



Effects of a literary intervention on interaction quality in small-group discussions in the upper elementary grades

Luciano Gasser^{a,*}, David Preisig^a, Anna Frei^a, Yvonne Dammert^a, Sara Egger^a,
P. Karen Murphy^b

^a Bern University of Teacher Education, Fabrikstrasse 8, 3012, Bern, Switzerland

^b The Pennsylvania State University, 102 CEDAR Building, University Park, PA, 16802, USA

ARTICLE INFO

Keywords:

Interaction quality
Social development
Literary education
Discourse
Argumentation

ABSTRACT

Background: Literary discussions represent a promising interaction context for the development of social, linguistic, and cognitive skills among children and adolescents. The one-year intervention of this study is based on a critical-analytic approach (Gasser et al., 2022; Murphy et al., 2009), focusing on the argumentative and inclusive quality of interactions in small group discussions about high-quality children's literature.

Aims: The goal is to study the effectiveness of this literary intervention on observed socio-emotional and instructional interaction quality in small group discussions about a moral dilemma text.

Sample: The sample included 51 teachers and 159 small groups from fourth- and fifth-grade classrooms.

Method: The study is based on a cluster-randomized control group design with three measurement occasions, considering the multi-level structure of the data (L1: measurement occasions, L2: small groups, L3: teachers). Interaction quality in discussions was measured by the Classroom Assessment Scoring System (CLASS).

Results: Multilevel growth curve analyses show positive changes in both socio-emotional and instructional interaction quality in small group discussions in the intervention group, but not in the control group.

Conclusions: The results are discussed with reference to the potential of literary discussions for an integrated approach to promoting socio-emotional and academic learning.

1. Introduction

Discussions about narrative fiction in family and school present a unique developmental context for socio-emotional and cognitive growth. For instance, shared book reading in early childhood is associated with better reading and social outcomes (Dowdall et al., 2020; Venkadasalam et al., 2022). Moreover, research indicates that shared book reading contributes more effectively to language or social-cognitive development than other conversational contexts (e.g., narrative talk or play; Ece Demir-Lira et al., 2019; Noble et al., 2018; Tompkins et al., 2018).

The effectiveness of this developmental context can be explained by its unique interaction quality (Clingenpeel & Pianta, 2007a; Landry et al., 2011; Nyhout & O'Neill, 2013; Schrijvers et al., 2019). In discussions about narrative fiction children analyze the problem of the story, speculate about the inner lives and relationships between characters, and make inferences about the text's moral messages. This

higher-order thinking is embedded in a collaborative process, where children and their peers, parents, or teachers work towards a shared perspective on the text. This includes children listening to each other, arguing for their own positions, being open to other perspectives, and respectfully disagreeing.

Various text-based discussion approaches for upper elementary school children use this important developmental context to promote argumentative skills and critical-analytic thinking (Lawrence et al., 2015; Lin et al., 2022; Wilkinson et al., 2023). For example, in Quality Talk (QT; Murphy et al., 2022) or Collaborative Reasoning (CR; Lin et al., 2022), children in small groups of four to six engage in text-based discussions with children taking interpretative authority over the text and cooperatively establishing balanced text interpretations.

This study examines the effects of a literary intervention on observed interaction quality in small-group discussions among fourth and fifth graders. Previous text-based discussion approaches predominately focus on promoting instructional interaction quality. The literary intervention

* Corresponding author. Bern University of Teacher Education, Switzerland.

E-mail addresses: luciano.gasser@phbern.ch (L. Gasser), david.preisig@phbern.ch (D. Preisig), anna.frei@phbern.ch (A. Frei), yvonne.dammert@phbern.ch (Y. Dammert), sara.egger@phbern.ch (S. Egger), pkm15@psu.edu (P.K. Murphy).

<https://doi.org/10.1016/j.learninstruc.2024.101996>

Received 28 March 2024; Received in revised form 9 August 2024; Accepted 10 August 2024

Available online 9 September 2024

0959-4752/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

of this study builds on QT, but additionally integrates social learning objectives that address children's social inclusion in peer groups. We therefore investigate effects of the literary intervention on both socio-emotional and instructional interaction quality. Moreover, previous research on collaborative small-group discussions often neglected the nested structure of the data (i.e., small groups are nested within classrooms), mostly due to small samples (Janssen et al., 2013). This study examines changes in interaction quality both at the level of small group and teacher level based on a randomized control group design.

1.1. Interaction quality in text-based discussions

From the perspective of ecological and transactional developmental theories (Bronfenbrenner & Morris, 1998; Sameroff, 2009), child development evolves through a dynamic interplay between systems at different levels. More distal systems (e.g., teachers' professional training) exert their influence on children via the immediate interactions between a child and significant others (e.g., teachers and peers; Pianta et al., 2021). These interactions are conceptualized as the primary mechanisms underlying social, cognitive, and language development (Bronfenbrenner & Morris, 1998). From this theoretical perspective, the primary goal of intervention efforts should focus on leveraging the quality of interactions in educational settings (Allen et al., 2011; Pianta & Hamre, 2009). Moreover, these theories call for multi-level designs to study how systems at different levels contribute to interaction quality (e.g., small groups vs. classrooms). Interaction quality is often described within two different broad domains – the socio-emotional and instructional domain (e.g., Landry et al., 2011; Pianta et al., 2007). We next characterize high-quality interaction in these domains in the specific context of text-based discussions.

1.1.1. Socio-emotional interaction quality in text-based discussions

Text-based discussion in the family and school is an inherently interpersonal activity, which is most likely to succeed within positive social relationships (Bus et al., 1997; Clingenpeel & Pianta, 2007b; Kraatz et al., 2020). Based on attachment theory, researchers defined emotional interaction quality in text-based discussions by characteristics such as warmth (e.g., praise and enthusiasm), sensitivity and responsiveness (e.g., recognizing a lack of interest), or support for autonomy (e.g., following the interests of the child) (Bus et al., 1997). High emotional support enhances children's perceptions of their parents or teachers as a source of security and support, which results in higher attention, engagement, and cooperation during shared readings (e.g., Blewitt & Langan, 2016; Landry et al., 2011).

The affective quality of interactions is relevant not only in early childhood but also in late childhood (Gasser et al., 2018; Sette et al., 2020). Peer groups characterized by positive affect, inclusion, respect, and fairness show higher productivity in collaborative work (Baucal et al., 2023). Microgenetic studies of interaction patterns in discussions of fourth and fifth graders about a moral dilemma story showed that both the positive communication of the teacher (i.e., praise for evidence-based references) and the children (i.e., social support, mutual respect, equal participation) contributes to the cognitive depth of the discussions (e.g., more elaboration and cumulative talk) (Chen et al., 2023; Jadallah et al., 2011). Finally, children and adolescents develop an increasing interest for autonomy and leadership in discussions with peers and adults (Allen et al., 1994). Various studies show that students who can take responsibility and control over the content and form of text-based discussions (e.g., control over turn-taking) develop higher argumentative and critical-analytic thinking skills than if the teacher dominates and controls the talk (Murphy et al., 2009).

1.1.2. Instructional interaction quality in text-based discussions

Text-based discussions also present a unique interaction context regarding cognitive and language development. High-quality discussions about children's books in early childhood include abstract and

decontextualized talk that transcends the here and now (i.e., the immediate and obvious elements of a story) by elaborating on the mental states of characters, making connections to the children's lives, and contrasting readers' different perspectives on the text (Adrián et al., 2007; Aram et al., 2013; Fitton et al., 2018). These interactions effectively stimulate basic language and social-cognitive abilities such as vocabulary or emotion understanding in younger children (Tompkins et al., 2018; Wasik et al., 2006).

For older children, it is increasingly important to engage in interactions that not only stimulate basic social-cognitive and language competencies but also complex cognition such as critical thinking, argumentative skills, and social reasoning (Gasser et al., 2022; Murphy et al., 2009). Such interactions are characterized by children identifying their own questions and answering them with reference to reasons and evidence. Children speculate about alternative narrative trajectories, evaluate characters' approaches for problem-solving and make inter-textual and personal references (Li et al., 2016; Schrijvers et al., 2019; Wilkinson et al., 2023). The children's responses build cumulatively on each other, extending the arguments of others and critically examining their argumentative validity (Reznitskaya & Wilkinson, 2021).

Moreover, discussions that focus on the big question of the narrative fiction (e.g., "Is it okay or not okay how the protagonist solves the problem?") and explore it from various perspectives, have high potential to promote deep insights into social complexities (Gasser et al., 2022; Lin et al., 2019). In high-quality text-based discussions, the teacher's feedback is characterized by scaffolding, where the teacher supports the children's argumentation quality as little as possible but as much as necessary (Lin et al., 2015). This includes facilitation strategies such as modeling effective discourse elements (e.g., "I now provide textual evidence for my claim. On page 54, I read ..."), eliciting justifications and evidence (e.g., "How do you know that?"), and inviting the children to consider alternative perspectives on the text (e.g., "Did you also consider the possibility that ... ?", Li et al., 2016). These characteristics of text-based discussions significantly relate to children's argumentation skills, higher-level text-comprehension, perspective-taking and social reasoning (Li et al., 2016; Lin et al., 2012; Walker et al., 2000).

1.2. Intervention research on text-based discussion approaches

Text-based discussion approaches strongly differ in their strategies and objectives for promoting cognitive and social outcomes. A taxonomy with three reader stances, which specifies how readers approach a text, has proven useful in systematizing different text-based discussion approaches (Gasser et al., 2022; Murphy et al., 2009). In efferent discussion approaches, the primary goal is to establish basic text comprehension, such as understanding vocabulary or the main message and supporting details of the text. Typically, in these approaches, the teacher is more likely to take control of the discussion and guide students towards an accurate understanding of the text. Expressive-oriented discussion approaches are more student-centered because they encourage spontaneous and personal responses of children to the text. However, they do not address the argumentative quality of these responses. Finally, critical-analytic discussion approaches (e.g., CR, QT) are also student-centered but additionally strongly support quality of questioning and argumentation and thus most effectively promote interactions relevant for the development of critical thinking and social reasoning (Murphy et al., 2009). For example, these approaches often introduce discussion rules that establish norms about how students can rationally engage in discussions (e.g., "Provide reasons for your claims.;" Baker et al., 2017). Moreover, teachers explicitly teach children how to ask authentic questions about the text and how to answer with arguments (Murphy et al., 2022). Teachers further facilitate discussions with talk moves that support effective questioning and argumentation (e.g., prompting justifications; Lin et al., 2015).

A key feature of many critical-analytic discussion approaches is the small-group format. Small groups are viewed as a privileged social

context for actively involving children and giving them responsibility and control over the discussion (Chen et al., 2023). An important advantage of small groups compared to larger groups is their ability to establish more equal participation and to diminish the prevalence of monologic conversation patterns, where dominant speakers impose their views on passive listeners (Fay et al., 2000). For withdrawn children, small groups provide a safer space to contribute to discussions than large groups (Mundelsee & Jurkowski, 2021). Additionally, teachers in small groups can respond more sensitively to the children's specific academic and social needs (Li et al., 2016).

An example of a critical-analytic discussion approach with a small-group format is QT (Murphy et al., 2022). QT supports teachers' facilitation of small-group discussions and students' development of critical-analytic thinking skills, as demonstrated through various single-group studies that have shown instructional interaction quality effectively changes through QT (e.g., more elaboration and cumulative talk; Li et al., 2016; Murphy et al., 2017). The professional development model includes an initial training during which teachers are introduced to the basics of the discussion approach, as well as ongoing coaching as they implement QT. During coaching, the teachers watch and analyze discussions from their classes and then collaborate with a discourse coach on how to further support students' high-level thinking in future discussions.

The QT model includes four parts: pedagogical principles, an instructional frame, discourse elements, and talk moves. The pedagogical principles clarify beliefs that undergird QT, such as the belief that *talk is a tool for thinking and interthinking* and that teachers should *embrace space and diversity* during discussions. The instructional frame lays out the practical elements of implementing QT discussions, including factors such as determining groupings, ground rules for discussion, and teacher and student roles during discussion. For example, students are expected, over time, to take on responsibility for the flow of the discussion and interpretive authority of the content. The QT approach provides support as students learn to take on these responsibilities not only through teacher facilitation of the discussions, but also by providing explicit instruction to students about the discourse elements, which include question and response types. QT emphasizes the use of authentic questions, or open-ended questions that promote talk by prompting students to engage in reasoning and make connections between the content and prior knowledge or experiences. The QT response types are rooted in argumentation and teach students to support their ideas using evidence and reasoning as well as to challenge their own and others' thinking during discussion. Finally, teacher discourse moves provide teachers with ways to support students' practice with the discourse elements and to push their thinking, such as modeling of the discourse elements or prompting students for further explanation.

1.3. The current study

In the present study, we examine the effects of a literary intervention on interaction quality, considering that small groups are nested in teachers. We test two hypotheses. First, we expect that small groups of the intervention group will show a stronger increase in interaction quality than small groups of the control group (H1). Second, we expect positive effects of the intervention on both socio-emotional and instructional interaction quality because the literary intervention integrates social and academic learning (H2).

The literary intervention of this study is based on the QT approach but extends it in two important respects. First, the intervention program specifically addresses the literary fiction genre. Over one school year, children read and discuss four literary fiction books. It has been shown that literary fiction, compared to other genres (e.g., nonfiction, popular fiction, comics), represents a privileged context for stimulating social cognitions in children and adolescents (Kumschick et al., 2014; Lenhart et al., 2023). Literary fiction is often characterized by complex characters, gaps and ambiguities in the story line, and outstanding language

which support immersion experiences and invite readers to actively engage with the text (Kidd & Castano, 2013; Oatley, 2016).

Second, the intervention program not only focusses on facilitating argumentation and cognitive depth in discussions, but also on positive group dynamics among children. Thus, the social learning goal of the intervention was to enhance social-emotional interaction quality in literary discussions and contribute to social inclusion in the classroom.

We considered two different strategies to promote socio-emotional interaction quality in literary discussions. First, we selected children's books that thematically focus on social inclusion and exclusion, particularly in intergroup contexts. Before discussing the books in small groups, children engaged in activities (see "Social conceptual mini lessons" in Table A1 in Appendix A for examples of activities) that aimed to enhance their understanding of social concepts and personal connections regarding the book's themes (e.g., characteristics of stereotypes). During small-group discussions, teachers prompted students to relate the stories to the classroom social life by asking connection questions (e.g., "Does this book remind you of an issue we had in our classroom?"). Consequently, we expect that high-quality discussions about the social themes of the children's books will transfer to improved social interactions in small groups and the classroom.

Secondly, based on research on the role of teacher feedback (e.g., praise, critics) for peer dynamics in the classroom (Hendrickx et al., 2017; Sette et al., 2020), teachers engaged in video-based reflections on how they can enhance their emotional support (e.g., warmth, sensitivity, autonomy support) and use talk moves (e.g., see "Teacher training on instructional interaction quality in literary discussions" in Table A1 for examples of talk moves) not only to strengthen argumentative quality but also inclusive group dynamics. Studies revealed that the quality of interactions between the teacher and individual students affects how well these students are socially included by their classmates (Endedijk et al., 2022). Teachers who are trained to model positive relationships with excluded children can positively contribute to improved peer relationships of these children in the classroom (Mikami et al., 2013). Accordingly, teachers from this study were trained to improve their attunement to the social hierarchies in the classroom (e.g., who has most control in the classroom? who is an outsider?) and reflected on strategies for subtly elevating the social standing of excluded children. For instance, this can be achieved through the talk move "extending" which invites children to listen to and build on each other's contributions. Similarly, they can use the talk move "marking" to praise a shy child for using an effective discourse element (e.g., "I liked how you referred to textual evidence to support your claim"). By encouraging peers to listen to and build on the contribution of an excluded child, a teacher can influence the peers' perception that the child's contribution is valuable.

In this study we assess socio-emotional and instructional interaction quality through the Classroom Assessment Scoring System Upper Elementary (CLASS UE; Pianta et al., 2012). CLASS explicitly addresses interaction quality between teachers and children as well as among peers. According to the CLASS, emotional interaction quality is assessed by the Emotional Support domain, which captures the extent to which teachers and children demonstrate positive relationships, teachers are sensitive and responsive towards children's social and academic needs, and allow children to take control over the classroom. Instructional interaction quality is assessed by the Instructional Support domain and addresses how effectively teachers engage the children and use high quality scaffolding to expand learning. Furthermore, Instructional Support assesses the extent to which interactions focus on deep content understanding, higher-order thinking and instructional dialogues that are characterized by cumulative talk. Since CLASS assessments are efficient and straightforward to implement, it is suitable for testing intervention effects on interaction quality in large-scale studies (Brown et al., 2010).

As research revealed that composition of groups importantly contributes to interaction quality (Murphy et al., 2017; Partee et al., 2020), we included several control variables at the group level. More

specifically, we considered group composition regarding children's gender, first language, and prosocial behavior.

2. Method

2.1. Sample and design

2.1.1. Participants

Fifty-one teachers from 41 classrooms participated in the study (18 fourth grade, 23 fifth grade). Teachers were predominantly female (86.27%) with an average age of 37.58 years ($SD = 10.31$) and an average professional experience of 14.76 years. In 31 classrooms, teachers participated alone, while in 10 classes, teachers participated in pairs. The average class size was 19.00 students ($SD = 2.86$). The sample further consisted of 752 children (48.54% girls, $M_{age} = 10.33$, $SD = 0.68$). 13.24% of the children were born abroad, and 32.16% did not speak German as the first language. Written consent for participation in the study was provided by the children and parents (rejection rate 2.97%). The study was approved by the Ethics Committee of Bern University of Teacher Education.

2.1.2. Power analysis

We conducted an a priori power analysis ($\alpha = .05$) with R (R Core Team, 2021) that tested a more complex model than the one presented in this article. In this analysis we assumed that interaction quality functions as a mediator between the intervention and individual student outcomes. Simulated two-level models with 40 classes (20 students per classroom) and with small to large intervention effects from $\beta = 0.26$ to $\beta = 0.91$ on interaction quality (Chinn et al., 2001), revealed power values between 0.79 and 0.95. Equivalent simulation models with 30 classes resulted in unsatisfactory power values. Given the more complex and restrictive nature of this mediation model, we concluded that for the hypotheses tested in this paper, 40 classrooms with three to four small groups each (approximately 160 in total) are sufficient to detect small to large intervention effects on interaction quality.

2.1.3. Control group design

The study design is based on a cluster-randomized control trial and included three measurement occasions (T0: September 2022, T1: January 2023, T2: June 2023). Classrooms were randomly assigned to intervention or control groups, stratified by grade level to ensure balanced representation across grades four and five, and by the percentage of non-Swiss residents in the community to balance the representation of non-Swiss and Swiss populations. Teachers in the wait-list control group received the professional training one year after the intervention.

For sample recruitment, schools across various German-speaking cantons of Switzerland were contacted. Classrooms that did not correspond to the target grade or did not agree to randomization were not considered further. The study started with 42 classrooms. One classroom from the control group withdrew at the T0 assessment. All remaining 41 classes stayed in the study. Despite the loss of one class, intervention condition (intervention vs. control) was unrelated to the percentage of non-Swiss children, $\chi^2(1) = 1.58$, $p = .21$, or children's first language (German vs. non-German), $\chi^2(1) = 0.03$, $p = .86$. No age difference between the control and intervention groups was found, $t(724) = 0.71$, $p = .48$. Moreover, teachers participating in pairs vs. alone were equally distributed across intervention and control group, $\chi^2(1) = 1.92$, $p = .17$.

2.1.4. Small groups

Overall, the sample consisted of 159 small groups with an average size of 4.53 children ($SD = 0.83$, range = 2–7). Each class included three to four small groups (five classes with three and 36 classes with four groups) and teachers were responsible for one to four discussion groups ($M = 3.45$, $SD = 0.86$). The groups remained the same across all three measurement occasions. Teachers were reminded of the group compositions

before each assessment and compliance was monitored by test administrators on-site. The small groups were formed by the teachers before the first measurement occasion, with instructions to create heterogeneous groups in terms of gender, migration background, and academic achievement. We computed the Simpson-Index regarding gender and first language to estimate the level of diversity in the small groups. For variables including two conditions, the Simpson index ranges from 0 to 0.5, whereby 0 represents full group homogeneity (e.g., only girls in the group) and 0.5 maximal diversity (equal number of boys and girls in the group). The small group diversity indexes were close to the classroom diversity indexes, indicating that diversity in the groups reflected classroom diversity (Gender: $SI_{group} = .46$ vs $SI_{class} = 0.47$; First language: $SI_{group} = .31$ vs $SI_{class} = 0.37$). The study design is summarized in Fig. 1.

2.2. Procedure and measures

2.2.1. Text-based discussions for assessment of interaction quality

At the three measurement occasions, children and their teacher discussed one of three stories with a similar moral dilemma (story types A, B, and C). All stories focused on a conflict between a moral norm (e.g., keeping a promise or not lying) and loyalty towards friends (e.g., not disappointing one's own friends). The three texts had a similar length ($M = 598$ words) and a comparable difficulty index (A: 26.8, B: 27.16, C: 26.92) which was calculated by the text analysis tool LIX (Lesbarkeit-index; Lenhard & Lenhard, 2011) based on indicators such as length of words and sentences. We systematically varied the story order, resulting in six story orders that were equally distributed across control and intervention conditions. No significant effects of story order were found on interaction quality. We included story type as a control variable in the analyses.

Group discussions took place in a room separate from the regular classroom. The stories were played to children on tablets. Additionally, they received the printed text of the stories. This dual approach supports comprehension in situations where students must understand under time constraints (Clinton-Lisell, 2023). Moreover, the printed text helped children reference story details during the discussion. The three stories ended with a dilemma question ("What should x do?"). The teachers were instructed to conduct the discussions with the children as they typically discuss texts with the children and to discuss for approximately 15 min. On average, the groups discussed for 14.98 min ($SD = 2.19$, range = 6–24 min). Small-group discussions were videotaped at all three measurement occasions. Children who participated in the discussions but could not be filmed were positioned behind the camera so that they were not visible ($n = 147$).

2.2.2. Interaction quality

We assessed interaction quality by the CLASS UE. The CLASS covers 12 dimensions, which are rated on a seven-point Likert scale (1 = low, 7 = high). Five psychology students attended a two-day observation training and subsequently certified through an online reliability test. This requires that the ratings for five videos match the master code with ± 1 point in over 80%. Before the raters independently coded, they received a half-day training on applying the CLASS dimensions to the specific context of text-based small-group discussions. This included explanations on how the various indicators and behavioral markers of CLASS manifest in this specific instructional context. The raters were also instructed to omit behavioral markers or indicators that could not show variance due to the standardization of the instructional context (e.g., physical proximity).

The 12 CLASS dimensions are organized by three broad interaction domains: Emotional Support, Classroom Organization, and Instructional Support. There were few missing in the CLASS scores (1.68% of a total of 477 small-group discussions), and they occurred because of a different teacher leading the discussions ($n = 2$), teacher absence ($n = 4$) and video malfunctions ($n = 2$).

Socio-emotional interaction quality was assessed using the three

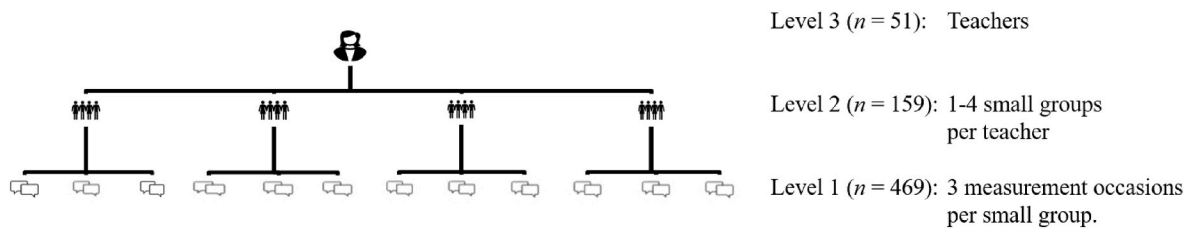


Fig. 1. Overview of the multilevel design of the study.

dimensions of Emotional Support: (1) Positive Climate (e.g., quality of relationships and affect, respect), (2) Teacher Sensitivity (e.g., awareness and responsiveness to students' needs, student comfort), (3) Regard for Students Perspectives (e.g., providing opportunities for autonomy, consideration of students' ideas). The Emotional Support dimensions were averaged, and Cronbach's alpha was high (0.80). We also considered students' engagement in discussions as an element of socio-emotional interaction quality, which is assessed by a single dimension and captures the degree to which students are focused actively engaged in the learning activity (e.g., active listening, asking questions, responding, sharing ideas, volunteering). Like the other dimensions, Student Engagement was assessed with a single rating on a seven-point Likert scale. Compared to the other dimensions of the CLASS, Student Engagement focuses exclusively on student (not teacher) behavior and therefore does not pertain to one of the three broad domains.

Instructional interaction quality was assessed by the five dimensions of the Instructional Support domain: (1) Instructional Learning Formats (e.g., active facilitation), (2) Content Understanding (e.g., discussions supporting deep and integrated understanding of facts, concepts, and principles), (3) Analysis and Inquiry (e.g., open-ended challenging questions, higher-order thinking), (4) Quality of Feedback (e.g., scaffolding), and (5) Instructional Dialogue (e.g., cumulative content-driven exchanges). The Instructional Support variable was created by averaging the five ratings (Cronbach's alpha = 0.93).

In this study, we excluded the three Classroom Organization dimensions (Behavioral Management, Productivity, Negative Climate) because these dimensions are unrelated to our study hypotheses. Additionally, Classroom Organization showed strong ceiling effects ($M = 6.90$, $SD = 0.35$). Negative Climate (within the Classroom Organization domain) differs from Positive Climate (within the Emotional Support domain) not only in value but also in content: Negative Climate indicates the presence of negative behaviors rather than the absence of positive behaviors. Confirmatory factor analyses in various large-scale studies in upper elementary and secondary grade classrooms supported a three-factor structure in which the Negative Climate dimension belongs to the Classroom Organization domain (Pianta et al., 2012). Conceptually, Negative Climate can be understood as an outcome of ineffective classroom organization, as poor behavior management or unproductive classrooms increase the risk of student misbehavior and negative affect among both teachers and students.

2.2.3. Group-level characteristics

At the group level, we included variables on first language, gender distribution, and prosocial behavior. All measures were assessed by tasks on tablets (see Table A2 in Appendix A for an overview of all the study instruments). First language was assessed by asking children to indicate the language they speak most at home. Prosocial behavior was assessed based on a peer-nomination procedure that requires children to nominate peers that show prosocial behaviors. This procedure was implemented at T0 and T2 and included two items ("Who in your class behaves fairly and is friendly to other children", "Who in your class helps and shares a lot with other children"). The nominations were counted for each child and divided by the number of potential nominators. We thus had four indicators of prosocial behavior (two at T0 and two at T2). Since an exploratory factor analysis with the four indicators yielded only

one factor with loadings between 0.85 and 0.88, we used the factor score from this analysis as a measure for children's individual prosocial behavior. To form the group variables, the variables gender, first language, and prosocial behavior were aggregated across the children of a small group.

2.3. Implementation of the intervention

The intervention contained three elements: (1) Onsite training, (2) implementation of the intervention in the classrooms over a school year, and (3) video-based coaching sessions during the implementation. The design of the intervention is illustrated in Fig. 2 (see also Table A1 for a more detailed overview of the intervention; for sample materials see [Soziale und sprachliche Kompetenzen über Kinderliteratur fördern | PHBern](#)).

2.3.1. Onsite training

The training totaled 10 h. One training focused on instructional interaction quality, while the second addressed the social learning goals of the program. In the first training, teachers were introduced to QT and reflected on videotaped small-group discussions based on effective characteristics of discussion quality. In the second training, teachers were introduced to the social themes of the children's books (i.e., intergroup exclusion, group dynamics, and civil courage). Furthermore, they reflected on social inclusion and exclusion processes in their classroom and how they can use the QT talk moves to positively influence group dynamics in their small groups and classroom.

2.3.2. The school-based intervention

In the first weeks of the intervention, children received explicit instructions on effective elements of productive text-based discussions through QT mini-lessons (see "QT mini lessons" in Table A1 for a more detailed description). They were also introduced to discussion rules that set the normative framework for socio-emotional and instructional interactions during the discussions (e.g., "We question and argue about ideas not people"). The children next read the books in two to three parts (see "Book lessons" in Table A1 for a more detailed description). The book reading included activities before, during, and after the discussions in small groups. Before the discussions, the children engaged in social conceptual lessons to introduce them to the social themes of the children's books and to establish personal connections to the story theme (e.g., discussions about the bystander role in social exclusion). The children then read a part of the book and completed tasks on basic text comprehension in their literacy journals. In preparation for the discussion, the children set discussion goals (e.g., "I want to better listen to others," "I would like to practice disagreement") and identified authentic questions about the text in their literacy journals (e.g., "What would you do in this situation if you were the protagonist?"). Then the teachers discussed with the children in small groups that were equally composed as the groups of the data assessments. The small-group discussions lasted for 15 min. The teacher stayed with the small group for the whole discussion. While the teacher discussed with one group, the rest of the class often engaged in reading or silent work in the literacy journal. After discussing, the teacher reflected with the children on the discussion. All elements of the intervention were described in manuals

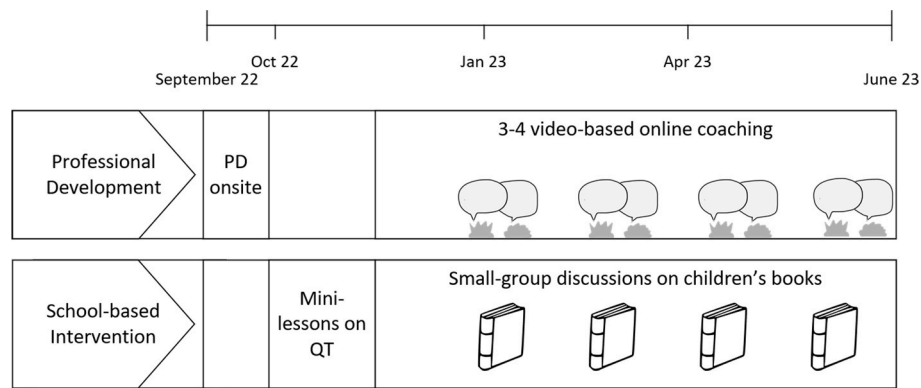


Fig. 2. Overview of the intervention design.

and lesson plans for the teachers.

The teachers documented the number of invested minutes and adherence to lesson plans after each unit. On average, they invested 33.03 h ($SD = 6.27$, range = 12.58 to 42.00), with 22.97% invested for the mini-lessons, 17.47% for the social conceptual lessons, and 59.56% for the reading and discussion activities. The teachers indicated on a four-point scale (0 = never, 3 = frequently) that they rarely deviated from the lesson plans ($M = 1.05$, $SD = 0.19$). On average, the teachers conducted 9.11 discussions with each of their small groups over the school year. Thirteen of the classes read and discussed three of children's books while seven read and discussed all four of the children's books. One teacher implemented less than 30% of the intervention (mini-lessons and one children's book, without participating in coaching). However, in line with the Intention-to-Treat approach, we decided not to exclude this teacher from the analyses.

2.3.3. Coaching

The coaching sessions began after the mini-lessons when the children discussed the first book in small groups. On average, the teachers attended 3.11 coaching sessions ($SD = 0.79$, range: 0–4). All coaching was conducted online and included reflection on videorecorded small-group discussions. Reflections were guided by a short observation guide, which contains descriptions and examples of indicators of instructional and socio-emotional interaction quality. Coaches rated teachers' motivation to develop their facilitation strategies after each coaching session on a five-point scale (0 = very low, 4 = very high). Their motivation was at a high level ($M = 3.26$, $SD = 0.63$) without significant changes over time. Additionally, after each session, coaches rated the quality of the videotaped discussions based on four items that focused on cognitive and social interaction quality. A multilevel model analysis revealed that coach-rated discussion quality improved over time, $F(1, 22.85) = 71.51$, $p < .001$, indicating successful implementation quality.

2.4. Statistical analyses

For the analyses of intervention effects, we used three-level multilevel models, where measurement occasions (time) at level 1 were nested within groups at level 2, and groups were nested within teachers at level 3. To select the model structure for each outcome, we adopted a stepwise approach by incrementally adding random effects at level 2 (group) and level 3 (teacher), testing for additional explained variance. The variance-covariance structure for the residuals was selected by comparing models with less restrictive variance-covariance structures (Peugh & Heck, 2017; West et al., 2022). We selected and applied a separate model for each outcome of interaction quality. Models were estimated in R (R Core Team, 2021) using restricted maximum likelihood (REML).

Fixed effects were tested by Wald test p -values, and random effects

were tested by likelihood ratio tests ($-2\Delta LL$) comparing nested models (Hoffman, 2015). To test for significance of slopes per intervention condition, we conducted follow-up simple slope analyses (Bauer & Curran, 2005). A significance level threshold of $p < .05$ was applied for all analyses. Cohen's d effect sizes for intervention effects were calculated using fixed effects (time*intervention condition) and baseline sample standard deviations (Feingold, 2009).

Maximum likelihood estimation in longitudinal multilevel modeling enables the analysis of all observed data on dependent variables (Grund et al., 2019). Consequently, groups with partially missing data on dependent variables were included in the analysis.

We added covariates at all three levels: story type and group size at level 1 (time), first language, prosocial behavior, and gender distribution at level 2 (group), and teacher experience at level 3 (teacher). For clarity, first language, group size, and teacher experience were grand mean centered, gender distribution was centered to a mean of 0.5, and prosocial behavior was z-standardized. Correlations amongst covariates and between covariates and outcomes of interaction quality were examined at each level.

3. Results

3.1. Preliminary analysis

We used unconditional models to assess the distribution of variance across levels. For Emotional Support, 53.4% of the variance occurred across time within groups (level 1), 11.2% across groups within teachers (level 2), and 35.4% between teachers (level 3). For Instructional Support, 52.0% of the total variance was found across time within groups (level 1), 3.2% across groups within teachers (level 2), and 44.6% between teachers (level 3). For Student Engagement, 53.5% of total variance occurred across time within groups (level 1), 11.4% across groups within teachers (level 2), and 35.1% between teachers (level 3).

The model selection process for Emotional Support and Student Engagement indicated that models with random intercepts at level 2 (group) and level 3 (teacher), and a random slope at level 3 (teacher) fit best. For both Emotional Support and Student Engagement, alternative variance-covariance structures for the residuals did not improve model fit, thus no alternative variance-covariance structures for the residuals were chosen. For Instructional Support, a model with random intercepts at level 2 (group) and level 3 (teacher), a random slope at level 3 (teacher), and a diagonal variance-covariance structure for the residuals, allowing for differing residual variances across occasions showed the best model fit. Refer to Appendix B for the equations of the final models and to Tables C1-C3 (Appendix C) for additional information on the model selection and a detailed overview of the final models. See Table 1a and Table 1b for means and correlations between variables per level of analysis.

Table 1a
Means, Standard Deviations (in Parentheses), and Correlations for Variables at Level 1 (Time)

Variable	M (SD)	1	2	3	4	5	6	7	8	9	10
Level 1 (time)											
1 Emotional Support T0	4.94 (0.65)	–									
2 Emotional Support T1	5.28 (0.85)	0.54***	–								
3 Emotional Support T2	5.49 (0.90)	0.44***	0.66***	–							
4 Student Engagement T0	5.02 (0.84)	0.48***	0.32***	0.39***	–						
5 Student Engagement T1	5.19 (0.91)	0.37***	0.59***	0.51***	0.47***	–					
6 Student Engagement T2	5.41 (0.89)	0.23**	0.37***	0.63***	0.43***	0.61***	–				
7 Instructional Support T0	4.00 (0.64)	0.56***	0.35***	0.44***	0.72***	0.42***	0.42***	–			
8 Instructional Support T1	4.45 (0.83)	0.41***	0.59***	0.58***	0.44***	0.78***	0.57***	0.54***	–		
9 Instructional Support T2	4.67 (0.97)	0.32***	0.48***	0.72***	0.44***	0.60***	0.78***	0.52***	0.75***	–	
10 Group size	4.53 (0.83)	0.01	0.04	0.14 [†]	0.10	0.07	0.02	0.23**	0.10	0.09	–

Note. Correlations between group size and outcomes of interaction quality are reported per measurement occasion.
*** $p < 0.001$. ** $p < 0.01$. [†] $p < .10$.

Table 1b
Means, Standard Deviations (in Parentheses), and Correlations for Variables at Level 2 (Group) and Level 3 (Teacher)

Variable	M (SD)	1	2	3	4	5	6
Level 2 (group)							
1 First language	0.32 (0.25)	–					
2 Prosocial behavior	0.01 (0.63)	–0.24**	–				
3 Gender distribution	0.51 (0.14)	0.06	0.06	–			
Level 3 (teacher)							
4 Grade	0.55				–		
5 Teacher experience (in years)	15.39 (12.13)				–0.38**	–	
6 Intervention condition	0.55				–0.11	0.01	–

Note. First language (German = 0, non-German = 1), Gender (girls = 0, boys = 1), Grade (4th = 0, 5th = 1), Intervention condition (control = 0, intervention = 1).
** $p < .01$.

3.2. Effects on interaction quality

The final models for the effects on interaction quality can be derived from Table 2. The intercepts indicate the predicted outcome of

interaction quality values at baseline for an averagely sized fourth grade group with a balanced gender distribution, an average proportion of students with German as their first language, and a teacher with average experience facilitating the discussion.

Table 2
Estimates (and Standard Errors) for Models Predicting Emotional Support, Instructional Support and Student Engagement

	Emotional Support		Student Engagement		Instructional Support	
	Est.		Est.		Est.	
Fixed effects						
Intercept	4.90***	(0.21)	5.07***	(0.24)	4.14***	(0.20)
Level 1 (time)						
Time	0.08	(0.05)	0.06	(0.06)	0.10 [†]	(0.06)
Story type A vs. B	–0.00	(0.06)	0.03	(0.07)	–0.01	(0.04)
Story type A vs. C	–0.07	(0.06)	–0.13*	(0.07)	–0.11*	(0.04)
Group size	0.06	(0.05)	0.05	(0.05)	0.06	(0.04)
Level 2 (group)						
First language	–0.11	(0.19)	–0.16	(0.21)	–0.26	(0.16)
Prosocial behavior	–0.02	(0.06)	0.13*	(0.06)	0.09 [†]	(0.05)
Gender distribution	0.10	(0.31)	0.07	(0.34)	–0.18	(0.25)
Level 3 (teacher)						
Grade	–0.13	(0.15)	–0.18	(0.18)	–0.11	(0.17)
Teacher experience	–0.02**	(0.01)	–0.01*	(0.01)	–0.01	(0.01)
Intervention condition	0.15	(0.15)	0.02	(0.18)	0.03	(0.16)
Cross-level effects						
Time intervention condition	0.40***	(0.07)	0.28**	(0.08)	0.50***	(0.08)
Random effects						
Level 1 (time): residuals	0.23	(0.48)	0.33	(0.57)	0.06	(0.25)
Level 2 (group): intercept	0.13	(0.36)	0.12	(0.35)	0.08	(0.28)
Level 3 (teacher): intercept	0.15	(0.39)	0.24	(0.48)	0.26	(0.51)
Level 3 (teacher): covariance	–0.18		–0.20		–0.30	
Level 3 (teacher): slope	0.03	(0.16)	0.04	(0.19)	0.05	(0.23)
Model summary						
–2LL	915.49		1047.02		803.61	
Marginal R ²	0.27		0.15		0.36	
Conditional R ²	0.68		0.61		0.90	

Note. Story type (A = 0, B = 1, C = 1), First language (German = 0, non-German = 1), Gender (girls = 0, boys = 1), Grade (4th = 0, 5th = 1), Intervention condition (control = 0, intervention = 1). No p-values are reported for random effects (standard deviations in parentheses). *** $p < 0.001$. ** $p < 0.01$. * $p < 0.05$. [†] $p < .10$.

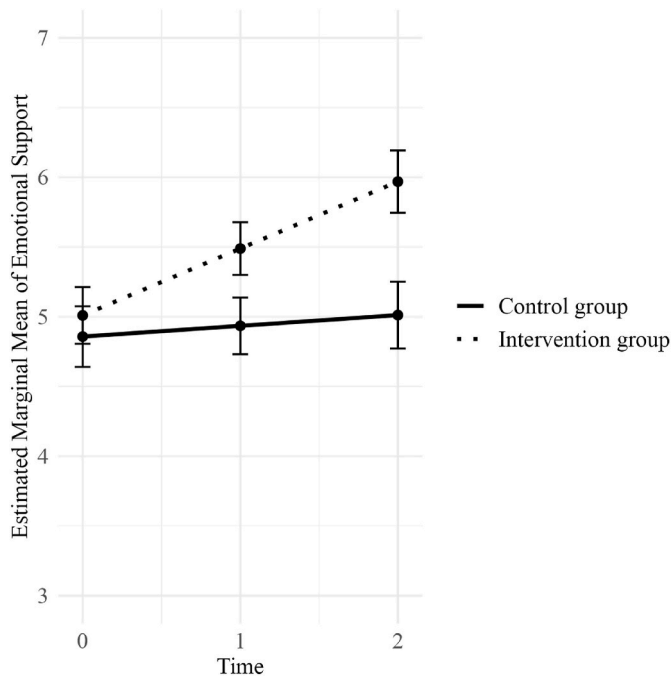


Fig. 3. Slopes of time on estimated marginal means of Emotional Support.

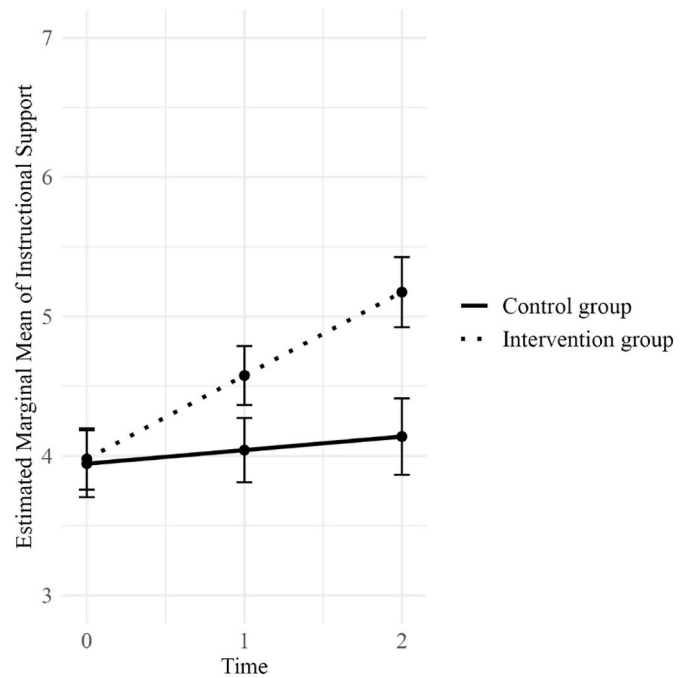


Fig. 5. Slopes of time on estimated marginal means of Instructional Support.

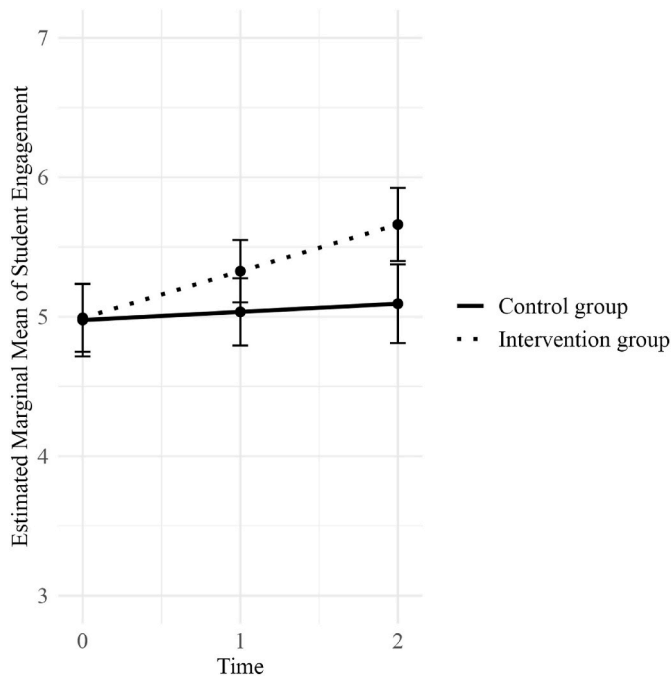


Fig. 4. Slopes of time on estimated marginal means of Student Engagement.

3.2.1. Socio-emotional interaction quality

An intervention effect was found for Emotional Support, $\beta = 0.40$, $SE = 0.07$, $p < .001$, $d = 1.24$ (see Fig. 3). Follow-up simple slope analyses revealed a positive slope of time on Emotional Support for the intervention condition, $\beta = 0.48$, $SE = 0.05$, $p < .001$. The slope for the control condition did not differ from zero, $\beta = 0.08$, $SE = 0.05$, $p = .143$. At Level 3 (teacher), groups led by teachers with more experience showed significantly lower Emotional Support, $\beta = -0.02$, $SE = 0.01$, $p = .010$.

Moreover, an intervention effect was found for Student Engagement, $\beta = 0.28$, $SE = 0.08$, $p = .001$, $d = 0.66$ (see Fig. 4). A positive slope of

time on Student Engagement was found for the intervention group, $\beta = 0.34$, $SE = 0.06$, $p < .001$, while the slope for the control condition does not indicate change over time, $\beta = 0.06$, $SE = 0.06$, $p = .346$. At Level 1 (time), groups discussing story type C showed significantly lower Student Engagement than groups discussing story type A, $\beta = -0.13$, $SE = 0.07$, $p = .046$. At Level 2 (groups), a significant positive effect of prosocial behavior on Student Engagement was found, $\beta = 0.13$, $SE = 0.06$, $p = .046$. At Level 3 (teacher), groups led by teachers with more experience showed significantly lower Student Engagement, $\beta = -0.01$, $SE = 0.01$, $p = .042$.

3.2.2. Instructional interaction quality

An intervention effect was found for Instructional Support, $\beta = 0.50$, $SE = 0.08$, $p < .001$, $d = 1.57$ (see Fig. 5). Simple slope analyses showed that the slope of time on Instructional Support for the intervention condition was positive, $\beta = 0.60$, $SE = 0.05$, $p < .001$, while the slope for the control condition did not deviate from zero, $\beta = 0.10$, $SE = 0.06$, $p = .086$. At Level 1 (time), groups discussing story type C showed significantly lower Instructional Support than groups discussing story type A, $\beta = -0.11$, $SE = 0.04$, $p = .018$. At Level 2 (groups), a marginally significant positive effect of prosocial behavior on Instructional Support was found, $\beta = 0.09$, $SE = 0.05$, $p = .072$.

4. Discussion

This study investigated the effects of a literary intervention on observed interaction quality in small group literary discussions. The main finding relates to the positive impact of the intervention on both socio-emotional and instructional interaction quality in discussions about a moral dilemma. The effect sizes were moderate for student engagement and large for emotional and instructional support. Extensive longitudinal studies based on the CLASS instrument demonstrate that such interactions longitudinally predict social, self-regulative, and cognitive competences in children and adolescents (e.g., Gasser et al., 2018; Mashburn et al., 2008; Rimm-Kaufman et al., 2015).

A distinctive feature of this literary intervention is the integration of cognitive and social learning objectives. Prevention research on social-emotional learning (SEL) clearly shows that SEL programs not only stimulate socio-emotional competencies but also contribute to positive

academic development (Cipriano et al., 2023). The ability of children and adolescents to build positive relationships with peers and teachers, or to interact cooperatively and fairly with others, is essential for successful academic participation and learning (Sette et al., 2020). Therefore, teaching practices that naturally take up opportunities in subject matters to simultaneously stimulate academic and social learning are of particular interest (Jones et al., 2011).

This study demonstrates that literary education presents a promising interaction context in both the socioemotional and instructional domain. Various intervention studies confirm that literary education effectively contributes to socio-cognitive and linguistic skills (Jones et al., 2011; Kumschick et al., 2014; Schrijvers et al., 2019). For example, the *Reading and Feeling* program uses literary fiction to stimulate social imagination in children (Kumschick et al., 2014). Over two months, children participate in creative activities such as staged readings, role play or rhyming games to enhance children's identification with characters and provide them an embodied experience of the characters' emotions. Gasser et al. (2022) characterize such literary programs as expressive, supporting immersion experiences, allowing children to explore the characters' psychology and make spontaneous connections between the text and their lives.

The current intervention differs from an expressive approach by focusing on the quality of interactions during literary discussions. Based on a critical-analytic perspective, we characterize the quality of literary interaction by the extent to which children (a) engage argumentatively with the text and the ideas of peers and (b) interact cooperatively, respectfully, and inclusively with each other and the teacher. While the effectiveness of text-based discussion approaches on instructional interaction quality is well documented (Murphy et al., 2009), studies with randomized control group designs, larger samples, and multiple measurement points remain rare (for exceptions see Lawrence et al., 2015; Lin et al., 2022; Wilkinson et al., 2023). The current study adds to this literature by focusing specifically on discussions about literary fiction and additionally including assessments of socio-emotional interaction quality in literary discussions.

Furthermore, this study is the first to systematically differentiate between the levels of teacher and small group. We understand small groups as social microcosms that provide children a safe space to practice their communicative and inclusive skills (Chen et al., 2023). At the group level, prosocial behavior among groups was associated with enhanced student engagement and marginally associated with instructional interaction quality. According to Janssen (2013), group-level effects are notoriously under-studied due to limitations in sample size of most studies on collaborative learning. Small groups with high prosocial behavior might exhibit higher engagement and instructional interaction quality because the children in these groups might be more likely to interact with each other positively and fairly (e.g., listen to each other, build on the ideas of others, and respectfully manage disagreements). Interestingly, other group characteristics did not relate to interaction quality. Thus, it seems that successful discussions are more likely to depend on the groups' capacity for positive interactions than socio-demographic characteristics.

The study also showed an effect of the text (story type) on student engagement and instructional interaction quality. Although the dilemma situations were very similar in content and form, this finding shows that the text is an important part of the literary interaction. Discussions between parents and children about narrative fictional and complex texts compared to nonfictional and simple texts are characterized by higher linguistic and socio-cognitive richness, extra-textual talk, and elaboration (Muhinyi et al., 2020; Nyhout & O'Neill, 2013). It is therefore important that future research investigates the effects of different genres and text characteristics on interaction quality in literary discussions.

Finally, teacher experience negatively predicted the quality of socio-emotional interactions. Previous research on the CLASS has revealed mixed findings regarding the relationship between teaching experience

and teaching quality, with some studies showing no associations (Graham et al., 2020; Sabol et al., 2013) and others reporting a decline in teaching quality (e.g., Pianta et al., 2007). For example, Partee (2020) found that high levels of disruptive behavior in the classroom were negatively associated with the observed emotional support provided by teachers, but this was only evident for experienced teachers, not for those with less experience. One explanation for our finding might be that emotional exhaustion and a controlling teaching style increase with years of teaching due to the profession's high demands (e.g., student misbehavior, Hellebaut et al., 2023). While a controlling teaching style might positively contribute to classroom organization, it could conflict with the quality of emotional support, particularly with regard for students' perspectives. However, as our study did not assess teachers' emotions or motivations, this explanation remains highly speculative.

The study is not without limitations. Firstly, it is unclear to what extent the CLASS is sensitive enough to capture specific interaction characteristics of text-based discussions, such as the quality of questioning and argumentation. An alternative to the CLASS is the Argumentation Rating Tool (ART; Reznitskaya & Wilkinson, 2021). Compared to CLASS, ART allows for a more specific observation of teacher facilitation and student argumentation during text-based discussions. However, the ART does not consider socio-emotional interaction quality. It is therefore important that future developments of observation systems on text-based interaction quality also take socio-emotional dimensions into account. Secondly, coding via CLASS may have led to an underestimation of group dynamics. Even though a substantial portion of variance in socio-emotional interaction quality referred to the group level, the group level played a minor role for instructional interaction quality. Moreover, the random slope of time was only relevant at the teacher level not the group level. Although the CLASS framework conceptually covers every classroom interaction, including interactions among peers, the instrument is typically used to measure teaching performance. Other observation or coding systems might be better suited to analyze the complex interplay between teachers and small groups. Thirdly, student engagement was assessed with a single rating on one dimension. Observation scales with multiple items generally have superior psychometric properties compared to a one-item scale.

Despite these limitations, the present study significantly extends previous research on literary and text-based discussion approaches in terms of study design and a comprehensive analysis of interaction quality. Particularly, it adds to previous research that literary education represents a promising developmental context to integrate academic and socio-emotional learning.

Funding resource

This work was supported by the Swiss National Science Foundation [100019_200841].

CRedit authorship contribution statement

Luciano Gasser: Writing – original draft, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **David Preisig:** Writing – review & editing, Writing – original draft, Formal analysis. **Anna Frei:** Writing – review & editing, Investigation. **Yvonne Dammert:** Writing – review & editing, Supervision, Conceptualization, Project administration, Resources, Methodology. **Sara Egger:** Writing – review & editing, Project administration, Data curation. **P. Karen Murphy:** Writing – review & editing.

Declaration of competing interest

None.

Acknowledgments

The authors would like to thank Stiftung Mercator Schweiz [2017-0026] and éducation21 [17.R.066] for funding the development of the intervention described in this study. We also would like to thank the undergraduate students for their support in coding videos, as well as the teachers and their students for participating in the study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.learninstruc.2024.101996>.

References

- Adrián, J. E., Clemente, R. A., & Villanueva, L. (2007). Mothers' use of cognitive state verbs in picture-book reading and the development of children's understanding of mind: A longitudinal study. *Child Development, 78*(4), 1052–1067. <https://doi.org/10.1111/j.1467-8624.2007.01052.x>
- Allen, J., Hauser, S. T., Bell, K. L., & O'Connor, T. G. (1994). Longitudinal assessment of autonomy and relatedness in adolescent-family interactions as predictors of adolescent ego development and self-esteem. *Child Development, 65*(1), 179–194. <https://doi.org/10.2307/1131374>
- Allen, J., 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>
- Aram, D., Fine, Y., & Ziv, M. (2013). Enhancing parent-child shared book reading interactions: Promoting references to the book's plot and socio-cognitive themes. *Early Childhood Research Quarterly, 28*(1), 111–122. <https://doi.org/10.1016/j.ecresq.2012.03.005>
- Baker, A. R., Lin, T.-J., Chen, J., Paul, N., Anderson, R. C., & Nguyen-Jahiel, K. (2017). Effects of teacher framing on student engagement during collaborative reasoning discussions. *Contemporary Educational Psychology, 51*, 253–266. <https://doi.org/10.1016/j.cedpsych.2017.08.007>
- Baucal, A., Josić, S., Stepanović Ilić, I., Videnović, M., Ivanović, J., & Krstić, K. (2023). What makes peer collaborative problem solving productive or unproductive: A qualitative systematic review. *Educational Research Review, 41*, Article 100567. <https://doi.org/10.1016/j.edurev.2023.100567>
- Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research, 40*(3), 373–400. https://doi.org/10.1207/s15327906mbr4003_5
- Blewitt, P., & Langan, R. (2016). Learning words during shared book reading: The role of extratextual talk designed to increase child engagement. *Journal of Experimental Child Psychology, 150*, 404–410. <https://doi.org/10.1016/j.jecp.2016.06.009>
- Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In R. M. Lerner, & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (5th ed., pp. 993–1028). John Wiley & Sons Inc. Vol. 1.
- Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4rs program. *Journal of Educational Psychology, 102*(1), 153–167. <https://doi.org/10.1037/a0018160>
- Bus, A. G., Belsky, J., van Ijzendoorn, M. H., & Crnic, K. (1997). Attachment and bookreading patterns: A study of mothers, fathers, and their toddlers. *Early Childhood Research Quarterly, 12*(1), 81–98. [https://doi.org/10.1016/S0885-2006\(97\)90044-2](https://doi.org/10.1016/S0885-2006(97)90044-2)
- Chen, J., Lin, T.-J., Wilkinson, I. A. G., Ha, S. Y., & Paul, N. (2023). Linkages between cognitive and social dialogue patterns during collaborative small-group discussions. *Learning and Instruction, 87*, Article 101795. <https://doi.org/10.1016/j.learninstruc.2023.101795>
- 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>
- Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development, 94*(5), 1181–1204. <https://doi.org/10.1111/cdev.13968>
- Clingenpeel, B. T., & Pianta, R. C. (2007a). Mothers' sensitivity and book-reading interactions with first graders. *Early Education & Development, 18*(1), 1–22. <https://doi.org/10.1080/10409280701274345>
- Clingenpeel, B. T., & Pianta, R. C. (2007b). Mothers' sensitivity and book-reading interactions with first graders. *Early Education & Development, 18*(1), 1–22. <https://doi.org/10.1080/10409280701274345>
- Clinton-Lisell, V. (2023). Does reading while listening to text improve comprehension compared to reading only? A systematic review and meta-analysis. *Educational Research: Theory and Practice, 34*(3), 133–155.
- Dowdall, N., Melendez-Torres, G. J., Murray, L., Gardner, F., Hartford, L., & Cooper, P. J. (2020). Shared picture book reading interventions for child language development: A systematic review and meta-analysis. *Child Development, 91*(2), 383–399. <https://doi.org/10.1111/cdev.13225>
- Ece Demir-Lira, Ö., Applebaum, L. R., Goldin-Meadow, S., & Levine, S. C. (2019). Parents' early book reading to children: Relation to children's later language and literacy outcomes controlling for other parent language input. *Developmental Science, 22*(3), Article e12764. <https://doi.org/10.1111/desc.12764>
- Endedijk, H. M., Breeman, L. D., van Lissa, C. J., Hendrickx, M. M. H. G., den Boer, L., & Mainhard, T. (2022). The teacher's invisible hand: A meta-analysis of the relevance of teacher-student relationship quality for peer relationships and contribution of student behavior. *Review of Educational Research, 92*(3), 370–412. <https://doi.org/10.3102/00346543211051428>
- Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science, 11*(6), 481–486. <https://doi.org/10.1111/1467-9280.0029>
- Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods, 14*(1), 43–53. <https://doi.org/10.1037/a0014699>
- Fitton, L., McIlraith, A. L., & Wood, C. L. (2018). Shared book reading interventions with English learners: A meta-analysis. *Review of Educational Research, 88*(5), 712–751. <https://doi.org/10.3102/0034654318790909>
- 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>
- Gasser, L., Grütter, J., Buholzer, A., & Wettstein, A. (2018). Emotionally supportive classroom interactions and students' perceptions of their teachers as caring and just. *Learning and Instruction, 54*, 82–92. <https://doi.org/10.1016/j.learninstruc.2017.08.003>
- Graham, L. J., White, S. L. J., Cologon, K., & Pianta, R. C. (2020). Do teachers' years of experience make a difference in the quality of teaching? *Teaching and Teacher Education, 96*, Article 103190. <https://doi.org/10.1016/j.tate.2020.103190>
- Grund, S., Lüdtke, O., & Robitzsch, A. (2019). Missing data in multilevel research. In *The handbook of multilevel theory, measurement, and analysis* (pp. 365–386). American Psychological Association. <https://doi.org/10.1037/0000115-017>
- Hellebaut, S., Haerens, L., Vanderlinde, R., & Katrien, De C. (2023). Burnout, motivation, and (de-)motivating teaching style in different phases of a teaching career. *Teaching and Teacher Education, 129*, Article 104168. <https://doi.org/10.1016/j.tate.2023.104168>
- Hendrickx, M. M. H. G., Mainhard, T., Oudman, S., Boor-Klip, H. J., & Brekelmans, M. (2017). Teacher behavior and peer liking and disliking: The teacher as a social referent for peer status. *Journal of Educational Psychology, 109*(4), 546–558. <https://doi.org/10.1037/edu0000157>
- Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge.
- Jadallah, M., Anderson, R. C., Nguyen-Jahiel, K., Miller, B. W., Kim, I.-H., Kuo, L.-J., Dong, T., & Wu, X. (2011). Influence of at teacher's scaffolding moves during child-led small-group discussions. *American Educational Research Journal, 48*(1), 194–230. <https://doi.org/10.3102/0002831210371498>
- Janssen, J., Chinn, C. A., Chan, C., & O'Donnell, A. M. (2013). Multilevel analysis for the analysis of collaborative learning. In C. E. Hmelo-Silver (Ed.), *The international handbook of collaborative learning* (pp. 112–125). Routledge.
- Jones, S. M., Brown, J. L., & Aber, J. L. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development, 82*(2), 533–554. <https://doi.org/10.1111/j.1467-8624.2010.01560.x>
- 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>
- Kraatz, E., Nagpal, M., Lin, T.-J., Hsieh, M.-Y., Ha, S. Y., Kim, S., & Shin, S. (2020). Teacher scaffolding of social and intellectual collaboration in small groups: A comparative case study. *Frontiers in Psychology, 11*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.587058>
- Kumshick, I. R., Beck, L., Eid, M., Witte, G., Klann-Delius, G., Heuser, I., Steinlein, R., & Menninghaus, W. (2014). READING and FEELING: The effects of a literature-based intervention designed to increase emotional competence in second and third graders. *Frontiers in Psychology, 5*. <https://doi.org/10.3389/fpsyg.2014.01448>
- Landry, S., Smith, K., Swank, P., Zucker, T., Crawford, A., & Solari, E. (2011). The effects of a responsive parenting intervention on parent-child interactions during shared book reading. *Developmental Psychology, 48*, 969–986. <https://doi.org/10.1037/a0026400>
- Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word Generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal, 52*(4), 750–786.
- Lenhard, W., & Lenhard, A. (2011). Berechnung des Lesbarkeitsindex LIX nach Björnson. Unpublished <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), 1–19.
- 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., Anderson, R. C., Hummel, J. E., Jadallah, M., Miller, B. W., Nguyen-Jahiel, K., Morris, J. A., Kuo, L.-J., Kim, I.-H., Wu, X., & Dong, T. (2012). Children's use of analogy during Collaborative Reasoning. *Child Development, 83*(4), 1429–1443. <https://doi.org/10.1111/j.1467-8624.2012.01784.x>
- Lin, T.-J., Ha, S. Y., Li, W.-T., Chiu, Y.-J., Hong, Y.-R., & Tsai, C.-C. (2019). Effects of collaborative small-group discussions on early adolescents' social reasoning. *Reading and Writing, 32*(9), 2223–2249. <https://doi.org/10.1007/s11145-019-09946-7>
- Lin, T.-J., Jadallah, M., Anderson, R., Baker, A., Nguyen-Jahiel, K., Kim, I.-H., Kuo, L.-J., Miller, B., Dong, T., & Wu, X. (2015). Less is more: Teachers' influence during peer

- collaboration. *Journal of Educational Psychology*, 107, 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), Article e12442. <https://doi.org/10.1111/bjep.12442>
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732–749. <https://doi.org/10.1111/j.1467-8624.2008.01154.x>
- Mikami, A. Y., Griggs, M. S., Lerner, M. D., Emeh, C. C., Reuland, M. M., Jack, A., & Anthony, M. R. (2013). A randomized trial of a classroom intervention to increase peers' social inclusion of children with attention-deficit/hyperactivity disorder. *Journal of Consulting and Clinical Psychology*, 81(1), 100–112. <https://doi.org/10.1037/a0029654>
- Muhinyi, A., Hesketh, A., Stewart, A. J., & Rowland, C. F. (2020). Story choice matters for caregiver extra-textual talk during shared reading with preschoolers. *Journal of Child Language*, 47(3), 633–654. <https://doi.org/10.1017/S0305000919000783>
- Mundelsee, L., & Jurkowski, S. (2021). Think and pair before share: Effects of collaboration on students' in-class participation. *Learning and Individual Differences*, 88, Article 102015. <https://doi.org/10.1016/j.lindif.2021.102015>
- Murphy, P. K., Croninger, R. M. V., Baszczewski, S. E., & Tondreau, C. L. (2022). Enacting Quality Talk discussions about text: From knowing the model to navigating the dynamics of dialogic classroom culture. *The Reading Teacher*, 75(6), 717–731. <https://doi.org/10.1002/trr.2110>
- Murphy, P. K., Greene, J. A., Firetto, C. M., Li, M., Lobczowski, N. G., Duke, R. F., Wei, L., & Croninger, R. M. V. (2017). Exploring the influence of homogeneous versus heterogeneous grouping on students' text-based discussions and comprehension. *Contemporary Educational Psychology*, 51, 336–355. <https://doi.org/10.1016/j.cedpsych.2017.09.003>
- 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>
- Noble, C. H., Cameron-Faulkner, T., & Lieven, E. (2018). Keeping it simple: The grammatical properties of shared book reading. *Journal of Child Language*, 45(3), 753–766. <https://doi.org/10.1017/S0305000917000447>
- Nyhout, A., & O'Neill, D. K. (2013). Mothers' complex talk when sharing books with their toddlers: Book genre matters. *First Language*, 33(2), 115–131. <https://doi.org/10.1177/0142723713479438>
- Oatley, K. (2016). Fiction: Simulation of social worlds. *Trends in Cognitive Sciences*, 20(8), 618–628. <https://doi.org/10.1016/j.tics.2016.06.002>
- Partee, A. M., Hamre, B. K., & Williford, A. P. (2020). Examining the role of preschool classrooms' behavioral composition as a predictor of the quality of teacher-child interactions. *Early Education & Development*, 31(6), 873–891. <https://doi.org/10.1080/10409289.2019.1694813>
- Peugh, J. L., & Heck, R. H. (2017). Conducting three-level longitudinal analyses. *The Journal of Early Adolescence*, 37(1), 7–58. <https://doi.org/10.1177/0272431616642329>
- Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science (New York, N.Y.)*, 315(5820), 1795–1796. <https://doi.org/10.1126/science.1139719>
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119. <https://doi.org/10.3102/0013189X09332374>
- Pianta, R. C., Hamre, B. K., & Mintz, S. L. (2012). *Classroom Assessment Scoring System (CLASS) Upper Elementary Manual*. Teachstone.
- Pianta, R. C., Lipscomb, D., & Ruzek, E. (2021). Coaching teachers to improve students' school readiness skills: Indirect effects of teacher-student interaction. *Child Development*, 92(6), 2509–2528. <https://doi.org/10.1111/cdev.13600>
- R Core Team. (2021). *R: The R project for statistical computing*. <https://www.r-project.org/>
- Reznitskaya, A., & Wilkinson, I. A. G. (2021). The Argumentation Rating Tool: Assessing and supporting teacher facilitation and student argumentation during text-based discussions. *Teaching and Teacher Education*, 106, Article 103464. <https://doi.org/10.1016/j.tate.2021.103464>
- Rimm-Kaufman, S. E., Baroody, A. E., Larsen, R. A. A., Curby, T. W., & Abry, T. (2015). To what extent do teacher-student interaction quality and student gender contribute to fifth graders' engagement in mathematics learning? *Journal of Educational Psychology*, 107(1), 170–185. <https://doi.org/10.1037/a0037252>
- Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-K programs predict children's learning? *Science*, 341(6148), 845–846. <https://doi.org/10.1126/science.1233517>
- Sameroff, A. (2009). The transactional model. In *The transactional model of development: How children and contexts shape each other* (pp. 3–21). American Psychological Association. <https://doi.org/10.1037/11877-001>
- Schrijvers, M., Janssen, T., Fialho, O., De Maeyer, S., & Rijlaarsdam, G. (2019). Transformative Dialogic Literature Teaching fosters adolescents' insight into human nature and motivation. *Learning and Instruction*, 63, Article 101216. <https://doi.org/10.1016/j.learninstruc.2019.101216>
- Sette, S., Gasser, L., & Grütter, J. (2020). Links between teachers' liking of students, peer inclusion, and students' academic achievement: A two-wave longitudinal study. *Journal of Youth and Adolescence*, 49(3), 747–756. <https://doi.org/10.1007/s10964-019-01048-5>
- Tompkins, V., Benigno, J. P., Lee, B. K., & Wright, B. M. (2018). The relation between parents' mental state talk and children's social understanding: A meta-analysis. *Social Development*, 27(2), 223–246. <https://doi.org/10.1111/sode.12280>
- Venkadasalam, V. P., Jenkins, J. M., Ganea, P. A., & Wade, M. (2022). Linking early maternal input during shared reading to later theory of mind through receptive language and executive function: A within- and between-family design. *Journal of Experimental Child Psychology*, 223, Article 105469. <https://doi.org/10.1016/j.jecp.2022.105469>
- Walker, L. J., Hennig, K. H., & Krettenauer, T. (2000). Parent and peer contexts for children's moral reasoning development. *Child Development*, 71(4), 1033–1048. <https://doi.org/10.1111/1467-8624.00207>
- Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology*, 98(1), 63–74. <https://doi.org/10.1037/0022-0663.98.1.63>
- West, B. T., Welch, K. B., & Galecki, A. T. (2022). *Linear mixed models: A practical guide using statistical software*. CRC 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, Article 101732. <https://doi.org/10.1016/j.learninstruc.2023.101732>
- LUCIANO GASSER is a professor at the Bern University of Teacher Education. His research focusses on social and moral development and teacher practices for inclusive classrooms.
- DAVID PREISIG is a PhD student at the Bern University of Teacher Education. His research interests lie in school-based interventions focused on social and emotional learning.
- ANNA FREI is a PhD student at the Bern University of Teacher Education. Her research interests lie in the interplay of interpersonal processes and social-cognitive development.
- YVONNE DAMMERT is a PhD student at the Bern University of Teacher Education. Her research interests include facilitation of productive classroom discussions, and video-based coaching.
- SARA EGGER is a postdoc and research associate at the Bern University of Teacher Education. She studies effects of school interventions on intergroup processes.
- P. KAREN MURPHY is a distinguished professor at The Pennsylvania State University. Her current research focuses on the role of classroom discussion on students' high-level comprehension.