Instructional Science (2026) 54:57
https://doi.org/10.1007/s11251-026-09809-z

ORIGINAL RESEARCH

Effects of a literary intervention on teachers' monologic and dialogic practices and the quality of students' contributions

Yvonne Dammert[1] · Luciano Gasser[1] · P. Karen Murphy[2]

Received: 26 September 2024 / Accepted: 28 May 2026

Abstract
This quasi-experimental study examines the effects of a literary intervention based on the Quality Talk (QT) approach aimed at promoting dialogic discussions about a moral dilemma. Nine teachers and 104 fourth- and fifth-grade students from Switzerland participated in the study. Over the course of one school year, students in the intervention group read and discussed four literary fiction books in small groups. Teachers received video-based coaching to facilitate the discussions. The subsequent analyses focused on changes in (a) monologic teacher practices, (b) dialogic teacher practices, and (c) the quality of student contributions. Multilevel analyses show that the intervention led to a shift from monologic to dialogic teaching patterns. More specifically, teachers in the intervention group were less likely to ask test questions, control turn-taking, or evaluate students' responses, but more likely to engage in dialogic practices, as indicated by an increase in the use of talk moves. Due to these shifts in teacher practices, students in the intervention group exhibited higher-quality contributions to literary discussions, as demonstrated by increases in authentic questions, cumulative and exploratory talk, reasoning, and elaborated explanations. Ultimately, these findings provide evidence that QT facilitates a shift toward dialogic patterns, thereby enabling more student-centered and cognitively demanding literary discussions in primary education.

Keywords Quality Talk · Dialogic teaching · Monologic teacher practices · Primary education · Literary fiction · Small-group discussion

Introduction

The way teachers and students engage with a text significantly influences how and what students learn during discussions (Goodwin et al., 2021; Murphy et al., 2009). For instance, dialogic discussions about literary fiction—during which students express their thoughts,

✉ Yvonne Dammert
yvonne.dammert@phbern.ch

[1] Bern University of Teacher Education, Fabrikstrasse 8, Bern 3012, Switzerland

[2] The Pennsylvania State University, 102 CEDAR Building University Park, Pennsylvania, PA 16802, USA

Published online: 20 June 2026

critically engage with the text, and share responsibility with their teacher—have been shown to improve students' reading comprehension and higher-order thinking (Correnti et al., 2021; Murphy et al., 2009, 2022). Despite their clear potential, many classroom discussions continue to follow a monologic pattern, which limits students' active engagement and learning opportunities (Alexander, 2008; Mehan & Cazden, 2015).

Differences in discussion patterns may be largely attributed to how teachers negotiate responsibility for steering the direction and content of the discussion (Mortimer & Scott, 2003). Even when surface-level indicators of dialogic discussions (e.g., open-ended questions) are present, teachers tend to exert a high degree of control over discussions through practices such as turn allocation or evaluative feedback (Scott et al., 2006; Sedova et al., 2020), which inhibit the emergence of genuine dialogic interactions. Shifting toward more dialogic patterns requires a fundamental transformation of teacher and student roles. Teachers must effectively scaffold students' contributions while progressively fading control over the discussion to students (Reznitskaya & Gregory, 2013), allowing students to adopt more agentic roles through initiating questions, managing participation, and co-constructing meaning (Clarke et al., 2015; Kim, 2022).

Quality Talk (QT), a text-based discussion approach, addresses these challenges by promoting a dialogic classroom culture and encouraging students to talk about, around, and with the text (Murphy & Firetto, 2018). QT combines explicit instruction in dialogic strategies with sustained teacher scaffolding and student-led discussion formats to foster shared responsibility and high-level engagement with texts. Several studies—which have primarily employed single-group designs and were conducted in the United States—have demonstrated that QT enhances students' response quality, critical-analytic thinking, and high-level text comprehension (Davies et al., 2017; Li et al., 2016; Murphy et al., 2018, 2022).

However, to date, few studies have examined how both dialogic and monologic teacher practices change in response to intervention efforts. Since monologic practices often persist alongside emerging dialogic practices (Sedova et al., 2020; Zook-Howell et al., 2020), a closer examination of their simultaneous development offers deeper insight into how teachers develop the capacity to share responsibility for the content and direction of classroom discussions.

To address this gap, we adapted the QT approach to fourth- and fifth-grade Swiss classrooms and tested the effects of a one-year literary intervention on text-based small-group discussions. In contrast to prior research, this study considers not only indicators of dialogic teacher practices (e.g., teachers' talk moves) but also markers of monologic teacher practices, including closed-ended questions (i.e., test questions), formal evaluations of students' responses, and teachers' control over students' turns as speakers. Using a quasi-experimental design, we examined how the intervention influenced dialogic and monologic discussion patterns and how the promoted practices, in turn, affected the quality of students' contributions (e.g., frequency of open-ended questions, reasoning, exploratory talk).

Extending previous QT research, this study accounts for the hierarchical structure of the data by applying multilevel modeling to estimate intervention effects while accountingfor between-teacher variance. Moreover, we apply the QT approach in a new context through our focus on a Swiss sample of students and teachers holding discussions specifically about literary fiction. This represents a valuable contribution, as most prior studies were conducted in the U.S. and did not explicitly focus on literary texts.

Discussions about literary fiction

Literary fiction affords unique opportunities for students to engage with ambiguity, emotional complexity, and divergent perspectives. Unlike expository texts, it often presents unresolved dilemmas and conflicting viewpoints, prompting readers to weigh and prioritize different perspectives (Bakhtin, 1984; Mar & Oatley, 2008). Such narrative elements can activate socio-cognitive processes within students, such as perspective-taking, reasoning, and critical thinking (Kidd & Castano, 2013; Lenhart et al., 2023).

When reading individually, however, students may focus on surface-level meaning, overlooking subtler social and moral implications. Deeper involvement with the text often arises when students encounter and negotiate alternative interpretations in dialogue with peers. Literary fiction offers particularly fertile ground for such exchanges, as it invites readers to explain their viewpoints, consider others' perspectives, and negotiate conflicting understandings of the moral questions raised by the text (Gasser et al., 2022; Lin et al., 2022). The benefits of such discussions, however, are critically dependent on their design, as not all formats equally promote students' reasoning, depth of understanding, or quality of insight (Schrijvers et al., 2019).

According to the transactional reader theory, readers can adopt different stances toward a literary text, which shape how and what they learn from it (Chinn et al., 2001; Murphy et al., 2009; Rosenblatt, 1978). From an *efferent* stance, the primary goal is to extract information from the text, while an *aesthetic* stance emphasizes emotional and imaginative engagement during reading. A *critical-analytic* stance goes beyond these two stances, requiring readers to reflect critically on both the information within the text and their own emotional responses to said information. This literary stance is inherently interpersonal, as readers benefit from co-readers who help them reflect on their personal biases and collaboratively work toward well-balanced interpretations.

A meta-analysis by Murphy et al. (2009) demonstrated that discussion approaches that encourage a critical-analytic stance (rather than an efferent or aesthetic stance) more effectively promote higher-level text comprehension and critical thinking among students. Text-based discussion approaches that hold a critical-analytic stance are characterized by the use of authentic questions, reasoning, and shared responsibility for the flow and content of discussions. Such engagement reflects the dialogic principles underpinning the broader pedagogical approach of *dialogic teaching* (Alexander, 2008), which aims to foster teaching and learning that is collective, reciprocal, cumulative, supportive, and purposeful.

Monologic teacher practices

Classroom discussions of literary texts often follow a monologic discussion pattern (e.g., Fässler et al., 2019; Nystrand, 2006). By monologic, Bakhtin refers to interactional patterns in which a single authoritative voice dominates discourse, legitimizes specific forms of knowledge, and systematically restricts the space available for alternative perspectives (Bakhtin, 1984). Such patterns are characterized by a high degree of teacher control over both the content and the form of the discussion (Scott et al., 2006). A common manifestation is the IRE—initiation, response, evaluation—pattern, whereby the teacher poses a closed-ended question, the student responds briefly, and the teacher evaluates the answer (Mehan, 1979). While these patterns may efficiently transmit knowledge, they limit stu-

dents' opportunities for active engagement, reasoning, and critical thinking, often resulting in more superficial levels of understanding (Böheim et al., 2021; Reusser & Pauli, 2015).

Observable indicators of teacher control include a high proportion of teacher talk as well as closed questions, or so-called *test questions*, which elicit predetermined responses and typically prompt students to recall and reproduce information from the text (e.g., "What does the main character do at the beginning of the story?") (Nystrand et al., 2003). Although professional development in dialogic teaching has been shown to reduce the prominence of such questions (Murphy et al., 2018; Reznitskaya & Gregory, 2013), recent case studies (Sedova et al., 2020; Zook-Howell et al., 2020) have indicated that such changes are not comprehensive. Teachers may incorporate dialogic elements (e.g., open-ended questions) while still maintaining control by, for example, limiting students' opportunities to speak, managing turn-taking, and frequently evaluating their responses. This hybrid practice, known as *overfacilitation* (Zook-Howell et al., 2020), combines surface-level features of dialogue with an underlying monologic logic—ultimately leaving control over the discussion with the teacher.

To better capture how teachers' ability to share interpretive authority may be developed through the QT approach, this study considers indicators of both dialogic and monologic teacher practices. By capturing the co-occurance of dialogic and monologic discussion features, the study offers a more nuanced understanding of how control over textual interpretation persists or evolves throughout dialogic interventions.

Dialogic teacher practices

In contrast to monologic teacher practices, dialogic discussions foster an environment in which students view themselves as equal contributors to a collaborative learning community (Reznitskaya & Gregory, 2013). However, simply limiting instructional control or increasing student talk is insufficient to facilitate truly meaningful participation. Instead, scaffolding, defined as targeted, adaptive, and temporary support, is crucial for fostering active engagement and enabling students to perform beyond their current level of independent competence (Vygotsky, 1978). Scaffolding serves as transitional support that enables students to gradually adopt dialogic practices. To foster student agency in text-based discussions, teacher support must not only be responsive to learners' reasoning and argumentation but also be gradually withdrawn as their competence and autonomy develop (Nickl et al., 2024; van de Pol et al., 2010).

A central scaffolding strategy in dialogic teaching is the use of *authentic questions*, defined as open-ended and cognitively demanding questions. Unlike test questions, authentic questions (e.g., "Why does the main character act like this?") invite students to inquire, justify, and engage critically with the text. This, in turn, enhances the cognitive level of discussions and invites the expression of diverse voices and perspectives. Numerous studies have highlighted the positive effects of authentic questions on students' motivation and response quality (e.g., Jurik et al., 2014; Rapanta & Macagno, 2023).

Scaffolding also entails the strategic use of talk moves that support students' argumentation and reasoning skills (O'Connor & Michaels, 2019; Wei et al., 2018). For example, when a teacher uses prompting, such as asking for evidence or justification for a claim, this increases the likelihood that the student will provide a more elaborate response and may also encourage peers to engage in similar practices (Lin et al., 2015). Empirical research

has demonstrated that talk moves such as summarizing (e.g., "Can you briefly summarize what has been said?"), modeling (e.g., "I would like to ask you an uptake question: What do you think her father could do instead?"), marking (e.g., "I like how Lucas supported his claim with a reason."), prompting (e.g., "How do you know that?"), and challenging (e.g., "What counterarguments could you make to Neele's claim?") are particularly effective in encouraging critical-analytic thinking and high-level text comprehension (Correnti et al., 2021; Wei & Murphy, 2018).

Quality of student contributions

Monologic and dialogic teacher practices shape the extent to which students contribute actively and meaningfully to classroom discussions. High-quality contributions are characterized by students taking ownership of the discussion and engaging in reasoning and the co-construction of knowledge with their peers (Ruzek et al., 2016).

Ownership manifests in two ways. First, students demonstrate control over the form of the discussion when they speak without teacher prompting and allocate speaker turns themselves (e.g., "Ian, do you agree with what I said?"). Second, they demonstrate control over the content of the discussion by independently posing their own questions and introducing new topics (Kim, 2022).

The quality of students' contributions is further reflected in their reasoning practices, which reflectstudents'ability to clearly express their positions and support their claims with reasons and evidence (Iordanou et al., 2019). Research has shown that verbalizing and explaining one's own thinking can reveal gaps in understanding and, in turn, promote learning (Asterhan & Schwarz, 2016). Critically, reasoning is not a strictly individual activity; it is also an interpersonal one. Collaborative reasoning enables the co-construction of knowledge (Kuhn, 2018), as it requires students to engage actively with diverse perspectives, either by building on others' ideas (*cumulative talk*) or critically examining and comparing alternative positions over several turns (*exploratory talk*). Regular engagement in such forms of discussion enhances comprehension and supports the development of students' reasoning skills (Asterhan & Schwarz, 2016).

Quality talk: facilitating dialogic discussions about literary texts

This study is centered on the QT approach, which operationalizes dialogic pedagogy in a structured framework for discussion in language arts that fosters students' critical-analytic thinking and high-level text comprehension (Murphy et al., 2009; Murphy & Firetto, 2018; Soter et al., 2008; Wilkinson et al., 2010). Critical readers go beyond talking *about* texts or identifying information; rather, they engage with texts by talking *around* and *with* them, analyzing, for example, their relevance to personal experiences and the broader world. Through collaborative discussions with peers, students refine their initial responses and evaluate the validity of one another's arguments.

QT emphasizes three core dialogic elements: authentic questions, reasoning, and co-constructed responses (i.e., cumulative and exploratory talk). Authentic questions prompt students to generalize, analyze, speculate, and create affective or intertextual connections to the text (Murphy & Firetto, 2018; Nystrand et al., 2003). To scaffold students' active

participation and reasoning, the model integrates a set of teacher talk moves—summarizing, modeling, marking, prompting, and challenging (see Sect. 1.1.2).

Following an initial workshop that introduces the core principles of QT, teachers implement these strategies in small-group discussions and engage in ongoing reflection on and analysis of their instructional practices during coaching sessions. QT is explicitly designed for small-group discussions of approximately four to six students (Murphy et al., 2018), a format that maximizes opportunities for each student to contribute and allows for more targeted scaffolding by the teacher. Whole-class mini-lessons complement this structure: mini-lessons are conducted with all students to introduce and model key dialogic skills, such as asking authentic questions and providing elaborated responses. These skills are subsequently applied and refined in small-group literary discussions, during which control gradually shifts from teachers to students, who increasingly manage turn-taking and assume interpretive authority over the text.

Supporting structures reinforce these practices: discussion rules help establish normative expectations for a dialogic classroom culture (e.g., "We let each other finish speaking"). Additionally, students use prompts in their literacy journals to generate authentic questions and elaborated responses, which they work on during the reading process and bring to discussions.

The results of prior empirical studies on this topic, which have primarily been conducted in English-speaking classrooms, suggest that QT enhances discussion quality by reducing test questions and teacher talk while promoting authentic questions and talk moves, including prompting, challenging, and modeling (Li et al., 2016; Murphy et al., 2018). Notably, all studies reported increases in students' use of authentic questions as well as the quality of their responses, particularly in terms of elaborated explanations and cumulative or exploratory talk (Davies & Esling, 2024; Li et al., 2016; Murphy et al., 2018).

Despite these promising findings, key aspects of monologic teacher practices remain underexamined. Thus, analyzing both monologic and dialogic teacher practices enables us to achieve a more comprehensive understanding of how teachers gradually grant control over discussions to students. Moreover, most prior studies have relied on single-group pretest-posttest designs, limiting causal inference.

Literary education in Switzerland

In German-speaking primary and lower secondary education, literature instruction continues to be strongly shaped by efferent reading stances, most notably in the form of teacher-directed text discussions aimed at developing basic language skills (e.g., vocabulary, narrative comprehension). It is also influenced by aesthetic approaches, including production-oriented activities like dramatization, drawing, and creative writing, all of which are geared toward fostering imagination and emotional engagement with the text. Empirical studies from Germany and Switzerland consistently show that typical literary discussions are largely teacher-led, dominated by closed questions, and focus primarily on comprehension checks and motivational aims, while open-ended, interpretive engagement with texts remains rare (e.g., Böhme et al., 2018; Fässler et al., 2019).

This pattern aligns with international survey data (Schulz et al., 2010), indicating that Swiss students have below-average opportunities to participate in open classroom discussions. In Swiss schools, teachers often associate open discussions with democratic educa-

tion formats, such as class councils, rather than with subject-specific instruction. However, recent curricular developments (D-EDK, 2016) and international research (e.g., Gasser et al., 2022; Nucci & Ilten-Gee, 2021) have highlighted the importance of fostering transversal competencies like reasoning and dialogic skills within subject-specific instruction, including literary education. Despite this, systematic approaches to the integration of student-centered literary discussions into daily practice remain scarce, presenting a serious gap in the literature as well as in schools. Thus, Quality Talk represents a timely and promising opportunity to challenge established routines in German-language literary education and promote dialogic classroom discussions.

The present study and hypotheses

This study addresses this gap by adapting the text-based QT approach to literary education for fourth and fifth graders in Switzerland. The adapted approach exclusively focuses on literary fiction. Moreover, the mini-lessons included explicit instruction on authentic questions that specifically address the literary context. For example, children learned to ask Character Questions to explore characters' inner lives and social relationships (e.g., " How does Tommy feel when everybody wants to know what he is hiding?"). They were also introduced to Big Questions to encourage analysis of the story's main problem, attempts at problem-solving, and the moral message of the story (e.g., "Is it OK or not OK how Bruno reacts to the bullying? Why?").

The Swiss context is of particular interest, as results from international large-scale assessments such as PISA place Swiss students in the OECD mid-range for reading literacy, significantly below countries like the United States and New Zealand (OECD, 2023), where previous QT studies have been successfully implemented. Such performance levels may necessitate greater teacher support to ensure basic text comprehension, while also providing considerable scope for gains in both discussion skills and reading comprehension.

Using a control-group design, we investigated whether QT influences (a) monologic teacher practices, (b) dialogic teacher practices, and (c) the quality of student contributions. To effectively capture persistent monologic structures, we refined Murphy et al.'s (2017) coding scheme by introducing additional codes (e.g., evaluation, turn allocation), thereby enhancing the analytical depth of our study. To account for variability, we used multilevel modeling to estimate intervention effects while controlling for between-teacher differences, grade level, and the proportion of students who did not speak German at home, which may influence students' discursive competence and participation patterns. This allows us to identify which elements of QT are robustly adopted across teachers as well as where implementation challenges remain.

Based on previous QT research, we have formulated three hypotheses:

H1 Controlling for pretest levels, teachers in the intervention group (IG) will exhibit lower levels of *monologic teacher practices* (as measured by the proportion of teacher turns, test questions, evaluative feedback, and teacher-directed turn allocation) at posttest relative to those in the control group (CG).

H2 Controlling for pretest levels, teachers in the IG will exhibit more substantial use of *dialogic teacher practices* (as measured by the frequency of authentic questions and talk moves) at posttest relative to those in the CG.

H3 Controlling for pretest levels, students in the IG will demonstrate higher-quality *contributions* (as measured by the frequency of authentic student questions, student-directed turn allocation, reasoning, elaborated explanations, and cumulative and exploratory talk) at posttest relative to those in the CG.

Methods

Design

This one-year study was conducted during the 2020–2021 school year with eight primary school classes across five schools in the German-speaking part of Switzerland. Recruitment was done via a digital flyer sent to all primary schools in the German-speaking canton of Lucerne, Switzerland. The flyer outlined the program's aims and detailed the associated intervention's components and time requirements. To evaluate the efficacy of our intervention, we adopted a quasi-experimental design with cluster-level allocation. Following the registration process, classes were randomly assigned to the IG or CG. All classes from the same school were assigned to the same group to enhance the study's internal validity. To improve group comparability, we also ensured that fourth- and fifth-grade classes were equally represented across the IG and CG. Initially, ten classes had signed up, but two from the CG dropped out prior to the start of the intervention, leading to a skewed distribution between the IG (five classes) and CG (three classes). The intervention program was introduced to the IG as part of their language arts curriculum, whereas the CG carried on with business-as-usual instruction.

Sample

Nine teachers participated in the study, six of whom were assigned to the IG and three were assigned to the CG. Although the IG included only five classes, six teachers participated because, in one class, both the classroom teacher and a special education teacher contributed to the intervention. The special education teacher was periodically involved in instruction as part of a co-teaching arrangement and, thus, took responsibility for supporting one of the small groups throughout the intervention. Eight of the nine teachers were female. Their average age was 35.89 years ($SD = 11.66$, range = 24–58), and their average level of professional experience amounted to 13.44 years ($SD = 11.39$, range = 3–35). Their average workload was 70.44% ($SD = 18.95$, range = 40–100), with a 100% workload corresponding to 29 lessons per week.

The final sample consisted of 104 students: 72 in the IG and 32 in the CG. Of the 141 students enrolled in the participating classes, only those with written parental consent for video recording were included in the sample. As a result, 37 students participated in the small-group discussions during the intervention but were not included in the pretest and posttest discussions that formed the basis of the analyses.

The average class size was 17.63 students ($SD = 3.61$, range = 13–22). At the beginning of the intervention, students' average age was 9.60 years ($SD = 0.68$, range = 9–12). The sample was balanced in terms of gender (51.9% female, 46.2% male, and 2% undisclosed). In terms of linguistic background, 66.3% spoke only German at home, 29.8% spoke German alongside another language, and 3.8% spoke only a non-German language.

At the study's outset, teachers formed small groups of three to seven students, which were composed heterogeneously in terms of gender and language skills. The groups in the IG had an average size of 5.54 students ($SD = 0.88$, range = 4–7), while those in the CG had an average size of 4.57 students ($SD = 1.13$, range = 3–6). The study comprised a total of 20 small groups ($n_{IG} = 13$, $n_{CG} = 7$), none of which changed at any point during the intervention or assessment. IG teachers managed an average of 2.67 groups each ($SD = 1.51$, range = 1–4), while CG teachers managed an average of 2.33 ($SD = 1.73$, range = 1–4). Mann–Whitney U tests for independent samples revealed no significant differences between the small groups in the IG and those in the CG in terms of students' age ($U = 125.50$, $z = -0.88$, $p = .381$), gender ($U = 131.50$, $z = -0.40$, $p = .688$), or home language use ($U = 149.00$, $z = 1.00$, $p = .320$).

Procedure

Figure 1 presents an overview of the data-collection process and the intervention design. Further details on the intervention and teacher professional development (TPD) are provided in Table 7 and 8 in Appendix A.

Assessment

Video data were collected before and after the intervention (pretest and posttest; see Fig. 1). On both ends of the intervention, students engaged in discussions with their teacher in small groups (average duration of 12.34 min; $SD = 2.97$). Each discussion focused on the same short story featuring a moral dilemma, *Revenge Is Sticky!* (894 words), which was used in

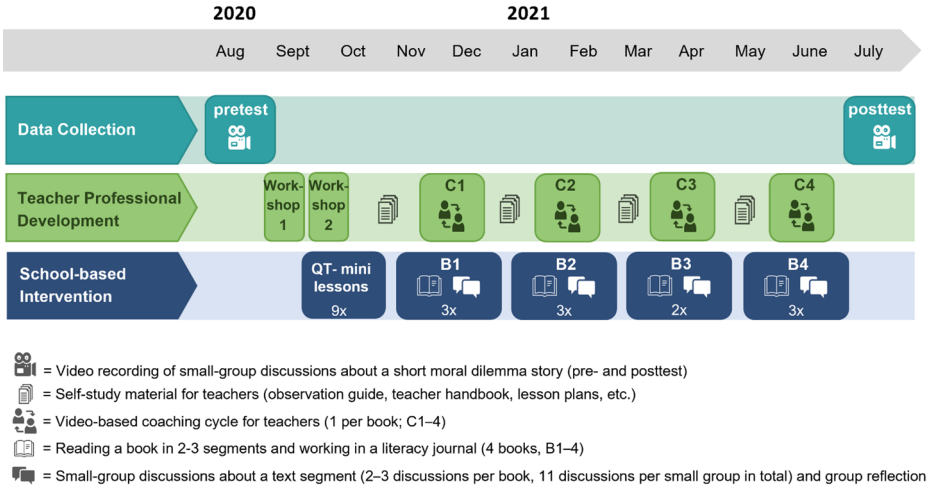

Fig. 1 Overview of Data Collection and Intervention Design

both the pretest and posttest to ensure comparability over time. The short story describes the dilemma of a girl who, upon arriving in a new class, must decide whether to tell the truth to the teacher about the harmful behavior of her new friends or lie to the teacher and maintain her friendships. This conflict between group loyalty and morality was also central to the four children's books used during the intervention. The final version of the short story is provided in Appendix B.

The difficulty level of the story and the intervention books was measured by the German readability index LIX (Lenhard & Lenhard, 2014), a quantitative indicator based on mean sentence length (number of words divided by number of sentences) and the proportion of long words (i.e., words with more than six letters). The short story's readability index score (29.97) is indicative of a low difficulty level (thresholds: < 40 = easy, 40–60 = medium, > 60 = complex academic texts; cf. Lenhard & Lenhard, 2014). The story was purpose-written for the study, piloted in a mixed-grade primary classroom (grades 4/5), and revised based on teacher and student feedback.

Shortly before the start of data collection, the teachers received the story and were instructed to guide text-based discussions as they usually do. Teachers read the story aloud to their classes, and students were invited to ask questions about difficult words. The discussions took place in small groups in a separate room and were all led by the teacher.

Intervention

Through this study's intervention based on the QT model, the students were instructed on how to identify various authentic questions, justify their claims, and collaboratively construct knowledge in discussions through nine mini-lessons (see "QT mini-lessons" in Table 7 in Appendix A for more detailed descriptions). They then read and discussed four high-quality children's books: *The Very Persistent Gappers of Frip* (Saunders & Smith, 2000), *Brown: My Alter Ego Is a Superhero* (Øvreås et al., 2019), *Rabbit & Bear: Attack of the Snack* (Gough & Field, 2018), and *Balaclava Boy* (Robson, 2014). All of these books (or the authors behind them) had received critical literary awards and met desired literary criteria, such as character multidimensionality, textual ambiguities, and outstanding language. The difficulty level of the books was age-appropriate, with an average LIX score of $M = 28.35$ ($SD = 2.83$, range = 24.65–31.30), indicating low complexity (cf. Lenhard & Lenhard, 2014).

The students also worked in literacy journals to activate their prior knowledge about story themes, establish basic comprehension of the texts, and prepare authentic questions for the discussion. Each book was read and discussed in two to three segments (see "Book lessons" in Appendix A, Table 7 for more detailed descriptions). Throughout the school year, students participated in 11 literary discussions and subsequently reflected on specific self-established discussion goals.

Professional development of teachers

The TPD program featured two professional development workshops and three to four individual coaching sessions (see Table 8 in Appendix A for a more detailed description). The first workshop introduced teachers to the conceptual foundations of QT, while the second workshop focused on the selected children's books, enabling teachers to become experts on their content, main themes, and literary characteristics.

During implementation, teachers participated in up to four video-based coaching sessions (C1–4), which were organized as a coaching cycle and grounded in a strengths-based approach (Allen et al., 2011). Throughout this iterative process, teachers analyzed their own classroom interactions, practiced talk moves, and reflected on their effectiveness. The overarching goal of these efforts was to transform teachers' discussion routines by bolstering their observational and reflective skills.

Coaching was conducted by three experienced teachers, one of whom is also the first author of this study. All three had participated in a three-day QT workshop that introduced the conceptual foundations of the approach and provided training in discussion coding. All employed materials were piloted by the coaches in their own classrooms prior to the intervention.

Measures

Demographic questionnaire

Before the pretest, teachers provided demographic information about themselves (age, gender, teaching experience, employment percentage, professional role, and grade level taught) and their students (age, gender, and primary home language).

Video analysis

Video data were transcribed and imported into MAXQDA Analytics Pro (version 22.3) for systematic coding. The coding process followed the coding manual from Murphy et al. (2017) but included additional codes to capture monologic teacher practices. For a detailed operationalization of all codes, see Appendix C, Table 9.

The coding process unfolded across four iterations: (1) identifying teacher and student utterances at the turn level; (2) delimiting question events (each comprising an initial question and its corresponding responses) at the cross-turn level, including the further classification of question types (authentic or test questions); (3) coding student responses at the turn level (reasoning, elaborated explanations) as well as the cross-turn level (cumulative talk, exploratory talk); and (4) analyzing additional utterances not categorized as a question or response (e.g., turn allocation, evaluation of student responses, teacher talk moves).

Two project members double-coded all small-group discussions recorded at pretest and posttest ($N=40$). Across all codes, interrater agreement resulted in a robust kappa ($\kappa=0.83$). Similarly, analyses of individual discussion elements resulted in good to very good intercoder reliability ($\kappa=0.63$–0.84; see Appendix D). Subsequently, the coding of all discussion elements was aggregated to the small-group level in line with the coding manual from Murphy et al. (2017). Considering the variance in discussion duration, which made it difficult to compare the frequencies of discussion elements in subsequent analyses, we adjusted the number of codes for each discussion element according to the duration of the discussion, transforming them into ten-minute interval variables (Hennessy et al., 2023). This approach standardized the data to facilitate more consistent comparisons. Teacher discussion practices and the quality of student contributions were coded as follows:

(a) *Monologic teacher practices*: To determine the degree of teacher control during text-based discussions, four indicators were analyzed, each reflecting different aspects of the regulation of participation and content. The proportion of *teacher turns* was measured by dividing the total number of teacher turns by the sum of all turns taken (students and the teacher) in the discussion, resulting in a measure that reflects the overall dominance of teacher talk. *Teacher turn allocation (TTA)* included all utterances in which the teacher designated the next speaker or directed how the discussion should continue. *Test questions (TQ)* were defined as closed questions with a predetermined answer. Finally, *teacher evaluation* (*EA*) included all utterances in which the teacher assessed the correctness or adequacy of a student's response.

(b) *Dialogic teacher practices*: We used two indicators to assess effective teacher scaffolding in dialogic discussions. *Authentic teacher questions (AQ teacher)* were defined as open-ended questions posed by the teacher without a predetermined answer. *Talk moves (TM)* captured a range of scaffolding practices, including instances of prompting, summarizing, or challenging students' contributions.

(c) *Quality of student contributions*: We used six indicators to assess the quality of student contributions. *Authentic student questions (AQ student)* referred to students' open-ended questions, and s*tudent turn allocation (STA)* referred to utterances in which students actively assigned or invited peers to contribute. We analyzed indicators of reasoning and argumentative engagement at both the individual and group levels (Murphy et al., 2017). *Reasoning* (*REA*) and *elaborated explanations* (*EE*) were coded at the individual (i.e., turn) level; the former counted student references to justifications, if-then rules, or analogies, while the latter counted instances of student responses including two reasoning elements. *Cumulative talk* (CT) and *exploratory talk* (ET) were coded at the group level; CT was identified when students built on and extended one another's ideas across at least three consecutive turns, while ET was coded when students challenged or rebutted others' contributions over a sequence of at least three turns. A detailed description of these codes is available in Appendix C.

Intervention fidelity

The teachers consistently documented the intensityof the intervention and their adherence to the manuals. They implemented the intervention with high fidelity. They spent more time on the intervention ($M = 34.07$ h) than the prescribed 32.92 h and reported minimal deviations from the manual ($M = 0.49$, $SD = 0.26$; 0 = *no deviations*, 3 = *many deviations*).

Data analysis

Analytical framework and data structure

The dataset exhibited a hierarchical structure, with 20 small groups (Level 1) nested within 9 teachers (Level 2). To account for dependencies within teachers, multilevel modeling was applied. With only nine Level-2 units, conventional estimation methods such as (restricted) maximum likelihood (REML) risk producing biased variance estimates and inflated Type I error rates (Hox et al., 2020; McNeish & Stapleton, 2016). Therefore, Bayesian multilevel modeling was employed. Recent methodological research has shown that Bayesian estima-

tion yields more stable and less biased parameter estimates under small-sample conditions, while also improving convergence and enhancing robustness via the specification of prior distributions (Hox et al., 2020; Smid et al., 2020; van de Schoot et al., 2021). All models were estimated in R using the *brms* package (R Core Team, 2023).

Model specification

Model development followed a stepwise procedure based on theoretical considerations and empirical model fit (Van Breukelen, 2023). Distributional choices were guided by descriptive inspection (histograms, Q–Q plots) and the comparison of competing distributional families using approximate leave-one-out cross-validation (PSIS-LOO).

We distinguished between different types of outcomes as follows:

(a) *Proportional outcomes* (e.g., proportion of teacher turns) were modeled using beta regression, which is appropriate for continuous variables bounded between 0 and 1.
(b) *Positively skewed continuous outcomes* were modeled using gamma regression when values were strictly positive (see Tables 2, 3 and 4).
(c) *Zero-inflated continuous outcomes* were modeled with hurdle gamma models (see Tables 2, 3, and 4), which decompose the outcome into (1) a binary *hurdle* component estimating the probability that the value is greater than zero, and (2) a *gamma* component estimating the distribution of the positive values. This approach separates the occurrence of an outcome from its magnitude, which is essential for zero-inflated continuous data.

All models included Condition (1 = intervention group) as the main predictor and controlled for the pretest value of the outcome variable. Two additional group-level covariates were included: the grand-mean centered proportion of students who did not speak German at home and grade level (dummy-coded: 1 = Grade 5). In hurdle gamma models, Condition was included in both the hurdle and gamma components, reflecting the assumption that the intervention would influence both the probability of occurrence and the magnitude of the outcome (Bürkner, 2017). All other covariates were restricted to the gamma component as a means of limiting model complexity. Random intercepts for teachers were modeled to account for Level-2 variance.

Prior specification

Priors were specified in line with established recommendations for Bayesian modeling under small-sample conditions (Zitzmann et al., 2021). For the intervention effect, we tested three levels of prior informativeness across all outcome models: *weakly informative* (e.g., Normal(0, 1)), *moderately informative* (e.g., Normal(–0.5, 0.5)), and *strongly informative* (e.g., Normal(–0.6, 0.2)). Prior means were derived from theoretical expectations regarding dialogic teaching, which consistently predicts reductions in monologic practices alongside increases in dialogic practices and the quality of student contributions (e.g., Alexander, 2008). Prior standard deviations were calibrated to reflect plausible ranges of intervention effects reported in previous empirical studies. Such studies typically report moderate to large effects on teacher practices and small to moderate effects on student outcomes (e.g.,

Chen et al., 2020; Matsumura et al., 2019; Wilkinson et al., 2023), given the scaling of the outcome variables and the respective model specifications. Prior plausibility was evaluated via prior predictive checks, and robustness was examined through sensitivity analyses comparing key model estimates (e.g., posterior means, 95% credible intervals, Bayesian R^2, ΔELPD) across prior conditions (available in Online Supplementary Material 1). Robustness was moderate to high, with deviations in posterior means ranging from 0.1 to 0.9, and the direction of the estimated intervention effects remained consistent within each outcome model across prior specifications. Overall, the specification of priors followed a combination of theoretical, empirical, and statistical considerations.

Estimation and model evaluation

All models were estimated using four Markov chains with 4000 iterations per chain, including 1000 warm-up iterations, resulting in 12,000 post-warm-up draws per model. No additional thinning was applied (thin = 1).

Convergence was assessed using the potential scale reduction factor ($\hat{R}$) and effective sample size diagnostics (Bulk ESS, Tail ESS). Across all models, $\hat{R}$ values were equal to 1.00, indicating full convergence. Effective sample sizes were consistently large (generally exceeding 2000), indicating stable and efficient posterior sampling (see Online Supplementary Material 2).

In addition to numerical diagnostics, convergence and sampling performance were evaluated visually using trace plots and autocorrelation plots. Trace plots were inspected for adequate mixing and stationarity across chains, while autocorrelation plots were used to assess the decay of serial dependence across lags. These diagnostics indicated satisfactory convergence and efficient sampling, with no evidence of non-convergence, poor mixing, or problematic autocorrelation structures. Exemplary diagnostic plots for each distributional family are presented in Online Supplementary Material 3.

Model fit was assessed using posterior predictive checks. Intervention effects are reported as posterior means with 95% credible intervals (CIs). CIs excluding zero were interpreted as indicating that the posterior distribution provided consistent evidence for a non-zero effect. Effect sizes are expressed as odds ratios (ORs) for beta and hurdle components and as rate ratios (RRs) for gamma components. ORs and RRs are reported as percentage changes in odds or expected values of the outcome. Evidence ratios (ERs) are reported as a Bayesian measure of the strength of evidence for a directional hypothesis, with larger values indicating stronger evidence (3 = moderate; 10 = strong; 30 = very strong; $\geq$ 100 = extreme).

Results

Descriptive statistics for all outcome variables, including monologic teacher practices, dialogic teacher practices, and student contributions, are presented in Table 1. These values represent observed data rather than model-based estimates.

Table 1 Means (M) and Standard Deviations (SD) of Outcome Variables by Condition and Measurement Point

Variable	Intervention Group ($n=13$)		Control Group ($n=7$)	
	Pretest M (SD)	Posttest M (SD)	Pretest M (SD)	Posttest M (SD)
Proportion of Teacher Turns	0.46 (0.04)	0.18 (0.13)	0.46 (0.03)	0.40 (0.05)
Teacher Turn Allocation	12.91 (5.21)	1.15 (0.91)	13.60 (4.81)	13.79 (3.63)
Test Question	10.64 (3.98)	0.21 (0.42)	6.22 (3.05)	4.11 (2.93)
Evaluation	25.74 (9.21)	4.25 (4.58)	18.16 (4.81)	16.68 (3.60)
Authentic Teacher Question	5.88 (2.56)	0.57 (0.94)	6.05 (1.38)	6.75 (2.94)
Talk Move	1.02 (0.71)	2.60 (1.61)	0.83 (1.04)	0.59 (0.78)
Authentic Student Question	0.04 (0.14)	4.93 (3.10)	0.00 (0.00)	0.00 (0.00)
Student Turn Allocation	0.10 (0.24)	2.08 (3.46)	0.00 (0.00)	0.22 (0.39)
Reasoning	4.55 (1.71)	8.21 (2.92)	3.90 (1.67)	3.22 (1.82)
Elaborated Explanation	0.69 (0.76)	3.39 (1.64)	0.51 (0.70)	0.82 (1.19)
Cumulative Talk	0.11 (0.29)	1.86 (1.19)	0.14 (0.38)	0.33 (0.58)
Exploratory Talk	0.07 (0.26)	0.79 (0.89)	0.00 (0.00)	0.09 (0.25)

Monologic teacher practices (H1)

For the first hypothesis, we examined how the intervention affected indicators of monologic teacher practices. Table 2 summarizes the posterior estimates from Bayesian multilevel regression models predicting these four indicators of teacher-centered discussion control.

Teacher turns Posterior estimates indicated that, at posttest, the proportion of teacher turns was 46% lower in the IG than in the CG, providing very strong evidence for a reduction (ER = 2999). In addition, higher baseline values, a higher proportion of students who did not speak German at home, and being in Grade 5 were all associated with fewer teacher contributions.

Teacher turn allocation (TTA) If TTA strategies were used (gamma part), their frequency was 73% lower in the intervention group, once again providing very strong evidence for a reduction (ER = 189). The likelihood of completely omitting such strategies (hurdle part) was 67% higher in the intervention group, but evidence for this effect was low (ER = 2)—although, notably, this estimate was highly uncertain.

Test questions (TQ) At posttest, IG teachers who continued to ask test questions during small-group discussions did so with a 63% lower average frequency than CG teachers (gamma part)—strong evidence of a reduction (ER = 196). At posttest, teachers in the IG were also more likely to omit test questions entirely (+ 869%, hurdle part), with very strong evidence for this effect (ER = 176), although the wide credible interval is indicative of high uncertainty regarding the exact size of the effect.

Evaluation (EA) Posterior estimates indicated that, when simple evaluative feedback was provided (gamma part), its average frequency was 34% lower in the IG than in the CG, supported by strong evidence (ER = 50). The likelihood of completely omitting simple evaluative feedback was 2105% higher in the IG (hurdle part). However, the very wide cred-

Table 2 Posterior Estimates from Bayesian Multilevel Regression Models Predicting Monologic Teacher Practices

	Teacher Turns[b] Est. [CI]	Turn Allocation (Teacher)[d] Est. [CI]	Test Question (Teacher)[d] Est. [CI]	Evaluation (Teacher)[d] Est. [CI]
Fixed Effects				
Intercept (Beta)	**1.23 [0.72, 1.75]**	-	-	-
Intercept (Hurdle/Gamma)	-	**–2.77 [–5.11, −0.84] / 1.40 [0.00, 2.72]**	–1.06 [–2.63, 0.35] / **1.90 [0.30, 3.50]**	**–3.68 [–7.52, −0.96] / 2.01 [0.50, 3.48]**
Pretest	**–0.60 [−0.99, −0.20]**	0.04 [–0.02, 0.11]	–0.04 [−0.21, 0.13]	0.02 [−0.04, 0.08]
First Language	**–0.59 [−0.96, −0.21]**	−0.19 [–1.17, 0.82]	0.21 [−1.62, 1.98]	−0.06 [–0.87, 0.76]
Grade	**–0.61 [−0.99, −0.21]**	−0.10 [–1.40, 1.22]	−1.69 [−3.68, 0.82]	−0.24 [–1.38, 0.95]
Condition (Beta)	**–0.62 [−1.01, −0.23]**	-	-	-
Condition (Hurdle/Gamma)	-	0.51 [–1.66, 2.88] / **−1.31 [–2.19, −0.31]**	**2.27 [0.55, 4.07] / –0.99 [–1.69, −0.25]**	**3.09 [0.17, 6.97] / −0.42 [–0.80, −0.03]**
Random Effect				
Teacher: SD (Intercept)	1.75 [1.03, 2.86]	0.79 [0.19, 1.68]	0.57 [0.02, 1.81]	0.81 [0.33, 1.53]
Model Summary				
Bayesian R^2	0.85 [0.68, 0.94]	0.79 [0.49, 0.93]	0.56 [0.17, 0.83]	0.68 [0.41, 0.87]
Odds Ratio	0.54 [0.37, 0.80]	1.67 [0.19, 17.89]	9.69 [1.69, 62.55]	22.05 [1.18, 1064.25]
Rate Ratio	-	0.26 [0.11, 0.74]	0.37 [0.18, 0.79]	0.66 [0.45, 0.97]

[b]Relative frequency, modeled using beta regression; [d]Frequency per 10-minute discussion, modeled using hurdle-gamma regression

Est. = Posterior mean; CI = 95% Credible interval. Bold values indicate that the 95% Bayesian credible interval does not include zero and are interpreted as evidence of an effect

Pretest = Baseline value of the respective outcome at t_0; First language (1 = Other than German, 0 = German); Grade (1 = 5th grade, 0 = 4th grade); Condition (1 = Intervention group, 0 = Control group). Bayes R^2 = Proportion of variance explained by fixed effects

ible interval is indicative of considerable uncertainty regarding the exact size of the effect despite the strong evidence for its direction (ER = 59).

Dialogic teacher practices (H2)

To evaluate the second hypothesis, we examined how the intervention affected indicators of dialogic teacher practices. Table 3 presents the posterior estimates from Bayesian multilevel regression models predicting these discourse-related outcomes.

Authentic teacher questions (AQ teacher) Contrary to the hypothesized increase, posterior estimates indicated a 73% lower frequency of authentic questions in the IG, with no evidence in favor of the hypothesized direction (ER < 1). In addition, the intervention increased the likelihood of completely omitting authentic questions by 1101%, with very strong supporting evidence (ER = 161).

Teacher talk moves (TM) Posterior estimates indicated that, among teachers who used talk moves during small-group discussions (gamma part), the intervention was associated with a 68% higher frequency compared to the CG, once again with strong evidence for this effect (ER = 323). The likelihood of completely omitting talk moves (hurdle component) was 93% lower in the IG, with very strong evidence for this effect (ER > 999).

Table 3 Posterior Estimates from Bayesian Multilevel Regression Models Predicting Dialogic Teacher Practices

	Authentic Question (Teacher)[d] Est. [CI]	Talk Move (Teacher)[d] Est. [CI]
Fixed Effects		
Intercept (Hurdle/Gamma)	**–1.81 [–3.73, −0.17] / 2.61 [0.85, 4.523]**	0.25 [–0.98, 1.47] / 0.33 [–0.61, 1.34]
Pretest	–0.11 [–0.34, 0.13]	0.07 [−0.60, 0.73]
First Language	–0.13 [–2.01, 1.71]	0.16 [–1.18, 1.53]
Grade	–0.55 [–2.47, 1.77]	−0.11 [–1.59, 1.47]
Condition (Hurdle/Gamma)	**2.49 [0.51, 4.68] /** –1.32 [–2.96, 0.68]	**–2.73 [–3.91, −1.53] / 0.52 [0.14, 0.89]**
Random Effect		
Teacher: SD (Intercept)	0.73 [0.03, 2.27]	0.38 [0.02, 1.10]
Model Summary		
Bayesian R^2	0.64 [0.29, 0.88]	0.43 [0.17, 0.69]
Odds Ratio	12.01 [1.67, 108.17]	0.07 [0.02, 0.22]
Rate Ratio	0.27 [0.05, 1.96]	1.68 [1.15, 2.43]

[d]Frequency per 10-minute discussion, modeled using hurdle-gamma regression

Bold values indicate that the 95% Bayesian credible interval does not include zero and are interpreted as evidence of an effect

Pretest = Baseline value of the respective outcome at t_0; First language (1 = Other than German, 0 = German); Grade (1 = 5th grade, 0 = 4th grade); Condition (1 = Intervention group, 0 = Control group). Bayes R^2 = Proportion of variance explained by fixed effects

Changes in quality of student contributions (H3)

To evaluate the third hypothesis, we examined changes in the quality of students' contributions to discussions. Table 4 presents the posterior estimates from the Bayesian multilevel regression models predicting these six outcome variables.

Authentic student questions (AQ student) The intervention group exhibited clear effects in both model components. The frequency of AQs (gamma part) was 102% higher among students in the IG than among those in the CG, with very strong evidence for this effect (ER = 461). Moreover, the probability of not asking any AQs (hurdle part) was far lower (–97%) in the IG, indicating a substantial increase in occurrence (ER > 999).

Student turn allocation (STA) The intervention was associated with a 64% higher frequency of turn allocation strategies among students (gamma part), with strong evidence (ER = 22). In the hurdle component (probability of completely omitting such strategies), the intervention was associated with a 66% lower likelihood of non-use (ER < 1). However, in both cases, the wide credible intervals encompass effects in both directions, indicating substantial uncertainty regarding the intervention effect.

Reasoning (REA) Posterior estimates indicated that students in the IG exhibited a 122% higher frequency of reasoning in their contributions (gamma part). The evidence ratio (ER = 249) indicated very strong Bayesian support for a positive intervention effect, despite some uncertainty with regard to the exact effect size.

Elaborated explanations (EE) Students in the IG also elaborated more in their explanations. The frequency of EEs was 93% higher among students in the IG than among those in the CG (ER = 230, gamma part). The likelihood of completely omitting elaborated explanations was 92% lower in the IG (hurdle part). The high evidence ratio signals very strong Bayesian support (ER = 5999).

Cumulative talk (CT) Posterior estimates indicated that, at posttest, cumulative talk occurred 58% more often in the IG than in the CG among those who engaged in it (gamma part); notably, there was very strong Bayesian support for this difference (ER = 137). The likelihood of no cumulative talk occurring at all was 87% lower in the IG than in the CG (hurdle part), also with very strong support (ER = 443).

Exploratory talk (ET) A similar pattern was observed for exploratory talk. At posttest, students in the IG produced 63% more exploratory talk than those in the CG (gamma part), with a strong effect (ER = 166). The likelihood that no exploratory talk occurred at all was 80% lower in the IG (hurdle part), again with strong evidence for this effect (ER = 67).

Table 4 Posterior Estimates from Bayesian Multilevel Regression Models Predicting Quality of Student Contributions

	Authentic Question (Students)[d] Est. [CI]	Turn Allocation (Students)[d] Est. [CI]	Reasoning[c] Est. [CI]	Elaborated Explanation[d] Est. [CI]	Cumulative Talk[d] Est. [CI]	Exploratory Talk[d] Est. [CI]
Fixed Effects						
Intercept (Hurdle/Gamma)	**2.09 [0.48, 3.97] / 0.80 [0.04, 1.55] /**	1.11 [–0.26, 2.63] / **1.13 [0.05, 2.48]**	- / **1.13 [0.40, 1.96]**	**0.39 [–0.85, 1.65] / 0.65 [0.09, 1.22]**	0.34 [–0.97, 1.65] / 0.19 [–0.37, 0.72]	1.31 [–0.04, 2.76] / –0.16 [–0.92, 0.59]
Pretest	–0.16 [–2.34, 2.11]	0.37 [–3.52, 5.47]	0.02 [–0.14, 0.17]	–0.10 [–0.60, 0.40]	0.71 [−0.67, 2.08]	−0.52 [−2.16, 1.18]
First Language	–0.42 [–1.60, 0.88]	−0.46 [–2.87, 2.35]	0.13 [–0.66, 0.96]	–0.03 [–1.10, 1.09]	0.72 [−0.20, 1.66]	0.28[–1.59, 2.15]
Grade	0.04 [–1.50, 1.74]	−1.35 [–3.57, 1.07]	0.43 [–0.34, 1.21]	–0.17 [–1.08, 0.83]	−0.79 [−2.29, 0.81]	−0.21 [–1.43, 1.02]
Condition (Hurdle/Gamma)	**–3.54 [–5.58, −1.66] / 0.70 [0.22, 1.18]**	–1.08[–2.86, 0.58] / 0.49 [–0.07, 1.06]	- / **0.80 [0.26, 1.27]**	**–2.49 [–3.99, −1.04] 0.66 [0.17, 1.13]**	**–2.05 [–3.54, −0.63] 0.46 [0.09, 0.83]**	**–1.62 [–3.11, −0.17] / 0.49 [0.10, 0.86]**
Random Effect						
Teacher: SD (Intercept)	0.50 [0.07, 1.15]	0.64 [0.03, 1.91]	0.19 [0.01, 0.59]	0.18 [0.01, 0.60]	0.37 [0.02, 0.98]	0.37 [0.01, 1.26]
Model Summary						
Bayesian R^2	0.57 [0.24, 0.85]	0.26 [0.03, 0.50]	0.55 [0.31, 0.70]	0.40 [0.20, 0.56]	0.52 [0.18, 0.77]	0.25 [0.04, 0.48]
Odds Ratio	0.03 [0.00, 0.19]	0.34 [0.06, 1.79]	-	0.08 [0.02, 0.35]	0.13 [0.03, 0.53]	0.20 [0.04, 0.85]
Rate Ratio	2.02 [1.24, 3.25]	1.64 [0.93, 2.88]	2.22 [1.30, 3.57]	1.93 [1.19, 3.11]	1.58 [1.09, 2.28]	1.63 [1.11, 2.37]

[c]Frequency per 10-minute discussion, modeled using gamma regression

[d]Frequency per 10-minute discussion, modeled using hurdle-gamma regression

Bold values indicate that the 95% Bayesian credible interval does not include zero and are interpreted as evidence of an effect

Pretest = Baseline value of the respective outcome at t_0; First language (1 = Other than German, 0 = German); Grade (1 = 5th grade, 0 = 4th grade); Condition (1 = Intervention group, 0 = Control group). Bayes R^2 = Proportion of variance explained by fixed effects

Discussion

This study investigated the effects of a QT-based literary intervention on teachers' monologic and dialogic practices and the quality of students' contributions in discussions about literary moral dilemmas in Switzerland. Analyses of 20 videotaped small-group discussions at pretest and posttest revealed that the intervention effectively reduced monologic teacher practices and strengthened scaffolding, which was, in turn, accompanied by a marked improvement in the quality of students' questions and responses during discussions. In contrast to prior research on text-based discussions, this study focused exclusively on literary fiction, thereby contributing to the literature by demonstrating that literary education represents a stimulating context not only for language and aesthetic learning but also for reasoning and critical thinking.

In the Swiss context, empirical studies have shown that everyday literature discussions are frequently characterized by a high degree of teacher control and few opportunities for open, student-led dialogue (Fässler et al., 2019; Schulz et al., 2010). Against this backdrop, the observed reduction in monologic practices and concurrent increase in dialogic scaffolding are particularly noteworthy and, as Freire (2000) argues, far from self-evident. Pedagogical approaches cannot simply be transferred wholesale from one educational context to another; rather, they must be attuned to the cultural and contextual realities of participating teachers and students. In the present study, the fundamental structure of the QT approach was preserved. However, we adapted the QT approach to the literary education context by selecting high-quality children's literature and training children to generate different types of authentic questions. These adapted questions encouraged children to explore characters' mental states and relationships, to personally connect with the story and to engage in moral reasoning by critically reflecting on the consequences of characters' attempts for problem-solving. Thus, our adapted intervention provides a model of how QT can be used to promote a critical-analytic stance specifically in literary education.

In line with our first hypothesis (H1), teachers in the IG asked fewer TQs, reduced their speaking time, and managed turn-taking less frequently than those in the CG. Notably, the intervention not only reduced overt forms of teacher control but also attenuated subtler mechanisms, such as evaluative feedback. The decrease in evaluations may point to a shift in teachers' underlying epistemic beliefs: They no longer viewed their textual interpretation as the only (or even primary) interpretation but rather as one among many potential interpretations (Pulles et al., 2022). More broadly, these changes reflect a meaningful transition from monologic to dialogic discussion patterns (Mehan, 1979; Nystrand et al., 2003) and align with dialogic principles of shared epistemic authority (Alexander, 2008).

While some teachers refrained entirely from using test questions, others, to a more limited extent, retained them alongside evaluative feedback and turn allocation. As Magirius et al. (2023) have explained, the occasional use of test questions may be pedagogically justified when misconceptions at the level of basic text comprehension cannot be resolved through open discussion alone. Nevertheless, such situations should be managed in ways that align with the overarching goal of fostering dialogic exchanges rather than reverting to monologic patterns. Tengberg et al. (2022) cautioned that teachers may be tempted in such moments to provide help by taking it over instead of offering scaffolded support, limiting students' opportunities to engage in their own meaning-making.

Therefore, monologic practices should not be regarded as mere by-products that automatically decline as dialogic practices become increasingly common. Rather, teacher training should explicitly address their coexistence, providing targeted support to help teachers modify entrenched patterns of evaluation and turn regulation while simultaneously expanding their dialogic repertoires. This study highlights the importance of encouraging teachers not only to strengthen their dialogic practices but also to critically reflect on less effective patterns that may persist in classroom or small-group discussions.

Regarding our second hypothesis (H2), the findings also indicate that, dialogic teacher practices improved significantly over the course of the intervention, with a strong evidence for this effect. Teachers in the IG increasingly employed talk moves that prompted students to elaborate, justify, or reconsider their positions, thereby fostering deeper engagement with the text. Beyond this increase in frequency, the likelihood of omitting talk moves entirely was substantially lower in the IG than in the CG. This improvement in effective scaffolding aligns with findings from prior QT studies (see Murphy et al., 2018). The findings on teachers' use of AQs strayed from our expectations. Teachers in the IG asked fewer AQs overall than those in the CG and were more likely to omit them entirely. One potential explanation for this result is that they deliberately reduced their questioning as students increasingly demonstrated agency by posing their own AQs. This interpretation aligns with the findings of Murphy et al.'s (2018) intensive longitudinal study, which tracked changes across 12 measurement points: Teachers initially increased their use of AQs following QT training but gradually stepped back as students assumed greater responsibility for questioning.

This shift among teachers from monologic to dialogic practices in small-group discussions was associated with improvements in the quality of students' contributions. Students in the IG were more likely to ask AQs than those in the CG. Notably, identifying AQs requires higher-order thinking skills, meaning that students' questioning during the discussions reflects deep cognitive engagement with the text. In addition to teachers' scaffolding during the discussions themselves, the explicit instruction given to students through mini-lessons on effective questioning further supported their ability to independently identify and generate AQs (Murphy & Firetto, 2018). QT's goal of prompting students to ask their own

Table 5 Example of a Dialogic Literature Discussion (Posttest, IG)

Speaker	Dialogue	Individual Code	Cross-Turn Code
S1	Which would you prefer: to be in a group with Anna and Max or with Lorenz? Why?	AQ	
S2	I think I would prefer to be with Lorenz because he's innocent. Besides, I don't want to get involved in the pranks of the two.	EE	ET (1)
S3	I understand what you mean; Anna and Max sometimes tease other kids, and I don't like that either. But if I were in Lorenz's group, I might end up as the outsider, and everyone in the class might also laugh at me.	EE	(2)
S4	Like S3 said, I would also go with Anna and Max. However, if they wanted to prank someone, I would try to stop them.	REA	(3)
T	Do you have any other reasons that might support your decision to go with Anna and Max?	TM	
S4	They have helped Susa before, when the class laughed at her.	REA	

T = teacher; S = student; AQ = authentic question; EE = elaborated explanation; TM = talk move; REA = reasoning; ET = exploratory talk (turns 1–3)

questions is quite distinctive. In other discussion approaches (e.g., Correnti et al., 2021), only teachers are trained to ask cognitively activating questions about the text.

However, our hypothesis that students would adopt their teacher's discourse-management strategies was not supported by the findings. More specifically, we found no evidence that students in the IG engaged in turn allocation more frequently. This could be attributable to the fact that explicit turn allocation was simply no longer necessary, as speaking rights were negotiated naturally through, for example, eye contact, interruptions, or pauses. More fine-grained analyses of nonverbal interactions are necessary to better understand how students allocate speaking rights during student-centered discussions.

Crucially, we also found that the intervention improved the argumentative quality of student contributions. Students in the IG produced more REAs and EEs, with compelling evidence for these effects, and demonstrated a heightened capacity to build on one another's contributions over several turns, collaboratively engaging in textual interpretation (CT). They also exhibited a stronger ability to adopt a critical stance toward or even counter their peers' responses (ET), supported by strong evidence. This is evident in Table 5, which presents an example of a posttest discussion in the IG. Students posed AQs, engaged in extended reasoning, and built on peers' contributions—all with minimal teacher intervention. The teacher's talk moves reflected contingent scaffolding rather than control. In contrast, the CG example presented in Table 6 is dominated by test questions, evaluative feedback, and monologic patterns (e.g., IRE sequences, overfacilitation), reflecting a teacher-led structure with limited opportunities for student ownership.

The overall positive results concerning the argumentative quality of students' contributions may be explained in three ways. First, the mini-lessons effectively equipped students with the skills necessary to support their claims with reasons and evidence and to critically engage with peers' arguments. Second, the reflective activities following each discussion encouraged students to assess both their individual and collective discussion practices and to set clear goals for future discussions. Such metacognitive reflection provides children

Table 6 Example of a Nondialogic Literature Discussion (Posttest, CG)

Speaker	Dialogue	Individual Code	Cross-Turn Code
T	What does the teacher want Susa to do?	TQ	IRE (1)
S1	To tell her the names.	A	(2)
T	Right. What if Susa had told the teacher Max and Anna's names?	EAAQ	(3)
S2	Maybe they would be the outsiders.	A	
S3	But they are popular in class and at school, so that is rather unlikely.	REA	
T	That's what the text said—that they're popular in class. So can you imagine that they would become outsiders?	AQ	OF (1)
S4	I don't think so, because it's the other way round: if you're always the nasty person and then you do something good, the others still won't like you as much because they know that you did all that.	REA	(2)
T	Okay. They keep it in mind. Mmm. What would you do in Susa's place?	EA AQ	(3)

T = teacher; S = student; TQ = test question; AQ = authentic question; A = simple answer; REA = reasoning; EA = evaluation; IRE = initiation–response–evaluation pattern (turns 1–3); OF = overfacilitation pattern (turns 1–3)

with individualized feedback, which helps them set new goals for future discussions and refine their discussion skills over time (Hattie et al., 2021). Third, by employing talk moves, teachers in the IG were able to facilitate discussions more effectively. These teacher talk moves are designed to prompt students to justify and substantiate their viewpoints more thoroughly. Such forms of scaffolding not only improve the clarity, acceptability, and logical validity of students' arguments but also serve as a catalyst for further argumentation, supporting the development and consolidation of knowledge (Wilkinson et al., 2023). We attribute teachers' improved ability to support students more purposefully to the employed strengths-based coaching approach, which has previously been demonstrated to improve teachers' observational skills, self-efficacy, autonomy, and professional competence (Zwart et al., 2015). Through its focus on leveraging teachers' existing strengths and providing constructive feedback, the coaching enabled teachers to effectively refine their facilitation techniques.

In the present study, apart from the proportion of teacher talk, no effects of students' home language or grade level were observed on teachers' discourse practices or the quality of student contributions. The absence of such effects may be attributable to methodological constraints, such as the small sample size and strong intervention effects masking smaller influences, especially given the sparse distribution of several count-based outcomes. Alternatively, it may indicate that dialogic discussions can be fostered regardless of students' language background or grade level when teachers receive targeted support in dialogic teacher practices.

Limitations and future research

This study yielded significant insights into the effectiveness of the text-based QT discussion approach in German-language literary education. Several limitations, however, must be considered when interpreting the results.

While the small-group format aligns with the design principles of QT and facilitates high-quality interaction and targeted scaffolding of individual students, it may limit the generalizability of our findings to whole-class contexts. Moreover, small-group discussions pose challenges for classroom organization, as teachers must design meaningful learning activities and effectively manage the behavior of students who are not participating in the discussions. Future research should focus on adapting and further developing the intervention for whole-class settings and on examining its sustained impact over time.

The inclusion of only two measurement points (pretest and posttest) does not allow fine-grained analyses of changes in discussion practices such as potential non-linear growth patterns. Longitudinal designs with multiple time points would allow for a more nuanced analysis of the dynamic interplay between teachers' monologic and dialogic teaching practices. Moreover, we do not know how sustainable the intervention effects are because the study did not include a follow-up measurement.

Although the use of identical texts in both groups' pretest and posttest could ensure measurement reliability, it carried the risk of a test–retest effect stemming from increased

familiarity with the material. Future research could employ a crossover design to systematically rotate texts within the same cohort to increase internal validity.

Finally, the small and imbalanced sample across the IG and CG limits the statistical power and generalizability of our results. Although we employed a Bayesian hierarchical modeling approach to account for the clustered data structure and maximize the use of available data, very small samples remain a methodological challenge (Hox et al., 2020). Such models partially "shrink" estimates for smaller or underrepresented clusters toward the overall mean, reducing random error but potentially attenuating genuine between-teacher differences. While the sensitivity analyses with alternative priors produced consistent results, the findings should nevertheless be interpreted with caution. However, the issue of small samples is not necessarily a specific limitation of our study; small samples are characteristic of research on dialogic classroom discussions, which usually include an average of 84.28 participants (Murphy et al., 2009). This limitation stems from the resource-intensive nature of TPD, which is necessary to successfully implement dialogic classroom discussions. Future research should examine how effective interventions may be adapted for broader implementation without sacrificing effectiveness.

Conclusion

Despite its limitations, our quasi-experimental study replicated and extended prior findings on the effectiveness of the QT approach. Consistent with earlier QT studies (e.g., Li et al., 2016; Murphy et al., 2018), our results showcase substantial improvements in teachers' discussion practices and the quality of student contributions. Over the course of the intervention, fourth and fifth graders evolved into engaged learners and thoughtful interpreters of literary texts (Wei & Murphy, 2018).

What sets this study apart from those that came before it is its detailed exploration of the transformation of teacher practices facilitated by a dialogic discussion approach to literary texts. Our research delves deeply into the dynamics of classroom interactions, demonstrating how QT empowers teachers to minimize their control over discussion practices and enables students to take responsibility for discussions. By stepping back and adopting a facilitative role, the teachers encouraged their students to develop their own questions, construct arguments, and critically engage with literary texts. Consequently, the results offer valuable insights into how teachers can effectively transition from monologic to dialogic discussions in order to foster a more student-centered learning environment in literary education.

Appendix A: Overview of the intervention and teacher professional development program

Table 7 Elements of the School-Based Intervention Program

Time	Element	Content	Material	Duration
Sep–Oct 2020	QT-mini lessons: Discourse elements	– Students explore various QT discourse elements that promote productive discussions – Each discourse element is introduced through a PowerPoint presentation in which students receive definitions, observe examples in animated video clips, and create their own examples. The following discourse elements are included: – *Types of authentic questions*: Questions that invite speculation, analysis, generalization, or connections to students' emotions, thoughts, and experiences, as well as uptake questions – *Types of student responses*: Responses in which students support their claims with evidence, build on others' contributions, or critically challenge them.	PowerPoint presentations Lesson plans Teacher handbook QT poster (with question types)	9 mini lessons
Nov 2020–Jun 2021	Book lessons	– Each of the four literary books[a] was read in two to three segments – The procedure for each segment of the text was consistent and included the following steps: – *Pre-reading activities*: Students received brief introductions to the main themes of the book (e.g., bullying) and made predictions about the text – *During-reading activities*: Quizzes and questions in the literacy journal supported basic text comprehension – *Post-reading activities*: Students formulated authentic questions – *Small-group discussions*: Students engaged in discussions about the text in their small groups, using their questions as a guide, with support from the teacher – *Post-discussion reflection*: Following the discussion, students reflected on their own and their group's discussion practices and formulated new discussion goals.	Lesson plans Teacher handbooks (summary, psychological and literary analysis, reading and discussion instructions) Talk move cards Literacy journals	8 to 12 lessons per book

[a]The intervention included the following literary books (German version): *The Very Persistent Gappers of Frip* (by George Sanders); *Brown (My Alter Ego Is A Superhero)* (by Håkon Øvreås); *Rabbit & Bear: Attack of the Snack* (by Julian Gough), *Balaclava Boy* (by Jenny Robson)

Table 8 Elements of the Teacher Professional Development Program

Time	Element	Content	Material	Duration
Sep 2020	Workshop I	– Provision of conceptual knowledge to enhance dialogic literature discussions – Introduction to the QT model and its application in literature classrooms – Analysis of two transcripts using the QT coding and observation manual – Practice and discussion of talk moves to promote cognitive depth in discussions, supported by video examples	PowerPoint presentation Transcripts Observation manual Video examples	1 half-day workshop 5 h
Sep 2020	Workshop II	– Introduction and exploration of the content, main themes, and literary characteristics of the four books – Practice and discussion of talk moves to promote dialogic discourse patterns in discussions, supported by video examples	PowerPoint presentation Teacher handbooks and project materials	1 half-day workshop 5 h
Nov 2020–Jun 2021	Individual coaching cycles	– The teacher recorded small-group discussions and submitted these videos to the coach – The coach reviewed the recordings, selected one discussion for in-depth analysis, transcribed it, and created short video clips of three effective segments. For each clip, the coach formulated a specific observation and reflection prompt – The teacher analyzed the transcript and video clips based on the coach's prompts, reflecting on their discourse practices and evaluating their impact on student learning – The teacher and coach engaged in digital coaching sessions to deepen the teacher's reflection, collaboratively develop strategies for more effective and frequent use of these discourse practices and set clear goals for future discussions – The coach provided the teacher with a summary of the coaching cycle and an action plan to guide subsequent discussions	Video Observation manual Transcript and video clips	3 to 4 cycles 2 h each

Appendix B: Short story

Revenge Is Sticky!

Nervously, Susa looked into the faces of her new classmates. She hoped she would soon make some friends in this class.

"This is Susa, your new classmate," said Mrs Schubert, and asked Susa to introduce herself.

"Well, yes, I'm Susa and… um…" Susa's words didn't really want to come out, "…and I'm new here!

"Oh really!" hissed a voice from the back row.

"And I like riding my skateboard around," Susa quickly added.

"Ah, that explains why you can't think straight. Too many falls on the head!" mocked another child. Soon, giggling came from all around the classroom. Well, almost all. A confi-

dent voice called out before Mrs Schubert could say anything: "Not at all! She's really good! I saw her jump five stairs before school. Super cool!"

It was Anna. And just like that, the laughter stopped. She and her friend Max were well-respected in class. They were popular with the other kids. Susa noticed that right away. She was happy Anna and Max had helped her. She had almost become an outsider from the very first minute. But luckily, things turned out differently. During recess, Susa still felt a bit lost. Until Anna and Max came over to look at her skateboard. The three kids talked and soon realized they all loved sports. Even better—they all enjoyed cruising through the city. Susa with her skateboard, Max with his BMX bike, and Anna with her scooter. Susa felt relieved. After that messed-up start, things were now going even better than expected. And when Anna and Max suggested meeting later at Anna's place to listen to music, Susa could hardly believe her luck. Time flew by with Anna and Max. They laughed until their stomachs hurt and sang along to their favorite songs. Susa felt happy. She had found new friends very quickly. With Anna and Max by her side, she felt safe and strong.

Nothing at this new school could go wrong now. It was easy to see—they were a good team. Like real friends, the kind who help you when you're in trouble. And they were hilarious, too. There was always something to laugh about with them. That also included making fun of others from time to time. For example, Lorenz was a boy in the last row who wore unusual clothes. Max pointed at him: "That's Lorenz. He's a total loser, even in sports. That's why we call him a Loser. If he's on your team, you'll lose every game!" Anna added, "He's also a traitor. He ran straight to Mrs. Schubert to tell her we'd forgotten our homework. That was mean. Lorenz doesn't care about his classmates at all!"

And Susa thought she remembered that one of the mean comments from yesterday had come from Lorenz's direction. In her head, she still heard the laughter from the day before. She immediately told the two of them what she suspected. "He doesn't even know you! That's just mean of him!" said Max. And Anna agreed angrily: "It's time he learned a lesson."

The plan against Lorenz was made quickly: Before the first lesson, the three stuck a big, sticky chewing gum on the edge of Lorenz's desk. A few moments later, Mrs Schubert cried: "Eww! What's that?!" Bullseye! She had touched it while checking homework, and now a long strand of gum hung between her fingers. Lorenz was immediately told to clean the desks that afternoon. At first, Susa couldn't help but smirk, but later she felt a little sorry for the loner. Not Anna and Max.

Annoyed that Mrs Schubert now kept a close eye on them during every test, they decided to go ahead with prank number two. Susa didn't know anything about it. When Mrs Schubert got up from her chair and walked to the blackboard, everyone could see it clearly: Bullseye number two! A piece of neon-yellow chewing gum was stuck to Mrs Schubert's backside. The class went wild. It took a moment before the teacher noticed what had happened. With her face turning red, she sent Lorenz, whom she also blamed for this second gum disaster, out of the room. She followed him angrily.

"But it wasn't me, Mrs Schubert! Really, really not!" he begged—and sounded quite convincing. The noise in the classroom grew louder and louder, so Mrs Schubert had to come back inside. After that, she didn't take her eyes off the class for a second. So she didn't miss it when a note slid under the desks toward Susa. She quickly grabbed it. She had caught it. "Revenge is sticky" was written on it. Susa was shocked. And not even the school bell offered an escape: "Susa, you stay here, please. Everyone else may go," said Mrs Schubert.

Max brushed past Susa on his way out and whispered, “You’re with us, okay? We’re friends and helped you on your first day, too!”

Now, Susa was alone in the room with Mrs Schubert, being questioned. She didn’t know her teacher yet, but she knew one thing: There was no easy way out. She would have to give names. What should Susa do?

Appendix C: Coding system

Table 9 Overview of the Coding System

Discourse indicator	Initiator	Description	Example
Test Question (TQ)	Teacher	A TQ presupposes a particular answer. The answer is explicitly stated in the text or is generally known	– What prank was played on Lorenz? – What is Cherise’s nickname?
Authentic Question (AQ)	Teacher or student	An AQ is an open-ended question that requires thinking about, around, and with the text. The responses are not pre-specified. It usually allows for a range of answers and generates several responses	– What does the author want to tell us about this story? – What do Bruno’s friends mean to him? – How else could Tara have solved the problem?
Talk Moves (TM)	Teacher	TM refers to questions or responses intended to support or scaffold students’ discourse, encourage students to engage with each other’s ideas, ask questions, and think critically (e.g., modelling, summarizing, marking, prompting, challenging)	– Why do you think that? (Prompting) – Let’s just pause and summarize what we heard. (Summarizing) – Really nice explanation. You used evidence and excellent reasoning. (Marking) – If I were to respond to Tom’s question with a strong argument, I would say... (Modelling) – Did you also consider that… (Challenging)
Evaluation (EA)	Teacher	EA is coded for statements in which the teacher assesses a student’s response in terms of correctness or quality. EAs also include active listening signals from the teacher	– Yes/No – Hm (agreeing) – That’s right (not quite) – Very good
Reasoning (REA)	Student	REA is coded when students explain how they arrive at a conclusion or idea and support their claim with a reason or a piece of evidence	– It is not fair to punish him, because it wasn’t done on purpose. It was an accident
Elaborated explanation (EE)	Student	An EE is an individual response in which a student explains how they arrive at a conclusion or idea and supports their claim with at least two reasons and/or pieces of evidence	– If I were her, I wouldn’t help the neighbors because they didn’t help her when she was overwhelmed. The neighbors even said that it was her fault
Cumulative talk (CT)	Students	CT is characterized by speakers building positively but uncritically on what others have said over at least three turns. Participants use talk to construct “common knowledge” by accumulation	– S1: As a superhero, Bruno stands up for himself S2: I agree with Ruth. He also behaves differently when he puts on the costume S3: Yes, and his friends also become more confident.

Table 9 Overview of the Coding System

Discourse indicator	Initiator	Description	Example
Exploratory talk (ET)	Students	ET occurs when students share, evaluate, and build knowledge over at least three turns. Students reason collectively by challenging each other and responding to challenges with reasons and evidence. Episodes of exploratory talk are coded starting with the utterance that prompts the challenge (i.e., the turn immediately preceding the challenge)	– S1: I think it's okay that Bruno painted their bikes because they destroyed his hut S2: But when everyone takes revenge for things done to them, it always goes on like that S3: If it's just a small revenge, I find it acceptable.
Teacher/ Student Turn Allocation (TTA/STA)	Teacher or student	TTA/STA refers to discourse practices used to organize, direct, and control the flow of a discussion	– I'm interested in what you think about this, Anne? – Who else has a question about this topic? – Anyone else want to add something?

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11251-026-09809-z.

Acknowledgements We would like to thank the teachers and their classes who participated in our intervention. Additionally, we express our gratitude to our research team, especially Denise Krummenacher and Hanspeter Inauen, for their support in piloting the lessons, and coaching teachers.

Author contributions *Conceptualization*: Yvonne Dammert, Luciano Gasser, P. Karen Murphy; *Methodology*: Yvonne Dammert; *Formal analysis and investigation*: Yvonne Dammert; *Writing - original draft preparation*: Yvonne Dammert; *Writing - review and editing*: Yvonne Dammert, Luciano Gasser, P. Karen Murphy; *Funding acquisition*: Luciano Gasser; *Resources*: Yvonne Dammert, Luciano Gasser, P. Karen Murphy; *Supervision*: Luciano Gasser.

Funding Open access funding provided by University of Teacher Education of Bern. This work was supported by Stiftung Mercator Schweiz [grant number 2017-0026]; Stiftung éducation21 [grant number 17.R.066] and Fachstelle für Rassismusbekämpfung Schweiz.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability The codes used during this study are available from the corresponding author upon request.

Declarations

Competing Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval The study described in this paper was conducted in accordance with the ethical standards of the APA.

Consent to participate Written consent was obtained from the participating teachers and the parents or legal guardians of the children prior to the study. They were assured that participation was voluntary and could be withdrawn at any time without consequences.

Consent to publish The authors confirm that no identifiable personal data were collected or published in this study. All data have been anonymized, and it is impossible to trace back to the participants. Therefore, no additional consent for publication was required.

References

Alexander, R. (2008). *Essays on pedagogy*. Routledge.

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *333*(6045), 1034–1037. https://doi.org/10.1126/science.1207998

Asterhan, C. S. C., & Schwarz, B. B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist*, *51*(2), 164–187. https://doi.org/10.1080/00461520.2016.1155458

Bakhtin, M. (1984). *Problems of Dostoevsky's poetics*. University of Minnesota Press.

Böheim, R., Schnitzler, K., Gröschner, A., Weil, M., Knogler, M., Schindler, A. K., Alles, M., & Seidel, T. (2021). How changes in teachers' dialogic discourse practice relate to changes in students' activation, motivation and cognitive engagement. *Learning Culture and Social Interaction*, *28*, 100450. https://doi.org/10.1016/j.lcsi.2020.100450

Böhme, K., Bertschi-Kaufmann, A., Pieper, I., Fässler, D., Depner, S., Kernen, N., & Siebenhüner, S. (2018). Leseverstehen und literarische Bildung – Welche Schwerpunkte setzen Lehrpersonen in ihrem Deutschunterricht und welche Texte wählen sie aus? Erste Befunde der TAMoLi-Studie. *Leseforum*, 1–23. https://doi.org/10.58098/LFFL/2018/3/642

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Chen, G., Chan, C. K. K., Chan, K. K. H., Clarke, S. N., & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and student learning. *Journal of the Learning Sciences*, *29*(4–5), 642–680. https://doi.org/10.1080/10508406.2020.1783269

Chinn, C. A., Anderson, R. C., & Waggoner, M. A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly*, *36*(4), 378–411. https://doi.org/10.1598/RRQ.36.4.3

Clarke, S. N. (2015). The right to speak. In L. B. Resnick, C. S. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 167–180). American Educational Research Association.

Correnti, R., Matsumura, L. C., Walsh, M., Zook-Howell, D., Bickel, D. D., & Yu, B. (2021). Effects of online content-focused coaching on discussion quality and reading achievement: Building theory for how coaching develops teachers' adaptive expertise. *Reading Research Quarterly*, *56*(3), 519–558. https://doi.org/10.1002/rrq.317

Davies, M. J., & Esling, S. (2024). The use of Quality Talk to foster critical thinking in a low socio-economic secondary geography classroom. *The Australian Journal of Language and Literacy*, *47*(1), 143–143. https://doi.org/10.1007/s44020-024-00057-7

Davies, M., Kiemer, K., & Meissel, K. (2017). Quality Talk and dialogic teaching—an examination of a professional development programme on secondary teachers' facilitation of student talk. *British Educational Research Journal*, *43*(5), 968–987. https://doi.org/10.1002/berj.3293

Deutschschweizer Erziehungsdirektoren-Konferenz (D-EDK). (2016). *Lehrplan 21*. D-EDK.

Fässler, D., Bertschi-Kaufmann, A., Pieper, I., & Böhme, K. (2019). *Student reading motivation and teacher aims and actions in literature education in lower secondary school*. *RISTAL: Research in Subject-matter Teaching and Learning, 2*(1), 118–139.

Freire, P. (2000). *Pedagogy of the oppressed* (30th anniversary edition). Bloomsbury Publishing.

Gasser, L., Dammert, Y., & Murphy, P. K. (2022). How do children socially learn from narrative fiction: Getting the lesson, simulating social worlds, or dialogic inquiry? *Educational Psychology Review*, *34*(3), 1445–1475. https://doi.org/10.1007/s10648-022-09667-4

Goodwin, A. P., Cho, S. J., Reynolds, D., Silverman, R., & Nunn, S. (2021). Explorations of classroom talk and links to reading achievement in upper elementary classrooms. *Journal of Educational Psychology*, *113*(1), 27–48. https://doi.org/10.1037/edu0000462

Gough, J., & Field, J. (2018). *Rabbit & bear: Attack of the snack*. Hodder & Stoughton.

Hattie, J., Fisher, D., Frey, N., & Clarke, S. (2021). *Collective student efficacy: Developing independent and inter-dependent learners*. Corwin.

Hennessy, S., Calcagni, E., Leung, A., & Mercer, N. (2023). An analysis of the forms of teacher-student dialogue that are most productive for learning. *Language and Education*, *37*(2), 186–211. https://doi.org/10.1080/09500782.2021.1956943

Hox, J., McNeish, D., & Miočević, M. (2020). Small samples in multilevel modeling. In R. van de Schoot & M. Miočević (Eds.), *Small sample size solutions* (pp. 215–225). Routledge. https://doi.org/10.4324/9780429273872-18

Iordanou, K., Kuhn, D., Matos, F., Shi, Y., & Hemberger, L. (2019). Learning by arguing. *Learning and Instruction*, *63*, 101207. https://doi.org/10.1016/j.learninstruc.2019.05.004

Jurik, V., Gröschner, A., & Seidel, T. (2014). Predicting students' cognitive learning activity and intrinsic learning motivation: How powerful are teacher statements, student profiles, and gender? *Learning and Individual Differences*, *32*, 132–139. https://doi.org/10.1016/j.lindif.2014.01.005

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, *342*(6156), 377–380. https://doi.org/10.1126/science.1239918

Kim, M. (2022). Student agency and teacher authority in inquiry-based classrooms: Cases of elementary teachers' classroom talk. *International Journal of Science and Mathematics Education*, *20*(8), 1927–1948. https://doi.org/10.1007/s10763-021-10233-7

Kuhn, D. (2018). A role for reasoning in a dialogic approach to critical thinking. *Topoi*, *37*(1), 121–128. https://doi.org/10.1007/s11245-016-9373-4

Lenhard, W., & Lenhard, A. (2014). *Berechnung des Lesbarkeitsindex LIX nach Björnson.* Retrieved from http://www.psychometrica.de/lix.html. https://doi.org/10.13140/RG.2.1.1512.3447

Lenhart, J., Richter, T., Appel, M., & Mar, R. A. (2023). Adolescent leisure reading and its longitudinal association with prosocial behavior and social adjustment. *Scientific Reports*, *13*(1), 9695. https://doi.org/10.1038/s41598-023-35346-7

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. https://doi.org/10.1016/j.cedpsych.2016.05.002

Lin, T. J., Jadallah, M., Anderson, R. C., Baker, A. R., Nguyen-Jahiel, K., Kim, I. H., Kuo, L. J., Miller, B. W., Dong, T., & Wu, X. (2015). Less is more: Teachers' influence during peer collaboration. *Journal of Educational Psychology*, *107*(2), 609–629. https://doi.org/10.1037/a0037758

Lin, T. J., Kraatz, E., Ha, S. Y., Hsieh, M. Y., Glassman, M., Nagpal, M., Sallade, R., & Shin, S. (2022). Shaping classroom social experiences through collaborative small-group discussions. *British Journal of Educational Psychology*, *92*(1), e12442. https://doi.org/10.1111/bjep.12442

Magirius, M., Scherf, D., & Steinmetz, M. (2023). Instructive dialogues on literary texts: A framework for dialogic teaching promoting high-level comprehension in the literature classroom. *L1-Educational Studies in Language and Literature*, *23*(2), 1–27. https://doi.org/10.21248/l1esll.2023.23.2.584

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, *3*(3), 173–192. https://doi.org/10.1111/j.1745-6924.2008.00073.x

Matsumura, L. C., Correnti, R., Walsh, M., Bickel, D. D., & Zook-Howell, D. (2019). Online content-focused coaching to improve classroom discussion quality. *Technology Pedagogy and Education*, *28*(2), 191–215. https://doi.org/10.1080/1475939X.2019.1577748

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, *28*(2), 295–314. https://doi.org/10.1007/s10648-014-9287-x

Mehan, H. (1979). What time is it, Denise? Asking known information questions in classroom discourse. *Theory Into Practice*, *18*(4), 285–294. https://www.jstor.org/stable/1476655

Mehan, H., & Cazden, C. B. (2015). The study of classroom discourse: Early history and current developments. In L. B. Resnick, C. S. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 13–34). American Educational Research Association.

Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Open University.

Murphy, P. K., & Firetto, C. M. (2018). Quality Talk: A blueprint for productive talk. In P. K. Murphy (Ed.), *Classroom discussions in education* (pp. 101–133). Routledge.

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' comprehension of text: A meta-analysis. *Journal of Educational Psychology*, *101*(3), 740–764. https://doi.org/10.1037/a0015576

Murphy, P. K., Greene, J. A., Firetto, C. M., & Butler, A. M. (2017). Analyzing the talk in Quality Talk discussions: A coding manual. *Pennsylvania State University*. https://doi.org/10.18113/s1xw64

Murphy, P. K., Greene, J. A., Firetto, C. M., Hendrick, B. D., Li, M., Montalbano, C., & Wei, L. (2018). Quality Talk: Developing students' discourse to promote high-level comprehension. *American Educational Research Journal*, *55*(5), 1113–1160. https://doi.org/10.3102/0002831218771303

Murphy, P. K., Greene, J. A., Firetto, C. M., Croninger, R. M. V., Duke, R. F., Li, M., & Lobczowski, N. G. (2022). Examining the effects of Quality Talk discussions on 4th- and 5th-grade students' high-level comprehension of text. *Contemporary Educational Psychology*, *71*, 102099. https://doi.org/10.1016/j.cedpsych.2022.102099

Nickl, M., Sommerhoff, D., Radkowitsch, A., Huber, S. A., Bauer, E., Ufer, S., Plass, J. L., & Seidel, T. (2024). Effects of real-time adaptivity of scaffolding: Supporting pre-service mathematics teachers' assessment skills in simulations. *Learning and Instruction*, *94*, 101994. https://doi.org/10.1016/j.learninstruc.2024.101994

Nucci, L., & Ilten-Gee, R. (2021). *Moral education for social justice*. Teachers College.

Nystrand, M. (2006). Research on the role of classroom discourse as it affects reading comprehension. *Research in the Teaching of English*, *40*(4), 392–412. https://www.jstor.org/stable/40171709 JSTOR.

Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes*, *35*(2), 135–198. https://doi.org/10.1207/s15326950dp3502_3

O'Connor, C., & Michaels, S. (2019). Supporting teachers in taking up productive talk moves: The long road to professional learning at scale. *International Journal of Educational Research*, *97*, 166–175. https://doi.org/10.1016/j.ijer.2017.11.003

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. PISA, OECD Publishing.

Øvreås, H., Torseter, Ø., & Dickson, K. (2019). *Brown (My alter ego is a superhero)*. Enchanted Lion books.

Pulles, M., Berenst, J., De Glopper, K., & Koole, T. (2022). Children's discussions about texts: Integrating and evaluating practices. *Linguistics and Education*, *69*, 101051. https://doi.org/10.1016/j.linged.2022.101051

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rapanta, C., & Macagno, F. (2023). Authentic questions as prompts for productive and constructive sequences: A pragmatic approach to classroom dialogue and argumentation. *Dialogic Pedagogy: An International Online Journal*, *11*(3), A65–A87. https://doi.org/10.5195/dpj.2023.546

Reusser, K., & Pauli, C. (2015). Co-constructivism in educational theory and practice. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 913–917). Elsevier.

Reznitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist*, *48*(2), 114–133. https://doi.org/10.1080/00461520.2013.775898

Robson, J. (2014). *Balaclava boy*. Little Island Books.

Rosenblatt, L. M. (1978). *The reader, the text, the poem: The transactional theory of the literary work*. Southern Illinois University.

Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2016). How teacher emotional support motivates students: The mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction*, *42*, 95–103. https://doi.org/10.1016/j.learninstruc.2016.01.004

Saunders, G., & Smith, L. (2000). *The very persistent Gappers of Frip* (1st ed.). Villard.

Schrijvers, M., Janssen, T., Fialho, O., & Rijlaarsdam, G. (2019). Gaining insight into human nature: A review of literature classroom intervention studies. *Review of Educational Research*, *89*(1), 3–45. https://doi.org/10.3102/0034654318812914

Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). *ICCS 2009 international report: Civic knowledge, attitudes, and engagement among lower-secondary students in 38 countries*. IEA.

Scott, P. H., Mortimer, E. F., & Aguiar, O. G. (2006). The tension between authoritative and dialogic discourse: A fundamental characteristic of meaning making interactions in high school science lessons. *Science Education*, *90*(4), 605–631. https://doi.org/10.1002/sce.20131

Sedova, K., Salamounova, Z., & Svaricek, R. (2020). *Getting dialogic teaching into classrooms: Making change possible*. Springer.

Smid, S. C., McNeish, D., Miočević, M., & Van De Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 131–161. https://doi.org/10.1080/10705511.2019.1577140

Soter, A. O., Wilkinson, I. A., Murphy, P. K., Rudge, L., Reninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal of Educational Research*, *47*(6), 372–391. https://doi.org/10.1016/j.ijer.2009.01.001

Tengberg, M., Blikstad-Balas, M., & Roe, A. (2022). Missed opportunities of text-based instruction: What characterizes learning of interpretation if strategies are not taught and students not challenged? *Teaching and Teacher Education*, *115*, 103698. https://doi.org/10.1016/j.tate.2022.103698

Van Breukelen, G. J. P. (2023). Cluster randomized trials with a pretest and posttest: Equivalence of three-, two- and one-level analyses, and sample size calculation. *Multivariate Behavioral Research*, *59*(2), 206–228. https://doi.org/10.1080/00273171.2023.2240779

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

Van de Schoot, R., Depaoli, S., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), 1–26. https://doi.org/10.1038/s43586-020-00001-2

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

Wei, L., & Murphy, P. K. (2018). Teacher and student roles: Walking the gradually changing line of responsibility. In P. K. Murphy (Ed.), *Classroom discussions in education* (pp. 30–53). Routledge.

Wei, L., Murphy, P. K., & Firetto, C. M. (2018). How can teachers facilitate productive small-group talk? An integrated taxonomy of teacher talk moves. *The Elementary School Journal*, *118*(4), 578–609. https://doi.org/10.1086/697531

Wilkinson, I., 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). Guilford Press.

Wilkinson, I. A. G., Reznitskaya, A., & D'Agostino, J. V. (2023). Professional development in classroom discussion to improve argumentation: Teacher and student outcomes. *Learning and Instruction*, *85*, 101732. https://doi.org/10.1016/j.learninstruc.2023.101732

Zitzmann, S., Helm, C., & Hecht, M. (2021). Prior specification for more stable bayesian estimation of multilevel latent variable models in small samples: A comparative investigation of two different approaches. *Frontiers in Psychology*, 11. https://doi.org/10.3389/fpsyg.2020.611267

Zook-Howell, D., Matsumura, L. C., Walsh, M. W., Correnti, R., & Bickel, D. D. (2020). Developing adaptive expertise at facilitating dialogic text discussions. *The Reading Teacher*, *74*(2), 179–189. https://doi.org/10.1002/trtr.1921

Zwart, R. C., Korthagen, F. A. J., & Attema-Noordewier, S. (2015). A strength-based approach to teacher professional development. *Professional Development in Education*, *41*(3), 579–596. https://doi.org/10.1080/19415257.2014.919341