in the course of interacting with others, which encouraged them to reexamine their own ideas, to consider new ideas, and to seek more information in order to reconcile the conflicts, leading to higher levels of reasoning and understanding (Piaget, 1932).

Likewise, the discourse described in Figure 2 was representative of QT discourse more broadly. Figure 4 illustrates how students, influenced by their enhanced epistemic cognition, engaged in critical-analytic thinking by elaborating their explanations, supporting their claims with reasons and evidence, challenging each other, generating counterarguments, and critically examining and weighing alternative perspectives. Compared to the midyear excerpt, more incidences of exploratory talk were found as the year progressed. Importantly, while the nested nature of the group discourse data did not afford the ability to statistically examine relationships with individual outcomes, the trends over time with respect to the group-level discourse showed a positive trajectory in line with the reported findings from Research Questions 2, 3, and 4.

Discussion

The literacy challenges of the 21st century, coupled with concerning results from national assessments of K–12 students, have led researchers, policymakers, educators, and parents to call for increased focus upon fostering students’ ability to move beyond basic comprehension by engaging in argumentation to support high-level comprehension, including critical-analytic thinking and epistemic cognition (Greene & Yu, 2016; Iordanou et al., 2016; Lee et al., 2016; Murphy & Alexander, 2016; National Education Association [NEA], 2014; National Governors Association Center for Best Practices [NGA Center], 2010; Organisation for Economic Co-

operation and Development [OECD], 2013). Classroom discussions have been identified as a promising means of modeling and fostering critical-analytic thinking and epistemic cognition about texts, but research has shown that those discussions must involve shared responsibility for talk between teachers and students, explicit instruction on productive discourse and argumentation, and feedback and support for teachers as they use discussion in their classroom (Murphy et al., 2009). QT is a classroom discourse model that leverages findings on productive discourse models to promote high-level comprehension (Murphy et al., 2009; Murphy & Firetto, 2017; Soter et al., 2008).

In this study, we examined an initial, year-long implementation of a revised model of QT, tracking changes in both teachers' and students' discourse, as well as students' basic and high-level comprehension and argumentation. One of the unique aspects of this study is that QT was unfolded in an ecological context, which required us to work closely with the teachers within their existing curriculum. This allowed a more comprehensive understanding of how QT fit within authentic classroom practice, particularly how the teachers employed it within the local context—a facet often lacking in the discourse literature. Over the course of the academic year, we continually trained, supported, and coached two fourth-grade teachers to implement QT with high fidelity. Observing through an ecological lens, we witnessed expected changes in the kinds and frequency of teacher talk with shifts toward more authentic questions, fewer test questions, and fewer teacher moves overall as teachers faded their scaffolding of the discourse. Likewise, the student discourse changed with an increase in authentic questions, a decrease in test questions, and notable increases in students' frequency of engaging in elaborated responses as well as instances of exploratory talk involving group construction of knowledge.

According to Vygotsky (1981), all higher cognitive processes develop from external social interaction and are later incorporated into one's mind. Likewise, students who engage in QT discussions should not only exhibit high-level comprehension in the discourse, but they also should internalize the use of argumentation strategy that may transfer to other learning tasks. As expected, in conjunction with the enhanced discourse and oral argumentation evidenced over time, benefits with respect to students' postdiscussion comprehension and written argumentation performance also emerged. Despite expected differences in initial basic comprehension performance, with females outperforming males on average and fluency being positively related to performance, on average all students' basic comprehension of text improved over time. Average growth in basic comprehension over the course of fourth-grade year (i.e., Cohen's $d = .41$; Scammacca, Fall, & Roberts, 2014) is much less than the average growth we found for our QT students (i.e., Cohen's $d = 1.25$). With no control group, causal claims are not warranted, but this vast difference is evidence of QT's great promise.

Likewise, students' performance on a high-level comprehension and a transfer measure to written argumentation increased in statistically and practically significant ways. The overall ES estimate for our argumentation measure (i.e., $ES = .350$) was larger than those Murphy and colleagues (2009) found for comparable interventions such as P4C (i.e., $ES = .214$) and CR (i.e., $ES = .260$). Again, causal statements are not warranted given our design, and comparisons with intervention studies must acknowledge the longitudinal, not comparative, nature of our work. Nonetheless, given recent NAEP findings, the growth in argumentation performance over the course of QT was striking. Overall, these findings cohere and extend other empirical studies of QT (e.g., Wilkinson et al., 2008), indicating that it is a promising model for promoting the kinds of high-level comprehension skills needed in the modern world.

In this study, our outcome measures (i.e., basic and high-level comprehension) were at Level 1 of our multilevel model (MLM) because they repeated over multiple time points and were nested within students. In contrast, the results pertaining to the changes in teachers' and students' discourse (e.g., changes in teacher moves or changes in frequency of elaborated explanations during group discussion) were also repeated measures but at a different level of nesting: the group. Thus, it was beyond the scope of this manuscript to conduct the complex statistical analyses necessary to establish a direct link between the discourse students engaged in at the group level and each student's individual outcomes (see Curran, Lee, Howard, Lane, & MacCullum, 2012). Despite these limitations, the trends across the discourse codes, basic comprehension, and high-level comprehension for two illustrative cases, evidenced in Figures 1, 5, and 6, indicate that this may be an important area for future research.

Limitations

One strength of our study, the authentic classroom settings that lend the findings ecological validity, is also a limitation: This was a single-group, time-series design that precludes causal claims about the efficacy of QT. Likewise, the promising findings outlined in this study are specific to a particular school in a particular context. QT's design was based upon a thorough review of the literature on classroom discourse (e.g., Murphy et al., 2009), suggesting that it has strong potential for external validity, but this potential was not the focus of our current study.

Our focus upon ecological validity and our commitment to the teachers who partnered with us required us to work within the literacy curriculum that was currently in use by the school. Therefore, our assessments were specific to that curriculum, and we were unable to do the kind of randomization of texts across time points that can control for confounds such as variation in students' prior knowledge or interest. Nonetheless, the long-term,

longitudinal design of our study decreased the likelihood of prior knowledge or interest as alternative explanations for our findings; it seems unlikely that the texts were ordered in such a way that these potential confounds had a monotonic effect over an entire academic year.

Finally, both of our teachers wanted to implement QT in its entirety, and we honored their wishes. Therefore, we were unable to do the kind of implementation fidelity analyses that might have identified the active ingredients of QT or determined the minimal effective dosage (Greene, 2015; O'Donnell, 2008; Warren, Domitrovich, & Greenberg, 2009). At this stage of testing, we had to focus upon whether QT in its most supported and extensive form would have positive effects upon students' high-level comprehension. We believed that such testing was warranted, given the lack of empirical research on classroom discourse models with a focus upon fostering critical-analytic, efferent, and expressive stances toward texts.

Future Directions for Research and Practice

Our study has substantive implications for future research on classroom discourse with many deriving directly from the necessary limitations of this study. QT, and other classroom discourse models, should be studied using experimental designs to establish an estimate of its causal effect upon students' high-level comprehension. Scale-up of QT will require understanding which aspects of the pedagogy are necessary for efficacy and which can be discarded when resources are limited. Such understanding comes from systematic study of the various active ingredients of QT (e.g., teacher coaching or standardized instructional materials). Similarly, while the participating students were from families and communities characterized by relatively lower socioeconomic means, the majority of students were native English speakers. Future investigations of the effectiveness of QT with English language learners or those who speak dialects other than standard English as well as those learning in English language minority contexts would provide foundational understandings regarding the generalizability of QT to other settings or communities. At present, we have only begun to initiate such studies in Taiwan, mainland China, and South Africa (Murphy & Firetto, 2017).

One particular aspect of QT, like many other literacy and classroom discourse models, deserves particular mention. Educators and researchers continue to debate the merit of heterogeneous versus homogeneous ability grouping for literacy discourse (Slavin, 1987; Webb, Nemer, Chizhik, & Sugrue, 1998). In our study, we implemented heterogeneous ability grouping to increase the ecological validity of our study and to honor the preferences of our partner teachers. Further research is warranted to investigate potential differences in student performance depending upon whether homogeneous or heterogeneous ability grouping is used within small-group classroom discourse models such as QT (Murphy, Greene, et al., 2017). It

should also be noted that an examination of the impact of gender on students' participation and performance in the small-group discussions is a notable area for future study.

Finally, our assessments of argumentation addressed the product of students' discourse (Greene et al., 2016). The quality of the discourse itself, influenced by students' epistemic cognition, is likely a strong determinant of whether students internalize the argumentation skills needed to transfer group discourse products into individual written argumentation products (Iordanou et al., 2016). Therefore, another area of interest for future research is process-oriented: how students' epistemic cognition is enacted and shaped over the course of a long-term classroom discourse model such as QT. An analysis such as this might be conducted by employing other macro- or microanalytic frameworks for examining the changes in students' discourse over their participation in QT (see Elizabeth, Ross Anderson, Snow, & Selman, 2012). Studies of verbal argumentation and epistemic cognition have shown that students do engage in such processing (e.g., Herrenkohl & Cornelius, 2013), but there is a paucity of research on how that processing changes longitudinally as a result of multiple opportunities to engage in structured discourse.


The numerous paths for potentially generative future research does not mean that educators should refrain from implementing changes in their practice based upon current research. This study builds on the growing research base supporting the importance of small-group classroom discussion as a tool to promote student engagement and critical-analytic thinking (Murphy et al., 2009). Despite the lack of experimental research needed to make causal claims, there is compelling evidence that small-group discussion, coupled with explicit instruction, scaffolding, and fading of teacher discourse moves, is a superior method for promoting high-level comprehension, compared to whole-class discussion or lecture-style instruction. Our findings also support the growing literature that guided inquiry learning, where teachers play a supportive role as students lead the process of inquiry, affords students unique opportunities to develop and practice essential skills as well as receive feedback upon how to improve them in a timely and influential manner (Janssen, Westbroek, & van Driel, 2014; Loyens, Kirschner, & Paas, 2012).

Conclusion

The modern world has placed new demands upon students, requiring them to develop the knowledge and skills necessary to be critical consumers of the myriad of print and digital resources available to them (Goldman et al., 2010). Such knowledge and skills are not innate; rather, they must be taught, modeled, and supported in classrooms, starting in elementary school (Bennett, Maton, & Kervin, 2008). Dialogic approaches such as

classroom discussion can provide a rich opportunity to build students' skills in these areas (McKeown et al., 2009; Reznitskaya et al., 2008). However, teachers' extant discourse patterns may not be in alignment with the characteristics of productive discourse that lead to enhanced comprehension and critical-analytic thinking (e.g., shared control with students; Soter et al., 2008). Our research provides initial evidence that QT, with its comprehensive focus on supporting productive discourse among teachers and students, is a promising method of developing students' critical-analytic thinking, epistemic cognition, and subsequent high-level comprehension. Scale-up of promising methods, such as QT, is needed so that all students can be prepared to actively and thoughtfully engage with the complex, multidimensional challenges of the 21st century.

ORCID iD

P. Karen Murphy  <https://orcid.org/0000-0001-8872-0376>

Notes

This research was supported by the IES, USDE through Grant R305A130031 to the Pennsylvania State University. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not represent the views of the Institute or the USDE.

¹In January, one teacher requested shifting three students across two groups in order to better balance the size, heterogeneity, and cohesiveness of the groups. Except for this instance, all students remained in the same groups from Time 3 to Time 14.

²Analyses available upon request to the first author.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: Bolt, Beranek, & Newman.
- Alexander, P. A., Murphy, P. K., Guan, J., & Murphy, P. A. (1998). How students and teachers in Singapore and the United States conceptualize knowledge and beliefs: Positioning learning within epistemological frameworks. *Learning and Instruction, 8*, 97–116. doi:10.1016/s0959-4752(97)00004-2
- Almasi, J. F. (1995). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussions of literature. *Reading Research Quarterly, 30*(3), 314–351. doi:10.2307/747620
- Alvermann, D. E., & Hayes, D. A. (1989). Classroom discussion of content area reading assignments: An intervention study. *Reading Research Quarterly, 24*(3), 305–335. doi:10.2307/747772
- Alvermann, D. E., O'Brien, D. G., & Dillon, D. R. (1990). What teachers do when they say they're having discussions of content area reading assignments: A qualitative analysis. *Reading Research Quarterly, 25*(4), 296–322. doi:10.2307/747693
- Anderson, R. C., Chinn, C., Waggoner, M., & Nguyen, K. (1998). Intellectually stimulating story discussions. In J. Osborn & F. Lehr (Eds.), *Literacy for all: Issues in teaching and learning* (pp. 170–186). New York: Guilford Press.

- Anderson, R. C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., Tillmanns, M., & Gilbert, L. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction, 19*, 1–46. doi:10.1207/s1532690xci1901_1
- Applebee, A., 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 Education Research Journal, 40*(3), 685–730. doi:10.3102/00028312040003685
- Beck, I. L., & McKeown, M. G. (2006). *Improving comprehension with Questioning the Author: A fresh and expanded view of a powerful approach*. New York: Scholastic.
- Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British Journal of Educational Technology, 39*(5), 775–786. doi:10.1111/j.1467-8535.2007.00793.x
- Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education, 95*(2), 191–216. doi:10.1002/sce.20420
- Billings, L., & Fitzgerald, J. (2002). Dialogic discussion and the Paideia Seminar. *American Educational Research Journal, 39*(4), 907–941. doi:10.3102/00028312039004905
- Bråten, I., Anmarkrud, Ø., Brandmo, C., & Strømsø, H. I. (2014). Developing and testing a model of direct and indirect relationships between individual differences, processing, and multiple-text comprehension. *Learning and Instruction, 30*, 9–24. doi:10.1016/j.learninstruc.2013.11.002
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist, 46*(1), 48–70. doi:10.1080/00461520.2011.538647
- Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2014). Students working with multiple conflicting documents on a scientific issue: Relations between epistemic cognition while reading and sourcing and argumentation in essays. *British Journal of Educational Psychology, 84*, 58–85. doi:10.1111/bjep.12005
- Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist, 46*(3), 141–167. doi:10.1080/00461520.2011.587722
- Chinn, C. A., O'Donnell, A. M., & Jinks, T. S. (2000). The structure of discourse in collaborative learning. *Journal of Experimental Education, 69*, 77–97. doi:10.1080/00220970009600650
- Chinn, C. A., Rinehart, R. W., & Buckland, L. A. (2014). Epistemic cognition and evaluating information: Applying the AIR model of epistemic cognition. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 425–453). Cambridge, MA: MIT Press.
- Chiu, M. M., & McBride-Chang, C. (2006). Gender, context, and reading: A comparison of students in 41 countries. *Scientific Studies of Reading, 10*(4), 331–362. doi:10.1207/s1532799xssr1004_1
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1–35. doi:10.2307/1170744
- Curran, P. J., Lee, T., Howard, A. L., Lane, S., & MacCullum, R. (2012). Disaggregating within-person and between person effects in multilevel and structural equation growth models. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 217–253). Charlotte, NC: Information Age Publishing.

- Daniel, M. H. (2010). *Reliability of AIMSweb Reading Curriculum-Based Measurement (R-CBM) (Oral Reading Fluency)*. NCS Pearson, Inc. Retrieved from <http://www.aimsweb.com/wp-content/uploads/Summary-of-Reliability-Information-for-AIMSweb-R-CBM.pdf>
- Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of joint understanding in the classroom*. London: Methuen.
- Elizabeth, T., Ross Anderson, T., Snow, E., & Selman, R. (2012). Academic discussions: An analysis of instructional discourse and an argument for an integrative assessment framework. *American Educational Research Journal*, 49(6), 1214–1250. doi:10.3102/0002831212456066
- Elley, W. B. (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford, UK: Pergamon Press.
- Fuchs, L., Fuchs, D., Hosp, M., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239–256. doi:10.1207/s1532799xssr0503_3
- Goffreda, C., & DiPerna, J. (2010). An empirical review of psychometric evidence for the dynamic indicators of basic early literacy skills. *School Psychology Review*, 30(3), 463–483.
- Goldenberg, C. (1993). Instructional conversations: Promoting comprehension through discussion. *The Reading Teacher*, 46, 316–326.
- Goldman, S. R., Lawless, K. A., Gomez, K. W., Braasch, J., McLeod, S., & Manning, F. (2010). Literacy in the digital world: Comprehending and learning from multiple sources. In M. G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 257–284). New York, NY: Guilford.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576. doi:10.1146/annurev.psych.58.110405.085530
- Graham, S., & Harris, K. R. (2014). Conducting high quality writing intervention research: Twelve recommendations. *Journal of Writing Research*, 6, 89–123. doi:10.17239/jowr-2014.06.02.1
- Great Books Foundation. (1987). *An introduction to shared inquiry*. Chicago: Author.
- Greene, J. A. (2015). Serious challenges require serious scholarship: Integrating implementation science into the scholarly discourse. *Contemporary Educational Psychology*, 40, 112–120. doi:10.1016/j.cedpsych.2014.10.007
- Greene, J. A., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and methodological directions. *Educational Psychologist*, 43(3), 142–160. doi:10.1080/00461520802178458
- Greene, J. A., Sandoval, W. A., & Bråten, I. (2016). Introduction to epistemic cognition. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 495–510). New York: Routledge.
- Greene, J. A., & Yu, S. B. (2016). Educating critical thinkers: The role of epistemic cognition. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 45–53. doi:10.1177/2372732215622223
- Herrenkohl, L. R., & Cornelius, L. (2013). Investigating elementary students' scientific and historical argumentation. *The Journal of the Learning Sciences*, 22(3), 413–461. doi:10.1080/10508406.2013.799475
- Iordanou, K., & Constantinou, C. (2015). Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of Socrates learning environment. *Science Education*, 99(2), 282–311. doi:10.1002/scs.21152
- Iordanou, K., Kendeou, P., & Beker, K. (2016). Argumentative reasoning. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 39–53). New York: Routledge.

- Janssen, F. J. J. M., Westbroek, H. B., & van Driel, J. H. (2014). How to make guided discovery learning practical for student teachers. *Instructional Science*, *42*, 67–90. doi:10.1007/s11251-013-9296-z
- Johnson, E., Jenkins, J., Petscher, Y., & Catts, H. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice*, *24*(4), 174–185. doi:10.1111/j.1540-5826.2009.00291.x
- King, A. (1999). Discourse patterns for mediating peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 87–116). Mahwah, NJ: Erlbaum.
- Kucan, L. (2009). Engaging teachers in investigating their teaching as a linguistic enterprise: The case of comprehension instruction in the context of discussion. *Reading Psychology*, *30*, 51–87. doi:10.1080/02702710802274770
- Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, *15*, 309–328. doi:10.1016/s0885-2014(00)00030-7
- Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentative competence. *Cognition and Instruction*, *31*(4), 456–496. doi:10.1080/07370008.2013.830618
- Lee, C. D., Goldman, S. R., Levine, S., & Magliano, J. (2016). Epistemic cognition in literary reasoning. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of Epistemic Cognition* (pp. 165–183). New York: Routledge.
- Li, M., Murphy, P. K., Wang, J., Mason, L. H., Firetto, C. M., Wei, L., & Chung, K. S. (2016). Promoting reading comprehension and critical-analytic thinking: A comparison of three approaches with fourth and fifth graders. *Contemporary Educational Psychology*, *46*, 101–115. doi:10.1016/j.cedpsych.2016.05.002
- Logan, S., & Johnston, R. S. (2009). Gender differences in reading: Examining where these differences lie. *Journal of Research in Reading*, *32*(2), 199–214. doi:10.1111/j.1467-9817.2008.01389.x
- Loyens, S. M. M., Kirschner, P. A., & Paas, F. (2012). Problem-based learning. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), *APA educational psychology handbook: Vol. 3. Application to learning and teaching* (pp. 403–425). Washington, DC: American Psychological Association.
- Mason, L., & Boscolo, P. (2004). Role of epistemological understanding and interest in interpreting a controversy and in topic-specific belief change. *Contemporary Educational Psychology*, *29*, 103–128. doi:10.1016/j.cedpsych.2004.01.001
- Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction*, *16*, 492–509. doi:10.1016/j.learninstruc.2006.09.007
- McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, *44*(3), 218–253. doi:10.1598/rrq.44.3.1
- Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.
- Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Philadelphia: Multilingual Matters.
- Mercer, N. (2000). *Words and minds: How we use language to think together*. London: Routledge.
- Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence*

- through academic talk and dialogue (pp. 347–362). Washington, DC: American Educational Research Association.
- Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and civic life. *Studies in Philosophy and Education*, 27(4), 283–297. doi:10.1007/s11217-007-9071-1
- Murphy, P. K. (2015). Marking the way: School-based interventions that “work.” *Contemporary Educational Psychology*, 40, 1–4. doi:10.1016/j.cedpsych.2014.10.003
- Murphy, P. K., & Alexander, P. A. (2016). Interrogating the relation between conceptual change and epistemic beliefs. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 439–459). New York: Routledge.
- Murphy, P. K., Alexander, P. A., & Muis, K. R. (2012). Knowledge and knowing: The journey from philosophy and psychology to human learning. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *Educational psychology handbook: Vol. 1. Theories, constructs, and critical issues* (pp. 189–226). Washington, DC: American Psychological Association.
- Murphy, P. K., & Firetto, C. M. (2017). Quality Talk: A blueprint for productive talk. In P. K. Murphy (Ed.), *Classroom discussions in education* (pp. 101–133). New York: Routledge.
- Murphy, P. K., Firetto, C. M., Greene, J. A., & Butler, A. M. (2017). *Analyzing the talk in Quality Talk discussions: A coding manual*. Unpublished manuscript, The Pennsylvania State University, University Park, PA. doi.org/10.18113/S1XW64
- Murphy, P. K., Greene, J. A., Firetto, C. M., Li, M., Lobczowski, N. G., Duke, R. F., ... Croninger, R. M. V. (2017). Exploring the influence of homogeneous versus heterogeneous grouping on students' text-based discussions and comprehension. *Contemporary Educational Psychology*, 51, 336–355. doi:10.1016/j.cedpsych.2017.09.003
- Murphy, P. K., Rowe, M. L., Ramani, G., & Silverman, R. (2014). Promoting critical-analytic thinking in children and adolescents at home and in school. *Educational Psychology Review*, 26, 561–578. doi:10.1007/s10648-014-9281-3
- Murphy, P. K., Wilkinson, I. A. G., & Soter, A. O. (2010). Instruction based on discussion. In R. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 382–407). New York: Taylor & Francis.
- Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' high-level comprehension of text: A meta-analysis. *Journal of Educational Psychology*, 101, 740–764. doi:10.1037/a0015576
- National Assessment Governing Board. (2013). *Reading framework for the 2013 National Assessment of Educational Progress*. Washington, DC: U.S. Department of Education. Retrieved from www.edpubs.gov/document/ed005373p.pdf
- National Education Association. (2014). *Preparing 21st century students for a global society: An educators guide to the “Four Cs.”* Washington, DC: Author.
- National Governors Association Center for Best Practices. (2010). *Common Core State Standards*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.
- Nystrand, M. (2006). Research on the role of discussion as it affects reading comprehension. *Research in the Teaching of English*, 40(4), 392–412.
- Nystrand, M. (with Gamoran, A., Kachur, R., & Prendergast, C.). (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.

- Nystrand, M., Wu, L., Gamoran, A., Zeiser, S., & Long, D. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes, 35*(2), 135–198. doi:10.1207/s15326950dp3502_3
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33–84. doi:10.3102/0034654307313793
- Organisation of Economic Co-operation and Development. (2013). *Trends shaping education 2013*. Paris, France: OECD Publishing. doi:10.1787/trends_education-2013-en
- Pearson, P. D., & Gallagher, M. C. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology, 8*, 20–22.
- Piaget, J. (1932). *The language and thought of the child* (2nd ed.). London: Routledge & Kegan Paul.
- Popham, W. J. (2006). *Assessment for educational leaders*. Upper Saddle River, NJ: Pearson Education, Inc.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2010). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- Reninger, K. B., & Wilkinson, I. A. G. (2010). Using discussion to promote striving readers' higher level comprehension of literary texts. In J. L. Collins & T. G. Gunning (Eds.), *Building struggling students' higher level literacy: Practical ideas, powerful solutions* (pp. 57–83). Newark, DE: International Reading Association.
- Reznitskaya, A., Anderson, R. C., Dong, T., Li, Y., Kim, I. H., & Kim, S. Y. (2008). Learning to think well: Application of argument schema theory. In C. C. Block & S. Parris (Eds.), *Comprehension instruction: Research-based best practices* (pp. 196–213). New York, NY: Guilford Press.
- Reznitskaya, A., Glina, M., Carolan, B., Michaud, O., Rogers, J., & Sequeira, L. (2012). Examining transfer effects from dialogic discussions to new tasks and contexts. *Contemporary Educational Psychology, 37*, 288–306. doi:10.1016/j.cedpsych.2012.02.003
- Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education, 96*(3), 488–526. doi:10.1002/sc.21006
- Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education, 89*(4), 634–656. doi:10.1002/sc.20065
- Scammacca, N. K., Fall, A.- M., & Roberts, G. (2014). Benchmarks for expected annual academic growth for students in the bottom quartile of the normative distribution. *Journal of Research on Educational Effectiveness, 8*(3), 366–379. doi:10.1080/19345747.2014.952464
- Schwarz, B. B., & De Groot, R. (2007). Argumentation in a changing world. *The International Journal of Computer-Supported Collaborative Learning, 2*(2–3), 297–313. doi:10.1007/s11412-007-9020-6
- Shinn, M. R., Good, R., Knutson, N., Tilly, W., & Collins, V. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459–479.
- Shinn, M. R., & Shinn, M. M. (2002). *AIMSweb training workbook: Administration and scoring of reading Maze for use in general outcome measurement*. Eden Prairie, MN: Edformation.

- Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, 57(3), 293–336. doi:10.2307/1170460
- Soter, A. O., Wilkinson, I. A. G., Murphy, P. K., Rudge, L., Reninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal Educational Research*, 47, 372–391. doi:10.1016/j.ijer.2009.01.001
- 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, 3–28. doi:10.1086/499740
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2015). *The nation's report card: 2015 reading state snapshot report*. Washington, DC: Author. Retrieved from <http://nces.ed.gov/nations-reportcard/subject/publications/stt2015/pdf/2016008NP4.pdf>
- Van den Bergh, L., Ros, A., & Beijaard, D. (2014). Improving teacher feedback during active learning: Effects of a professional development program. *American Educational Research Journal*, 51(4), 772–809. doi.org/10.3102/0002831214531322.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 144–188). Armonk, NY: Sharpe.
- Wagemaker, H. (Ed.). (1996). *Are girls better readers? Gender differences in reading literacy in 32 countries*. The Hague, The Netherlands: International Association for the Evaluation of Educational Achievement.
- Warren, H. K., Domitrovich, C. E., & Greenberg, M. T. (2009). Implementation quality in school-based research: Roles for the prevention researcher. In L. M. Dinella (Ed.), *Conducting science-based psychology research in schools* (pp. 129–151). Washington, DC: American Psychological Association.
- Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, 13, 21–39. doi:10.1016/0883-0355(89)90014-1
- Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, 22(5), 366–389. doi:10.2307/749186
- Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal*, 35(4), 607–651. doi:10.2307/1163461
- Wei, L., & Murphy, P. K. (2017). Teacher and student roles: Walking the gradually changing line of responsibility. In P. K. Murphy (Ed.), *Classroom discussions in education* (pp. 30–53). New York: Routledge.
- Wei, L., Murphy, P. K., & Firetto, C. M. (in press). How can teachers facilitate productive small-group talk? An integrated taxonomy of teacher discourse moves. *The Elementary School Journal*.
- Weinstock, M. P., Neuman, Y., & Glassner, A. (2006). Identification of informal reasoning fallacies as a function of epistemological level, grade level, and cognitive ability. *Journal of Educational Psychology*, 89(2), 327–341. doi:10.1037/0022-0663.89.2.327
- Wells, G. (2007). Semiotic mediation, dialogue, and the construction of knowledge. *Human Development*, 50, 244–274. doi:10.1159/000106414

Murphy et al.

- Wilkinson, I. A. G., Soter, A. O., & Murphy, P. K. (2010). Developing a model of Quality Talk about literary text. In M. G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 142–169). New York: Guilford Press.
- Wilkinson, I. A. G., Soter, A. O., Murphy, P. K., & Li, J. (2008, March). *Promoting high-level comprehension of text through quality talk: A quasi-experimental study*. Paper session presented at the annual meeting of the American Educational Research Association, New York, NY.
- Wu, X., Anderson, R. C., Nguyen-Jahiel, K., & Miller, B. (2013). Enhancing motivation and engagement through collaborative discussion. *Journal of Educational Psychology, 105*(3), 622–632. doi:10.1037/a0032792

Manuscript received February 18, 2016

Final revision received March 10, 2018

Accepted March 19, 2018