Child Development, 2026, 1–16
https://doi.org/10.1093/chidev/aacag089
Published: May 18, 2026
Empirical Article

Society for Research in Child Development OXFORD

Improving children's collective and individual moral reasoning through small-group discussions about narrative fiction

Luciano Gasser[1,*], David Preisig[1], Yvonne Dammert[1], Anna Frei[1], Sara Egger[1], and P. Karen Murphy[2]

[1]Institute for Research, Development, and Evaluation, Bern University of Teacher Education, Bern, Switzerland
[2]Department of Educational Psychology, Counseling, and Special Education, The Pennsylvania State University, University Park, PA, United States
*Corresponding author: Institute for Research, Development and Evaluation, Bern University of Teacher Education, Fabrikstrasse 8, Bern CH-3012, Switzerland. Email: luciano.gasser@phbern.ch

Abstract

This randomized trial tested whether small-group discussions about narrative fiction improve children's collective and individual moral reasoning. The sample consisted of 704 Swiss children ($M_{age} = 10.32$, 49.3% girls, 29.8% with a non-Swiss German home language), 159 discussion groups, and 51 teachers across 3 waves in 2022–2023. Data included coded discussions and essays on conflicts between fairness and group loyalty. The intervention enhanced collective reasoning, indicated by increases in student talk, questions, reasons, challenges, and build-ons (*RR*s = 1.37–8.62) and decreases in unreasoned claims (*RR* = 0.49). For individual reasoning, children in the intervention group provided more moral justifications (*RR* = 1.67) but not more group-based justifications and shifted to higher justification levels, including more dual-domain responses (*OR* = 2.05).

Keywords narrative fiction, dialogue, moral development

Lay summary

This study examined whether reading and discussing children's books can help children reason about moral issues. Participants were 704 children who were assigned to either an intervention or a control group. Teachers in the intervention group received training to hand over more control to children and to scaffold children's argumentation during literary discussions. Over the school year, children read and discussed 4 children's books that addressed peer group exclusion. Children in the intervention group engaged more actively and argumentatively in moral dilemma discussions than children in the control group. Their written responses also showed more moral justifications and greater integration of different perspectives. The results indicate that regular literature lessons not only support language learning but also children's moral development.

Narrative fiction represents a unique context for cultivating moral judgments and emotions (Carroll, 2000; Nussbaum, 1985). Through artistic narration, it enables readers to immerse themselves in unfamiliar lives and to train social imagination and empathy. It stimulates moral reasoning by presenting conflicting perspectives on social conflicts without authorial resolution, requiring readers to weigh these perspectives on their own (Bakhtin, 1984). Over the past 15 years, empirical evidence has demonstrated that narrative fiction effectively stimulates theory of mind and empathy in young and adult readers (Mar et al., 2010; van der Kleij et al., 2025; Wimmer et al., 2024). Moreover, shared book reading in early childhood provides unique opportunities for high-quality conversations that contribute to children's social-cognitive development (Tompkins et al., 2018).

However, little is known about how children's literature can be effectively used to foster moral development in schools. Recently, intervention researchers have argued for social-emotional learning (SEL) programs that naturally take up the social learning opportunities within regular subject areas to promote both academic and social learning (Jones et al., 2011; Lin et al., 2022). Compared with extracurricular interventions, an integrated approach has the advantage that teachers and students do not perceive social learning as an "add-on" but as an integral part of high-quality instruction. This approach not only increases

Handling Editor: Audun Dahl
Editor: Shauna Cooper
Received: April 11, 2025. **Revised:** March 3, 2026. **Accepted:** April 14, 2026
© The Author(s) 2026. Published by Oxford University Press on behalf of the Society for Research in Child Development. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

instructional time ("two-for-one") but also supports a holistic learning experience by meaningfully connecting academic content to children's lives (Nucci & Ilten-Gee, 2021).

This randomized controlled trial investigated the effects of small-group discussions about narrative fiction on fourth and fifth graders' individual and collective reasoning about hypothetical moral dilemmas. Over one academic year, children in the intervention group discussed 4 children's books addressing peer group conflicts, while the control group continued with business-as-usual (BAU) language arts instruction. Changes in children's dialogic practices were assessed through coding of transcribed small-group discussions about a hypothetical story involving a conflict between fairness and group loyalty, and individual reasoning was measured through written essays on the same conflict. Although prior research has examined how shared readings in family and school settings support children's social-cognitive development (e.g., Tompkins et al., 2018), surprisingly few studies have addressed their role in moral reasoning.

Developmental trajectories in individual moral reasoning

Moral reasoning is defined as thinking about "right" and "wrong" based on principles such as justice, welfare, and equal rights (Dahl et al., 2022). It differs from basic social competencies such as perspective-taking or empathy. While these skills are essential for effective participation in social relationships and groups, they are insufficient for challenging the status quo. Moral reasoning evaluates social practices and norms based on whether they disadvantage and harm specific individuals or groups and serves as a catalyst for societal change toward greater justice (Killen & Dahl, 2021). Therefore, intervention programs should focus on moral competencies in addition to basic social skills (Nucci & Ilten-Gee, 2021).

Guided by the *Social Reasoning Development* (SRD) approach (Killen & Rutland, 2011), this study focuses on children's reasoning about conflicts between loyalty to friends (e.g., protecting the group) and fairness (e.g., preventing harm) in contexts of peer group exclusion. SRD research investigates how children evaluate and reason about exclusion based on group membership (e.g., ethnicity, disability) and interpersonal characteristics (e.g., shyness). Children are viewed as active agents in their social worlds who construct moral meaning through social interaction and reasoned exchanges about conflictual moral issues. In particular, peer discussions are theorized to play a key role in changes in individual moral reasoning and peer group norms (Killen et al., 2022).

SRD research has shown that even preschoolers predominantly reject unfair social exclusion from peer groups (Gasser et al., 2014; Killen et al., 2001; Theimer et al., 2001). This finding can be explained by children's early conceptual understanding of morality. Research from the Social Domain Theory perspective has shown that 3-year-old children understand that moral rules (compared to social conventions) are valid independent of context and authority demands (Dahl et al., 2022; Yoo & Smetana, 2022). This empowers young children to critique unjust group practices (Killen & Rutland, 2011).

However, peer group conflicts are often multifaceted, involving both moral and group-based concerns that children must weigh and prioritize. SRD research revealed that children develop an increasingly sophisticated understanding of group dynamics, particularly regarding the critical role of conformity in maintaining group cohesion and identity (Mulvey, 2016). Older children better understand that disloyal behavior threatens group norms and can justify excluding nonconforming members (Cooley & Killen, 2015). Accordingly, they increasingly consider the costs of disloyal behavior for themselves and the group.

While excluding an individual from a peer group may be justified in some situations—for example, when the individual and the group diverge substantially in basic interests—social exclusion can also undermine moral principles, especially when based on stereotypes or prejudice (Killen & Rutland, 2011). In such situations, children face a tension between their motivation for belonging and loyalty at the local group level and their principled sense of fairness and justice (Mulvey et al., 2023). With age, children improve in coordinating moral and group-based concerns in context-sensitive ways (Gasser et al., 2014). These growing coordination skills support nuanced judgments about when the moral implications of socially exclusionary group practices take precedence over group interests (Cooley & Killen, 2015).

Collective moral reasoning

According to constructivist theories, dialogic engagement in moral conflicts is the key mechanism through which individual moral reasoning develops (Killen & Rutland, 2011; Mammen & Paulus, 2023; Piaget, 1932). Building on the *transactive discourse* framework (Berkowitz & Gibbs, 1985; Kruger & Tomasello, 1986; Nucci, 2016), we define collective moral reasoning procedurally as collaborative attempts to reach uncoerced agreements on moral conflicts. This requires that all participants articulate their perspectives and negotiate the validity of claims by providing reasons and evaluating others' claims.

An important condition for such co-constructive discussions is that children can take an active role (Piaget, 1932). Children's agency is indicated by their spontaneous initiations, such as contributions and questions (Kruger & Tomasello, 1986; Mammen et al., 2019). This enables children to lead and shape the content and form of discussions, which have been positively linked to their moral development (Kruger, 1992). Conversely, if discussion partners minimize children's autonomy, for example, by moral lecturing, children show less growth in moral reasoning (Walker et al., 2000; Walker & Taylor, 1991).

Children's active roles in moral dilemma discussions are more likely to emerge in interactions with peers than with adults (Kruger & Tomasello, 1986; Mammen et al., 2019). In the classroom, discussions typically follow the Initiation–Response–Evaluation (IRE) pattern (Nystrand et al., 2003): teachers initiate with closed-ended questions (I), children respond with short answers (R), and teachers then evaluate them as right or wrong (E). Yet, social hierarchies in classroom discussions are not fixed and can be changed through professional development (Murphy et al., 2009). When teachers are trained to release control to students during dilemma discussions, students' active engagement increases, as indicated by more frequent student talk and questioning (Chinn et al., 2001).

Beyond active engagement, children's moral development hinges on the argumentative quality of exchanges in moral discussions (Nucci, 2016). Argumentation in moral discussions is often

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

described in terms of "transacts," defined as interactions in which partners collaboratively build on each other's reasoning through justification, extension, and challenge (Berkowitz & Gibbs, 1985). A recent short-term intervention study showed that both requests for justification and disagreements helped children advance their reasoning about fair sharing and punishment in an allocation task (Li & Tomasello, 2022). While requests for justification force children to publicly present their thinking, challenges create cognitive conflict that prompts reconsideration. However, challenges are not strictly necessary for gains in moral reasoning. When children listen to others and build on their contributions, they also contrast and integrate multiple perspectives, which supports more balanced moral reasoning (Damon & Killen, 1982). For example, a 4-year longitudinal study with 10-year-old children found that dilemma discussions with peers and parents—where children's reasoning was encouraged and acknowledged rather than criticized—were effective in fostering moral reasoning (Walker et al., 2000). Research in classrooms has shown that if teachers are trained in facilitating transactive discussions, the argumentative quality of discussions and students' individual moral reasoning improve (Nucci et al., 2015).

Narrative fiction and moral reasoning

Narrative fiction serves as a unique stimulus for rich individual and collective moral reasoning. Unlike hypothetical vignettes, which are commonly used in social-cognitive and moral learning programs, narrative fiction is characterized by a high level of detail (Mar & Oatley, 2008). This detail supports intense perceptual and emotional experiences, allowing readers to adopt the inner perspective of characters, which makes complex moral decisions more vivid and immediate. Accordingly, Nussbaum (1985) argued that narrative fiction enables deeper moral reflections than abstract moral dilemmas.

Unlike real-life experiences, narrative fiction provides a safe space for such reflections, as readers can disengage from overwhelming immersive experiences at any time (Vezzali et al., 2015). Moreover, narrative fiction is more structured than real-life experiences (Mar & Oatley, 2008). While real events are often shaped by randomness, narrative fiction exhibits a high degree of coherence, which facilitates meaning-making for readers. This might explain why shared book reading has been found to be a particularly effective context for promoting children's social-cognitive development (Tompkins et al., 2018).

Despite the strong consensus on the educational value of children's literature for socio-moral development, developmental and educational approaches present divergent views on how children learn from narrative fiction (Gasser et al., 2022). While some highlight that children acquire new moral insights by reading stories (*efferent* approach; e.g., Lee et al., 2014), others emphasize practicing social-cognitive skills through perspective-taking with characters during reading (*expressive* approach; Kidd & Castano, 2013; van der Kleij et al., 2025). The *critical-analytic* approach foregrounds argumentative discussions of narrative fiction as a rich context for collaborative peer deliberation on complex moral issues (Lin et al., 2022; Nucci & Ilten-Gee, 2021). According to this approach, the moral messages of the text and the readers' affective responses serve as a starting point for deeper engagement with the text. This engagement is inherently collaborative, as it requires peers to critically examine the legitimacy of differing affective responses and jointly develop well-examined interpretations. In such discussions, children learn key argumentative practices, such as justifying their opinions on moral issues, respectfully challenging others, and building on prior speakers' contributions. It is assumed that this social practice of articulating and balancing different perspectives serves as a model for individual moral reasoning.

The current intervention

The intervention in this study is based on the critical-analytic approach to social learning through narrative fiction. We use the text-based discussion approach, Quality Talk (QT), which supports children's argumentative and critical exchanges in small groups (e.g., Murphy et al., 2022). Through mini-lessons, including explicit instruction, modeling, and practice, children first learn different types of open-ended questions that help steer discussion and elicit rich talk about the story's social and moral issues. For example, these include exploring characters' mental states (*Character Questions*), engaging in moral reasoning about characters' problem-solving and the text's moral message (*Big Questions*), and asking how they would respond to a moral conflict as the protagonist (*Affective Questions*). Additionally, children learn to respond argumentatively to these questions by providing reasons for their claims, building on peers' responses, and challenging peers' viewpoints.

In the subsequent book lessons, children held small-group discussions throughout the school year on 4 children's books. These books addressed complex topics about group conflicts that encouraged children to consider group-based concerns (e.g., importance of group membership and loyalty) but to critically reflect on unfair group practices from a moral point of view. In addition to the thematic relevance, we also considered aesthetic criteria, such as complex characterization, ambiguity, and polyphony of perspectives, which invite children to adopt multiple viewpoints on the text (Bruner, 2011; Kidd & Castano, 2013). According to QT, teachers assume a new role as facilitators of student-centered discussions and relinquish control over the content and form of the discussion to the children.

This intervention was designed for late childhood because this period is characterized by increasing motivation for peer-group affiliation and a growing orientation toward group norms and conformity (Ellis & Zarbatany, 2017; Killen & Rutland, 2011). As a result, older children more frequently face situations in which maintaining group belonging may conflict with acting in line with moral principles (LaFontana & Cillessen, 2010). At the same time, their growing capacity to reason with peers provides a key resource for maintaining autonomy and resisting problematic peer pressure (Allen et al., 2006). Accordingly, strengthening children's moral reasoning in peer discussions constitutes a developmentally timely intervention target in late childhood and the transition to early adolescence.

The present study

This study examined intervention effects on individual and collective moral reasoning about hypothetical conflicts between fairness and group loyalty at 3 measurement points (T0 = pretest; T1 = midtest; T2 = posttest). Collective reasoning was assessed in a small-group discussion format, capturing children's active

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

engagement (word counts, questions) and argumentation (reasons, unreasoned claims, build-ons, challenges). Individual reasoning was assessed by coding children's written justifications to the same stories, resulting in 3 scores: (a) moral justifications, (b) group-based justifications, and (c) justification level as a proxy for domain coordination (0 = no justification; 1 = single-domain; 2 = dual-domain).

We hypothesized that, relative to the BAU control group, children in the intervention group would show greater growth in active engagement and argumentation during moral dilemma discussions. For individual reasoning, we hypothesized that moral developmental success would be evidenced by 2 criteria (cf. Nucci et al., 2015): (1) more moral justifications in written responses—taken as evidence of prioritizing fairness over group loyalty, given that the study's dilemma stories impose high moral costs for loyalty to friends; and (2) shifts in justification level from no justification to single-domain (moral or group-based) and dual-domain justifications, indicating progress in coordinating both concerns.

We implemented a multilevel approach by considering the hierarchical structure of the data (e.g., students nested in small groups) and included various control variables at the levels of measurement point (e.g., story type: Dilemma A, B, or C), child (e.g., reading comprehension, gender), and small group (e.g., size, group composition). In randomized controlled trials, average treatment effects can be misestimated when true change is curvilinear. Accordingly, testing a quadratic time effect is recommended even with 3 waves (Yang et al., 2024). We therefore ran exploratory tests of a quadratic time term and its interaction with the intervention to determine whether change was linear or curvilinear.

Method

Sample and design

The sample consisted of 704 Swiss children (49.3% girls, $M_{age} = 10.32$ years, $SD = 0.67$) from 41 fourth- and fifth-grade classrooms in German-speaking cantons of Switzerland ($n = 321$ and $n = 383$, respectively). Because ethnicity is not collected in Swiss schools or by the Federal Statistical Office, we report migration-related indicators instead. In total, 13.2% of the children were born abroad, and 29.8% did not speak Swiss German at home. Additionally, 51 teachers (86.3% women) participated in the study ($M_{age} = 37.58$ years, $SD = 10.31$), with an average professional experience of 14.76 years. In ten classrooms, teachers participated in pairs. Class sizes averaged 19.27 children ($SD = 2.90$). Parents and children provided active consent (3.0% refusal). The study was approved by the Ethics Committee of the Bern University of Teacher Education.

The study included 3 measurement points: T0 (M = September 2, 2022, $SD = 9.36$ days), T1 (M = January 26, 2023, $SD = 11.66$ days), and T2 (M = June 13, 2023, $SD = 10.95$ days). Before T0, teachers formed heterogeneous small groups with respect to academic achievement and migration background. These 159 small groups (including the teacher) remained constant across all waves and intervention discussions. Research assistants monitored adherence to group assignments at each wave.

Classrooms were randomly assigned to the intervention or BAU control group, stratified by grade and school-level migration share. One of the initial 21 control classrooms withdrew at pretest. The final sample comprised 20 control and 21 intervention classrooms, with no significant differences in gender distribution, grade level, home language, foreign-born share, reading comprehension, or group sizes between intervention and control groups (see Table A1 in the supplemental materials). However, a marginally significant effect revealed that discussion groups were smaller in the intervention than in the control group at T0 (but not at T1 and T2). Group size was included as a covariate in all analyses.

For sample-size justification, we conducted an a priori, simulation-based power analysis to inform the design-level decision regarding the number of classrooms. Effect-size inputs came from a meta-analysis of text-based discussion interventions (Murphy et al., 2009) and from primary studies reporting means and standard deviations (e.g., Chinn et al., 2001). Because these studies differed substantially in how they operationalized discussion quality and reasoning and often relied on small samples, the assumed intervention effects varied considerably, ranging from standardized $\beta = 0.26–0.91$ for discussion outcomes and standardized $\beta = 0.03–0.87$ for reasoning outcomes. The power analysis was implemented in R using a simplified 2-level planning model (with classrooms as the clustering unit). We evaluated a mediation-based (indirect-effect) planning model (Intervention → discussion outcomes → reasoning outcome) as a conservative benchmark for determining the required number of clusters. We compared designs with 30 versus 40 classrooms (20 students each) under $\alpha = .05$. Simulations indicated power of .79–.95 for 40 classrooms across prespecified effect-size combinations, whereas a 30-classroom design was underpowered. Accordingly, we targeted at least 40 classrooms.

Procedure

This study was not preregistered. The present report includes all measures that were assessed at all 3 measurement occasions, namely (i) written essays and (ii) video-recorded small-group discussions. We applied only those coding categories relevant to the current research questions and did not include codes outside this scope (e.g., teacher behaviors). At pretest and posttest (but not midtest), we additionally administered measures of children's perspective taking, intergroup attitudes, perceived intergroup norms, and peer-nominated relationships. These measures are not analyzed here. All manipulations are described below.

Stories

Drawing on vignette structures from SRD research (e.g., Malti et al., 2012; Mulvey & Killen, 2015), we developed 3 original dilemma stories (A, B, and C). Each story portrayed a conflict between loyalty to friend(s) and resistance to the group or friend because of their unfair behavior toward a socially marginalized child (for the full stories, see online supplemental materials, Chapter 2). For example, in Story A, the protagonist befriends a new neighbor who has recently immigrated to Switzerland and will attend the same school. The new friend asks the protagonist to show him around during recess, and the protagonist agrees and promises to wait for him the next day. However, while waiting during recess, other friends press the protagonist to join their soccer team, emphasizing their urgent need for his support and downplaying the importance of helping the newcomer.

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

The core conflict was derived from SRD vignettes, but in contrast to the typically very short vignettes used in SRD studies, our stories were intentionally written as longer, more detailed narratives. This was done to approximate the fictional narrative context of the intervention with richer situational cues for discussion. The stories were extensively piloted with children of the target age group to ensure that they offered sufficient complexity to stimulate controversial discussions. The stories were comparable in length (A = 576 words, B = 623 words, C = 595 words) and difficulty (i.e., average word and sentence length) as indicated by the LIX index (A = 26.8, B = 27.16, C = 26.92; Lenhard & Lenhard, 2011). At each measurement point, children read and discussed a different story. The 3 stories were counterbalanced across measurement points and conditions in 6 orders to control for potential order effects. Stimuli were presented as audio recordings on tablets, with printed texts provided to support comprehension and to allow students to refer back to the text during discussion. Small groups discussed in 4 successive morning slots in the same order across waves. To test for discussion-order (fatigue) effects, we computed group-level Spearman correlations between slot (1–4) and all outcomes. Correlations were near zero at each occasion (T0: $\rho = -0.13$ to 0.06; T1: −0.15 to 0.11; T2: −0.08 to 0.05; all $ps \geq .066$).

Essays and discussions

Each story ended with the question: "*What should the protagonist do?"* Children had 10 minutes to write an individual response ($M = 42.41$ words, $SD = 25.38$). They then engaged in teacher-facilitated small-group discussions (target duration: 15 minutes, observed duration: $M = 15.00$ minutes, $SD = 2.19$, range = 6–24), approaching the text as in their usual language arts lessons. All discussions were video-recorded. Children who declined filming sat outside the camera's view ($n = 137$). We transcribed 470 discussions via a professional service and analyzed them with MAXQDA 24 (VERBI Software, 2024). Essays and discussions were coded by 7 trained undergraduate coders who, after 13–16 training sessions, achieved overall interrater agreement above 80% for 3 consecutive sessions, assessed separately for the essay and the discussion coding system (see supplemental materials, Chapter 3, for coding examples).

Instruments

Collective reasoning

Consistent with the transactive discourse framework (Mammen et al., 2019; Nucci et al., 2015; Walker et al., 2000), we analyzed discussions according to formal rather than content criteria, assuming that moral reasoning develops through dialogues in which children actively participate and negotiate perspectives through argumentation. To operationalize these properties, we adapted categories from the QT Discussion Manual (Murphy, 2018), which map onto transactive discourse moves. In contrast to the original version, coding was conducted at the speaker-turn level rather than the question-event unit (which spans multiple turns), to enable child-level analyses. For each speaking turn (i.e., a single child utterance), each code could be assigned at most once. Fifteen percent of the transcripts (70 discussions) were double-coded. All Cohen's kappa values indicated good interrater agreement (questions: $\kappa = .80$; reasons: $\kappa = .71$; unreasoned claims: $\kappa = .73$; challenges: $\kappa = .73$; build-ons: $\kappa = .80$).

Active engagement

We used the total word count and the frequency of child-initiated questions as indicators of children's active engagement in discussions. A question was coded when a child asked a topic-related or procedural question inviting peers to share their ideas (e.g., "What would you do in this situation?" "What do you think?"). Clarification questions were not coded as questions, as they were closed-ended and mostly directed toward the teacher (e.g., "What's the name of this character?").

Argumentation

We classified reasons, challenges, and build-ons as high-quality indicators of children's argumentation, and unreasoned claims as a low-quality indicator.

Reasons were coded whenever a contribution provided a justification for its claim (i.e., explained why the claim should hold), including explicit justifications (e.g., "He should wait, because he promised"), if–then rules (e.g., "If you make a promise, then you should keep it—so he should wait"), or analogies, comparisons, categorizations that support the claim (e.g., "Telling on a friend is like betrayal—you stand by your friends," "Repeating pranks is bullying, and bullying is wrong").

Unreasoned claims were simple statements in which children expressed an idea without justification, without challenging or building on (e.g., "I think it's mean," "I would tell my friends to stop doing this").

Challenges referred to responses that questioned, criticized, or rejected a previous statement, often prompting the consideration of a neglected perspective. They could take the form of direct disagreement, partial opposition, or critical questions. Challenges needed to be expressed explicitly, for example through contrastive words like "but" or "however," or phrases such as "I disagree" (e.g., "But if they didn't help her, why should she help them?").

Build-ons were coded when children responded to a peer's contribution without teacher interruption or explicitly referred to a peer's argument using their name (e.g., Child1: "I think he wants to make her happy," Child2: "I think so too. He really surprised her with his gift"). Simple repetitions or brief completions of others' statements, such as finishing a sentence without adding new information, justification, or inference, were not coded as build-ons.

Individual reasoning

The coding system for individual reasoning combined a deductive, theory-driven approach with inductive, data-driven refinement. Guided by the SRD framework, we prespecified the 2 superordinate categories—moral and group-based justifications—and several subcategories (Cooley & Killen, 2015; Gasser et al., 2014; Theimer et al., 2001; Yüksel et al., 2025). Moral justifications reflect children's concerns about the unfairness and harm of excluding an outgroup member. Group-based justifications capture loyalty demands and anticipated social costs tied to either a peer group or a close friend. Through iterative coding, additional subcategories emerged and were mapped onto the moral or group-based domain. We segmented each essay into justification units, defined as the smallest meaning unit expressing one distinct reason relevant to the child's chosen course of action, including arguments against it. Units could consist of a clause, a sentence, or multiple sentences, as long as they conveyed the same reason. Each subcategory was coded at most once per

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

essay, yet essays could contain several moral or group-based subcategories. The 2 types are not mutually exclusive and may co-occur within the same response.

All essays were double-coded. Disagreements were discussed and resolved. For the analyses, all justification subtypes were collapsed into 2 broad categories (moral vs. group-based). Interrater reliability for individual reasoning frequencies was good to excellent, with ICCs of .92 for moral and .90 for group-based justifications.

Moral justifications

Moral justifications were coded if responses referred to: (i) justice and equal rights (e.g., "because she must tell the truth"), (ii) the victim's welfare (e.g., "because he would be sad," "because this would help him defend himself"), (iii) perspective-taking with the victim by expressing a first-person role shift into the victim's situation (e.g., "I know how it feels to be bullied," "You realize that it wasn't nice if you take the victim's perspective"), (iv) fairness norms in friendships (e.g., "because if she is a good friend, she will understand why I did this"), and (v) references to concrete experiences—direct or vicarious—highlighting harmful or beneficial consequences of (im)moral behavior (e.g., "because I once experienced how it hurts to be excluded," "I was once in a similar situation, and I would've appreciated it if someone had helped me"). Justifications that referred to low costs of a moral decision (e.g., "because my friend will not know that I helped the victim") or to sanctions (e.g., "she won't be punished if she tells the truth") were not coded as moral.

Group-based justifications

Group-based justifications were coded when children referred to: (i) loyalty to group or friend (e.g., "you should stick with your friend"), (ii) group or friend repercussions (e.g., "she could lose her friend," "her friends could take revenge"), (iii) risks for the ingroup friend(s) (e.g., "the teacher will punish his friends"), and (iv) references to similar experiences with the negative consequences of disloyalty (e.g., "because I once lost my friend after telling the truth"). Justifications that legitimized or downplayed the harm were not classified as group-based (e.g., "because it's the victim's own fault").

Justification level

To assess whether the intervention increased children's integration of moral and group-based justifications (see Kuhn & Crowell, 2011; Nucci et al., 2015), we created an ordinal scale: 0 = no justification; 1 = one or more justifications from a single domain (moral *or* group-based); 2 = justifications from both domains.

Reading comprehension

Reading comprehension was assessed using a beta version of ELFE III, a revision of the standardized reading comprehension test ELFE II, which is designed for students in Grades 1–7 (Lenhard et al., 2017). ELFE III is currently under development. The version used here was provided for research purposes. The test was administered on tablets in a multiple-choice format with fixed time limits for each subtask. It includes 3 subscales targeting reading comprehension at the word, sentence, and text levels. All 3 subscales were z-standardized, and a composite mean score was calculated (Cronbach's $\alpha = .84$).

Group-level characteristics

Group size

Group size was assessed by counting the number of children present during a discussion. Group size was calculated separately for each measurement point.

Group home language

At each measurement point, we calculated the group's proportion of children with a non–Swiss German home language (0 = *all children speak Swiss German at home*; 1 = *none of the children speak Swiss German at home*).

Intervention

Details of all intervention modules are provided in the online supplemental materials (Table A4). Sample materials and discussion videos are available at https://www.phbern.ch/skill.

Professional training for teachers

Before the pretest, teachers completed 2 on-site half-day training sessions. Session 1 introduced the QT approach and the intervention materials. The trainers guided teachers in evaluating the quality of children's questions and responses using videos of literary discussions. Teachers reflected on the dominant role of teachers in typical classroom discussions and how they can shift responsibility to children. They also learned to use specific scaffolding strategies, referred to as Talk Moves, to enhance the argumentative quality of discussions, such as eliciting justifications, encouraging children to build on each other's contributions, or challenging dominant perspectives. Session 2 focused on 4 children's books. Teachers were introduced to SRD and engaged in discussions about how moral, group-related, and personal concerns conflict in the children's books. All books addressed peer group conflicts, illustrating how inclusive versus exclusive group norms shape group members' behavior and depicting story characters who successfully or unsuccessfully challenged unfair group behavior. Teachers were trained to take the perspective of victims of exclusion and to reflect on the perspectives of excluders and bystanders. Teachers received detailed lesson plans and materials (Dammert & Gasser, 2025).

Teachers received up to 4 video-based coaching sessions ($M = 3.11$, $SD = 0.79$, range: 0–4). These sessions were designed to strengthen teachers' observational and reflective skills for assessing the quality of collective reasoning during their literary discussions. For each session, teachers submitted a video, and coaches selected short clips and provided prompts before the meeting. During a one-hour online meeting, teachers discussed their responses with the coaches and planned next steps. The coaching team consisted of 4 experienced teachers, including the third author, who trained the team and coached some classrooms.

Intervention lessons for children

The intervention consisted of 2 distinct phases. During the initial 2 weeks (phase 1), children completed 9 mini-lessons introducing 5 types of authentic (open-ended) questions and core elements of argumentative responses. For example, students learned to ask affective questions that connect the story to personal experiences and connection questions that link the story to other texts or media. In the lessons on argumentative responses, children learned the elements of a full and coherent argument, the

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

importance of counterarguments and co-construction. Each lesson followed 3 steps: a brief PowerPoint presentation that defined the targeted discourse element and provided linguistic prompts; an animated video modeling it; and guided practice activities, such as generating a specific question type in response to a comic video. In phase 2, across the school year, children then read and discussed 4 children's books. Each book was split into 2 to 3 segments, and students received a separate literacy journal for each book. Before reading, a social-conceptual lesson built vocabulary and background knowledge about the book's core themes. During reading, students identified authentic questions and completed playful activities in their literacy journals (e.g., quizzes, drawn scenes) to support comprehension and emotional engagement with the text. They also set personal discussion goals (e.g., better listening).

The literary discussions took place in teacher-facilitated small groups. The teacher led each discussion from beginning to end. Group composition remained consistent with that of the assessments. While one group discussed with the teacher, the remaining students worked independently on their literacy journals. The small-group format was chosen to foster active participation, particularly among shy or anxious children. Each discussion ended with a brief teacher-led reflection on how the discussion unfolded and whether students met their goals.

Implementation fidelity

After each session, coaches rated teachers' motivation to improve their discussion facilitation strategies (0 = *very low*, 4 = *very high*). Their motivation was consistently high ($M = 3.26$, $SD = 0.63$), with no change across sessions. Coaches also rated discussion quality (4 items) for each videotaped session, which increased significantly over time, $F(1, 22.85) = 71.51$, $p < .001$, indicating high implementation fidelity and professional growth. On average, teachers invested 33.03 hours ($SD = 6.27$, range = 12.58–42.00) in the program: 23.0% for mini-lessons and 77.0% for book lessons. Self-reported deviations from lesson plans (0 = *never*, 3 = *frequently*) were minor ($M = 1.05$, $SD = 0.19$). Teachers conducted 9.05 small-group discussions on average. Thirteen classes completed 3 children's books, and 7 completed all 4. One teacher discussed only one book and did not receive coaching. In line with the intention-to-treat principle, we retained this case in the analyses.

Missing data

At the small-group level, missingness occurred because (a) in 2 team-taught classes at T0, the teachers led the wrong small groups ($n = 3$) and (b) one teacher was on vacation at T1 ($n = 4$). Consequently, valid small-group discussions were available for 470 sessions (T0: $n = 156$, T1: $n = 155$, T2: $n = 159$). At the child level, data on baseline reading comprehension (ELFE) were missing for 41 children at T0 (e.g., due to illness). Because this covariate was a key predictor, primary analyses were restricted to children with observed ELFE scores at T0 ($n = 704$). Seven children who entered the study at the end of the school year were excluded because they did not contribute baseline data. To examine potential selection bias associated with restricting analyses to children with observed ELFE scores, we compared children with and without missing ELFE at T0 on all outcomes assessed at T1 and T2 using mixed-effects models that mirrored the primary model structures. We found no evidence for differences in outcome levels or change over time between groups (all $p > .05$), suggesting that restricting analyses to children with valid ELFE scores at T0 is unlikely to bias the T1–T2 results (Table A2 in the supplemental materials). In total, the dataset comprised 1,999 observations for the discussion outcomes. For the individual reasoning outcomes, 16 occasions lacked written responses, resulting in 1,983 observations.

Analyses

Except for the justification-level measure, all outcomes were counts and were analyzed with generalized linear mixed models (GLMMs) using the glmmTMB package in R (Brooks et al., 2017). Nakagawa's R^2 was computed with the *performance* package (Lüdecke et al., 2021). We report effect sizes as rate ratios (*RRs*), obtained by dividing the intervention group's proportional change from T0 to T2 by the control group's change. The ordinal justification-level outcome was analyzed with a cumulative link mixed model (CLMM; logit link) via the *ordinal* package (Christensen, 2018). R^2 was computed on the latent logit scale (Nakagawa et al., 2017), and effect sizes were reported as odds ratios (*ORs*). For justification level, means were estimated using the *emmeans* package (Wieditz et al., 2024).

For GLMM model selection, we followed Peugh et al. (2021) for multilevel longitudinal analyses of count data. We tested zero inflation with unconditional models that included only linear and quadratic effects of Time. The distribution family [i.e., Poisson, Negative Binomial 1 (nbinom1), Negative Binomial 2 (nbinom2)] was selected based on BIC comparisons of fully conditional models. The best-fitting distributions varied by outcome: Poisson for reasons, moral and group-based justifications; nbinom2 for questions, unreasoned claims, and build-ons; and nbinom1 for total word count and challenges. For the ordinal justification-level outcome, we fitted a CLMM and used a staged model-building strategy analogous to linear mixed-effects modeling.

Random intercepts were specified at the child and teacher levels in most of the models, with 2 exceptions where BIC favored omitting one of these intercepts. The group-level intercepts were retained only for build-ons and challenges. Random slopes for Time at the teacher level were included only when they improved model fit. For the collective reasoning outcomes, we included a fixed quadratic Time term (Time^2) to more precisely estimate intervention effects over time (Yang et al., 2024). For the individual reasoning outcomes, the quadratic Time term did not improve fit and was omitted. We did not include random quadratic Time effects, as these are not identifiable with only 3 waves.

In all models on collective reasoning, an offset term was included to adjust for discussion length and prevent duration bias. The offset accounted for deviations from the prescribed 15-minute discussion time. We included story (Story A vs. B, Story A vs. C), children's reading comprehension, and gender as covariates in all models. Using dummy coding with Story A as the reference category, this specification adjusts for differences across all 3 stories. The contrast between Stories B and C is implicitly captured as a linear combination of the 2 story coefficients. To keep models simple, we included only covariates significant for one outcome group. Therefore, group size and group home language were included only for collective reasoning outcomes, and grade only for individual reasoning outcomes. Gender, grade, reading

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

comprehension, group home language, and group size were grand-mean centered. For all other predictors, 0 denotes the reference (e.g., control, pretest). The analyses are confirmatory, given directional hypotheses for all outcomes. However, analyses of quadratic effects are exploratory.

Results

Observed means are reported in Table 1 (see Table A3, for correlations among all outcomes). Model estimates and fit indices appear in Table 2 for collective reasoning outcomes and in Table 3 for individual reasoning outcomes. In the text, we report estimated mean changes from T0 to T2 for intervention and control groups. Changes across all measurement points are in Table A4. Estimated means for collective reasoning outcomes refer to a child's output in a 15-min discussion. Figures 1 and 2 show the intervention effects based on the estimated means.

Collective reasoning

Active engagement: words and questions

Children in the intervention group showed larger gains in word production and child-initiated questions, as indicated by significant linear Time × Condition interactions (Table 2). From T0 to T2, word production increased by +79.75 words in the intervention group, compared with +15.68 words in the control group (Figure 1A). From T0 to T2, the intervention group's increase in expected word production was 1.37 times that of the control group (*RR* = 1.37). The frequency of children's questions increased by 0.42 from T0 to T2 in the intervention group, compared with 0.02 in the control group (*RR* = 8.62; Figure 1B). Significant quadratic Time × Condition interactions for word and question frequencies indicated stronger gains in the intervention group from T0 to T1 than from T1 to T2.

Children talked less in Story C than in Story A. They also asked fewer questions in Stories B and C than in Story A. Higher reading comprehension predicted more words and more questions. Children in larger groups talked less, while those in groups with a higher proportion of children with a non-Swiss German home language asked more questions.

Argumentation: reasons, unreasoned claims, challenges, build-ons

The linear Time × Condition interaction was significant for all argumentation outcomes (Table 2). Compared with children from the control group, children from the intervention group showed larger increases in reasons, challenges, and build-ons, and a larger decrease in unreasoned claims (Figures 1C–1F). From T0 to T2, children in the intervention group changed by +1.13 reasons, +0.34 challenges, +2.82 build-ons, and −1.72 unreasoned claims. The control group changed only modestly (+0.07, +0.03, +0.21, +0.12). Relative to the control group, *RR*s were 1.74 for reasons, 4.53 for challenges, 3.68 for build-ons, and 0.49 for unreasoned claims (indicating a larger decrease). A quadratic interaction effect also emerged for reasons, challenges, and build-ons, indicating that gains in the intervention group (vs. control) were stronger from T0 to T1 than from T1 to T2. For unreasoned claims, the nonlinear interaction suggested that most of the reduction occurred early (T0–T1), followed by a weaker change from T1 to T2.

Children challenged less often in Stories B and C than in Story A and were less likely to build on peers' contributions in Story B compared with Story A. Higher reading comprehension predicted all 4 argumentation outcomes. Boys provided fewer reasons and more unreasoned claims. Children in larger groups provided fewer reasons and made fewer unreasoned claims. In groups with a higher proportion of children with a non-Swiss German home language, children produced more unreasoned claims.

Individual reasoning

Moral justifications

A significant linear Time × Condition interaction effect was found, indicating that children in the intervention group showed

Table 1 Means (min–max) of collective and individual reasoning outcomes by condition and measurement point, with justification level reported as category proportions.

	Control group			Intervention group		
Outcome	**T0**	**T1**	**T2**	**T0**	**T1**	**T2**
Collective reasoning						
Words	189.61 (5–749)	189.03 (3–1186)	215.02 (2–849)	190.99 (1–663)	261.27 (3–1,133)	295.69 (6–1,059)
Questions	0.02 (0–1)	0.02 (0–1)	0.06 (0–2)	0.03 (0–2)	0.56 (0–13)	0.67 (0–12)
Reasons	1.79 (0–11)	1.84 (0–8)	1.87 (0–9)	1.70 (0–9)	3.07 (0–15)	3.05 (0–11)
Unreasoned claims	2.97 (0–14)	2.81 (0–17)	3.14 (0–18)	4.04 (0–21)	1.95 (0–12)	2.08 (0–14)
Challenges	0.05 (0–3)	0.07 (0–2)	0.09 (0–3)	0.10 (0–3)	0.56 (0–14)	0.63 (0–9)
Build-ons	0.69 (0–10)	1.13 (0–16)	1.15 (0–8)	1.03 (0–14)	3.71 (0–26)	4.38 (0–21)
Individual reasoning						
Moral justifications	0.40 (0–3)	0.51 (0–3)	0.46 (0–3)	0.36 (0–3)	0.52 (0–3)	0.68 (0–4)
Group-based justifications	0.13 (0–2)	0.20 (0–2)	0.26 (0–3)	0.14 (0–2)	0.28 (0–3)	0.32 (0–3)
Justification level[a]	0.60/0.35/0.06	0.50/0.40/0.10	0.50/0.41/0.09	0.62/0.34/0.04	0.47/0.42/0.11	0.35/0.54/0.11

Note.
[a]Justification level is reported as category proportions (0/1/2 = no justification/single-domain justification/dual-domain justification).

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

Table 2 Fixed effects, random effects, and model fit for models predicting collective reasoning outcomes.

	Active engagement		Argumentation			
Parameter	Words	Questions	Reasons	Unreasoned claims	Challenges	Build-ons
Fixed effects						
Intercept	5.08*** (0.05)	−3.87*** (0.40)	0.34*** (0.09)	0.97*** (0.10)	−3.66*** (0.43)	−0.80*** (0.19)
Story: B vs. A	−0.04 (0.02)	−0.48** (0.16)	0.05 (0.04)	−0.02 (0.04)	−0.36* (0.14)	−0.12* (0.05)
Story: C vs. A	−0.05* (0.02)	−0.38* (0.15)	0.05 (0.04)	−0.02 (0.04)	−0.42** (0.14)	−0.10 (0.05)
Reading comprehension	0.17*** (0.03)	0.33*** (0.08)	0.20*** (0.03)	0.08** (0.02)	0.25** (0.08)	0.17*** (0.03)
Gender[a]	−0.02 (0.05)	−0.16 (0.14)	−0.12* (0.05)	0.21*** (0.04)	0.18 (0.12)	0.01 (0.05)
Group size	−0.19*** (0.02)	−0.04 (0.11)	−0.15*** (0.03)	−0.18*** (0.03)	0.00 (0.12)	−0.01 (0.05)
Group home language	−0.00 (0.10)	0.80* (0.36)	−0.10 (0.11)	0.21* (0.11)	0.19 (0.41)	−0.28 (0.19)
Time[b]	0.05 (0.07)	−0.44 (0.95)	0.20 (0.12)	−0.02 (0.11)	1.08 (0.68)	0.79*** (0.20)
Time^2 (quadratic)	−0.00 (0.03)	0.45 (0.44)	−0.09 (0.05)	0.02 (0.05)	−0.34 (0.30)	−0.29*** (0.08)
Condition[c]	0.00 (0.07)	0.25 (0.51)	−0.05 (0.12)	0.31* (0.13)	0.60 (0.51)	0.40 (0.25)
Time × Condition	0.42*** (0.09)	4.79*** (1.08)	0.83*** (0.16)	−0.91*** (0.16)	2.07** (0.76)	1.58*** (0.25)
Time^2 × Condition	−0.13** (0.04)	−1.86*** (0.48)	−0.28*** (0.07)	0.28*** (0.07)	−0.66* (0.33)	−0.47*** (0.10)
Random effects						
Child (intercept)	0.30	0.36	0.22	0.10	0.30	0.14
Group (intercept)	—	—	—	—	0.44	0.08
Teacher (intercept)	0.02	0.36	0.10	0.17	1.14	0.58
Teacher (slope)	0.00	—	0.05	0.07	0.45	0.18
Teacher (corr intercept-slope)	−0.48	—	−0.85	−0.73	−0.85	−0.56
Model summary						
BIC	24,531.6	1,850.6	7,352.2	8,161.1	2,129.1	6,524.8
Marginal/conditional R^2	0.18/0.93	0.48/0.63	0.16/0.52	0.13/0.46	0.26/0.54	0.33/0.66

Note. Standard errors for fixed effect coefficients are reported in brackets. Random effects are reported as variances.
[a]Gender was coded 0 = girls, 1 = boys, and then grand-mean centered.
[b]0 = pretest, 1 = midtest, 2 = posttest.
[c]0 = control group, 1 = intervention group.
Grade was included only in the individual reasoning models; it was excluded from the collective reasoning models because it was not significant in any of those models.
*$p < .05$.
**$p < .01$.
***$p < .001$

a greater increase in moral justifications than those in the control group (Table 3). At T2 compared to T0, children in the intervention group provided 0.32 more moral justifications, whereas children in the control group showed a smaller increase (+0.06 justifications; see Figure 2A). The rate of change in the intervention group was 1.67 times that of the control group (*RR*). Additionally, children provided fewer moral justifications in Story B than in Story A. Children with higher reading comprehension used more moral justifications than children with lower reading comprehension. Boys provided fewer moral justifications than girls.

Group-based justifications

No intervention effect was found for group-based justifications (Table 3; Figure 2B). However, children referred more frequently to group-based justifications in Stories B and C than in Story A. Group-based justifications increased from fourth to fifth grade and over time. Children with higher reading comprehension produced more group-based justifications, and boys produced fewer than girls.

Justification level

Before fitting the mixed models, we tested the proportional-odds (PO) assumption in a baseline cumulative link model (CLM; logit link) for the focal predictors; no PO violations were detected for Time, Condition, or their Time × Condition interaction, suggesting comparable intervention effects across the transitions (0→1 and 1→2).

The final model revealed a significant Time × Condition interaction, indicating that the intervention group shifted more strongly toward higher justification levels over time than the control group (Table 3). The corresponding interaction effect was *OR* = 1.43 per time unit, which translates to a T0→T2 (2-unit) *OR* of 2.05 (Figure 2C).

With time, children were more likely to use higher justification levels. Higher reading comprehension was associated with higher justification levels. Responses by girls and responses to Story C also tended to fall into higher levels than responses by boys and responses to Story A (PO was violated for gender and for story; see the online supplemental materials, Table A5, for the partial-proportional-odds robustness test).

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

Table 3 Fixed effects, random effects, and model fit for individual reasoning outcomes.

Parameter	Moral justifications	Group-based justifications	Justification level
Fixed effects			
Intercept	−0.82*** (0.10)	−3.00*** (0.19)	—
Story: B vs. A	−0.62*** (0.09)	1.28*** (0.14)	0.16 (0.11)
Story: C vs. A	0.14 (0.07)	0.90*** (0.15)	0.55*** (0.11)
Reading comprehension	0.18*** (0.04)	0.13* (0.06)	0.28*** (0.06)
Gender[a]	−0.39*** (0.07)	−0.35** (0.11)	−0.59*** (0.10)
Grade[b]	0.02 (0.10)	0.28* (0.11)	0.18 (0.13)
Time[c]	0.07 (0.06)	0.32*** (0.09)	0.23** (0.08)
Condition[d]	−0.14 (0.13)	0.18 (0.18)	−0.14 (0.18)
Time × Condition	0.26*** (0.08)	0.05 (0.12)	0.36** (0.11)
Random effects			
Child (intercept)	—	0.24	0.17
Teacher (intercept)	0.06	—	0.10
Model summary			
BIC	3,497.6	2,178.9	3,555.2
Marginal/Conditional R^2	0.15/0.20	0.19/0.29	0.09/0.16

Note. Intercepts are reported for count models (moral and group-based justifications). For justification level (0 = no justification; 1 = single-domain justification; 2 = dual-domain justification), the model includes category thresholds instead of an intercept (0|1 = 0.60, 1|2 = 3.22; logit scale).
[a]0 = girls, 1 = boys.
[b]0 = fourth grade, 1 = fifth grade.
[c]0 = pretest, 1 = midtest, 2 = posttest.
[d]0 = control group, 1 = intervention group.
Gender and grade were grand-mean centered.
*$p < .05$.
**$p < .01$.
***$p < .001$.

Discussion

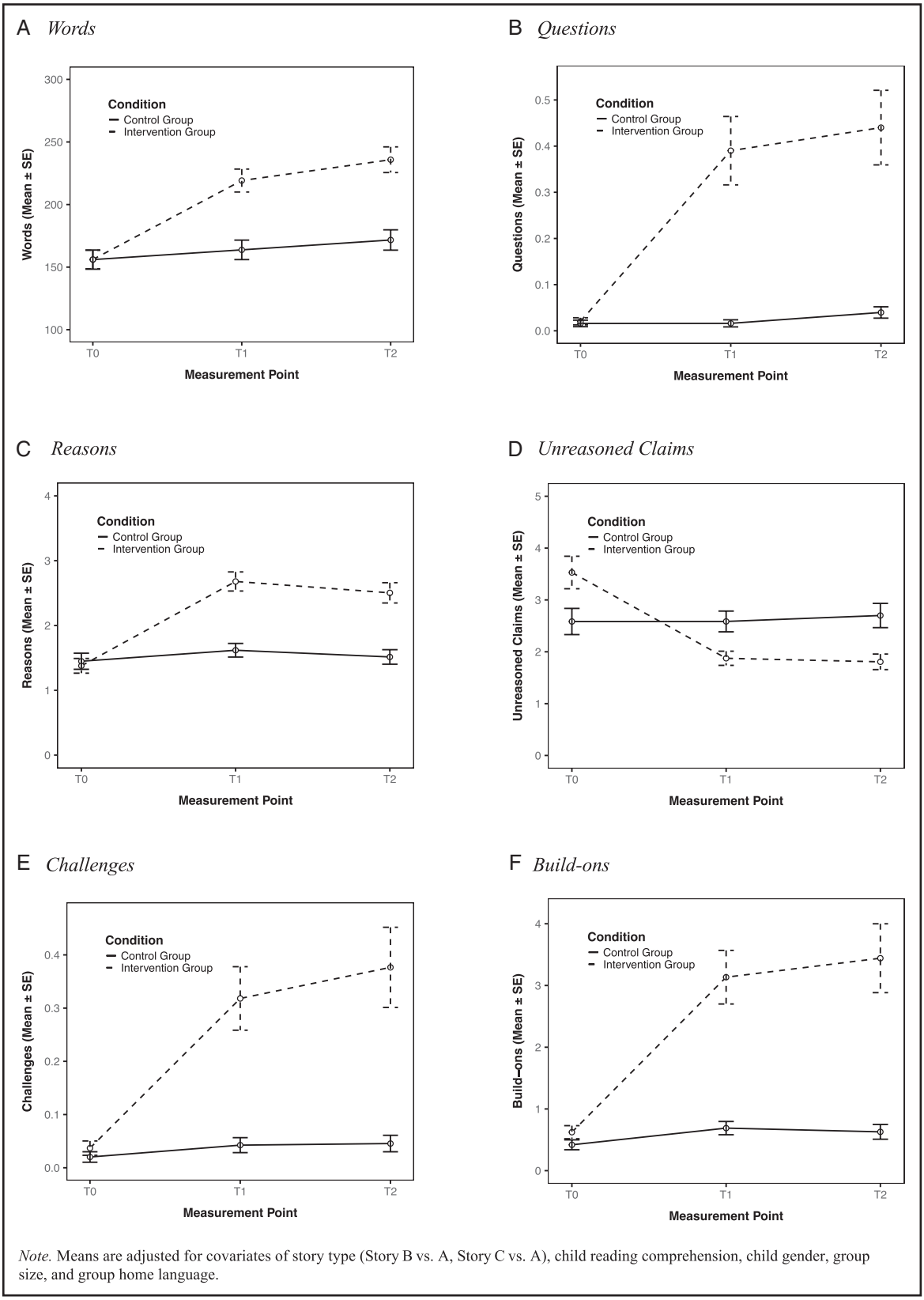
The study demonstrated that the intervention effectively enhanced both collective and individual moral reasoning in children. Regarding collective reasoning, children in the intervention group showed greater increases in their active engagement during discussions, as indicated by the amount of talk and their control over the content through spontaneous questioning. Moreover, these children showed greater gains in argumentative practices, such as reasoning, challenging, and building on others' contributions. At the same time, less effective argumentative practices declined, as reflected in fewer unreasoned claims among children in the intervention group. The intervention led to significant gains in moral, but not group-based, justifications, suggesting that children in the intervention group prioritized moral over group-based concerns when group loyalty entailed high moral costs. Finally, children in the intervention group were more likely to progress from giving no justification to providing single- and then dual-domain justifications.

While developmental research has emphasized the importance of shared book reading for young children's social-cognitive development (Tompkins et al., 2018), relatively little attention has been given to the role of school-based literary education in children's moral development. Drawing on a critical-analytic approach to social and moral learning through narrative fiction (Gasser et al., 2022), our intervention used narrative fiction as a developmental context to foster children's dialogic practices and moral reasoning.

A recent intervention study by Killen et al. (2022) employed teacher-facilitated discussions of hypothetical scenarios about exclusion based on group membership (e.g., ethnicity, gender) to promote children's moral reasoning about intergroup conflicts. Similar to our approach, teachers facilitated critical-analytic engagement with the stories by encouraging children to relate the stories to their own lives, connecting different stories, and considering moral and group-related perspectives. A randomized controlled trial with third- to fifth-graders showed significant improvements in social reasoning through these peer discussions. Our study extends this line of research by showing that developmentally stimulating discussions can also be effectively integrated into the regular language arts curriculum using narrative fiction. Argumentative discussions show positive effects on many academic outcomes (Murphy et al., 2009; Reznitskaya & Gregory, 2013). Integrating moral education into language arts thus has the potential to simultaneously stimulate children's social and academic development.

The positive findings for both collective and individual moral reasoning suggest that the social practices in the discussions serve as a model for individual moral reasoning. From a social-constructivist perspective, dialogic discussions serve as a vehicle for internal cognitive development. Tomasello (2014) argues that human thinking is inherently shaped by joint intentionality and shared meaning-making. Through dialogic engagement, children learn to justify their claims and consider alternative viewpoints—practices that, over time, become internalized as dialogic modes of thinking. Reznitskaya and Gregory (2013) explain this transition with reference to "argumentative stratagems"—such as reasoning, countering, or integrating opposing views—which initially emerge as social practices in discourse but gradually become cognitive strategies that learners apply independently. Dialogic

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

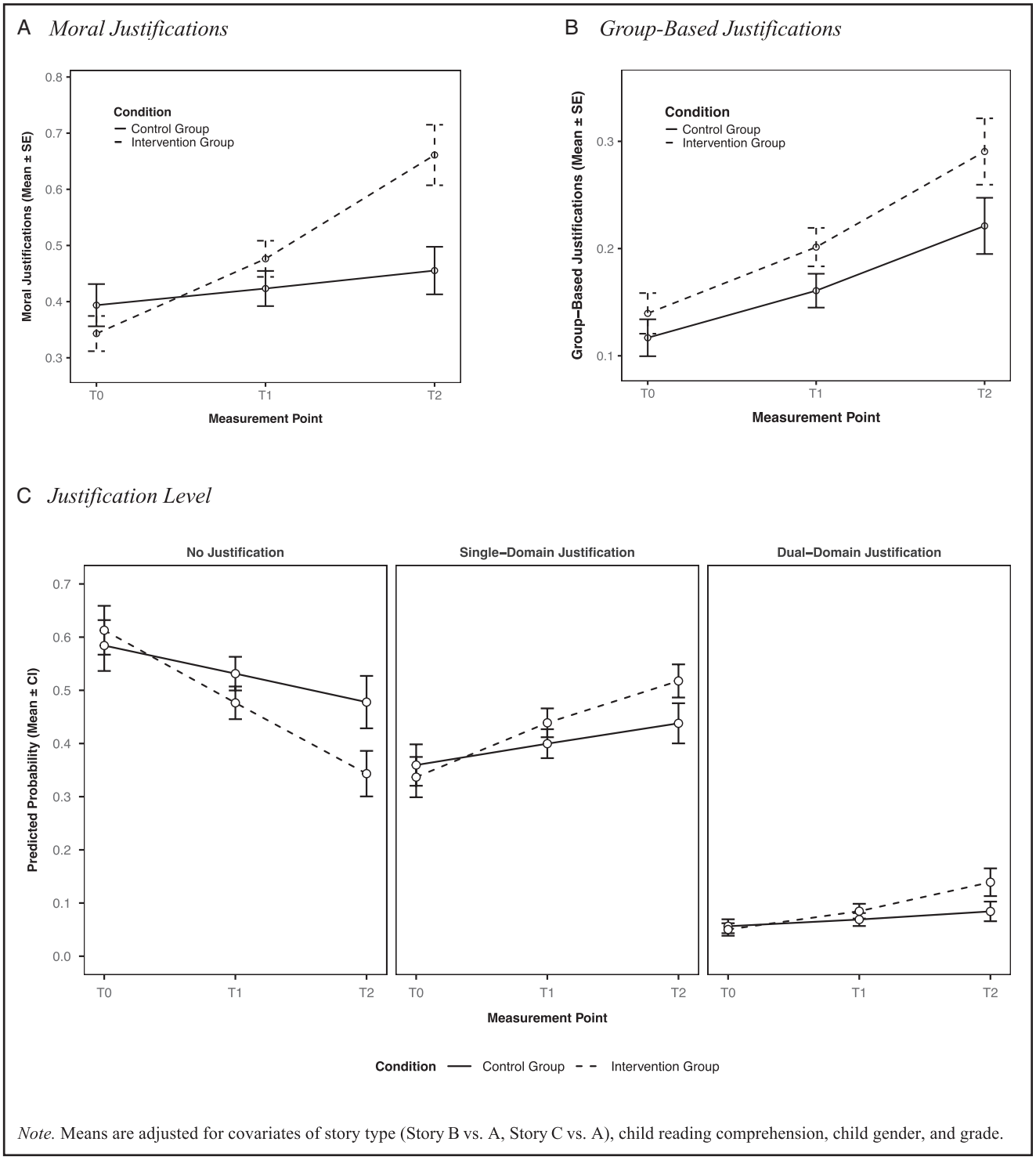


Note. Means are adjusted for covariates of story type (Story B vs. A, Story C vs. A), child reading comprehension, child gender, group size, and group home language.

Figure 1 Intervention effects on collective reasoning outcomes (model-predicted means).

participation thus functions as a model for higher-order reasoning: what is first co-constructed in discussions transfers into autonomous forms of reflective thinking. Consistent with these theoretical accounts, research by Deanna Kuhn and colleagues showed that when adolescents can take multiple perspectives during discussions on controversial issues, they subsequently shift from single- to multiple- and integrated-perspective arguments in their argumentative writing (Kuhn & Crowell, 2011).

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026



Note. Means are adjusted for covariates of story type (Story B vs. A, Story C vs. A), child reading comprehension, child gender, and grade.

Figure 2 Intervention effects on individual reasoning outcomes (model-predicted means).

Although our intervention likewise produced gains in justification level (no justification → single-domain → dual-domain), dual-domain responses were rare (Figure 2C), and children showed a less pronounced increase in considering multiple perspectives than adolescents in Kuhn and Crowell's study. Children might have shown stronger such responses during the individual writing task if the prompt had required them to generate counterarguments to their own positions (Wilkinson et al., 2023). However, it is also plausible that the open format of literary discussions is less effective for cultivating multi-perspective moral argumentation than dialogic programs that provide an explicit instructional format requiring students to anticipate opposing reasons, practice counters and rebuttals in dialogues and debates, and receive structured feedback (Kuhn & Crowell, 2011).

We did not find an intervention effect for group-based reasoning. As the professional development for teachers included

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

analyses of children's books through the lens of social domain theory, it is possible that the higher frequencies of moral (but not group-based) responses reflect demand characteristics (i.e., participants' awareness of the study's aims). In middle childhood, children show increasing concern with moral self-presentation (Rutland et al., 2005). Accordingly, some intervention children may have voiced more moral justifications because they perceived such responses as socially desirable. While we acknowledge this possibility, it is worth noting that the program adopted a minimal normative stance. In training, teachers were guided to use the books to discuss fairness, group dynamics, and personal issues (e.g., privacy and individual rights). We emphasized that narrative fiction's contribution lies less in evaluating acts as right or wrong than in engaging with multifaceted social conflicts and exploring the circumstances under which wrongdoing occurs. Accordingly, teachers were encouraged to adopt a devil's advocate role when children moralized too quickly. Central to this literary stance is a balance between empathic exploration of the contextual and mitigating circumstances of wrongdoing and critical distance in moral judgments—what Nussbaum (1985) calls "finely aware and richly responsible." This stance contrasts with the more strongly normative approaches of many SEL programs.

Interestingly, the analysis revealed a quadratic pattern in the development of collective reasoning in the intervention group: improvements were more pronounced during the first half of the intervention. This early acceleration may reflect the initial QT mini-lessons, which explicitly introduced effective discourse strategies. Thus, explicit instruction may foster children's participation in moral discussions. While such instruction is well established in reading comprehension interventions, it is rarely applied to higher-order reasoning. Building on the QT framework, our approach suggests that metacognitive components of moral discourse can be systematically supported through structured, explicit teaching.

From a critical perspective, the quadratic effect might also reflect the limited impact of the intervention during the second half of the school year. Across all collective reasoning outcomes, gains from mid- to posttest were small. This suggests that additional practice alone may be insufficient to maintain early gains. Unlike smaller-scale studies that offer extensive coaching to a few teachers (e.g., Murphy et al., 2022), our large-sample design limited the dosage of support. Given the persistence of monologic discourse patterns in classrooms (Reznitskaya & Gregory, 2013), ongoing professional development may be needed to achieve stronger and more lasting improvements in children's collective reasoning skills.

While we did not find intervention effects on group-based justifications, age and grade clearly predicted increases in this outcome. This finding aligns with the SRD research showing that older children and adolescents increasingly consider group loyalty and identity in their inclusion and exclusion decisions (Killen & Rutland, 2011; Mulvey & Killen, 2015). While SRD emphasizes the role of moral reasoning in opposing unfair group practices, it has been less explicit about children's capacity to articulate and justify their concerns in front of peers. Yet, argumentation skills in real-life social and moral conflicts may be an important resource for resisting negative peer influence. For example, Allen et al. (2006) showed that adolescents who could effectively manage disagreements with a close friend while maintaining positive relationships were less susceptible to negative peer influence. Accordingly, interventions should provide opportunities for children and adolescents to practice reasoning and argumentation in peer interactions as a means of supporting autonomy in moral conflicts involving group dynamics (Killen et al., 2022).

We also found significant effects of several control variables. Children's reading comprehension was positively related to all outcomes, because both the written and the oral tasks required understanding the dilemmas. Thus, children with reading difficulties may have struggled to effectively engage in the discussion and writing tasks due to limited text understanding. In groups with many children speaking a language other than Swiss German at home, children made more unreasoned claims but also asked more questions. Overall, children in linguistically diverse groups performed similarly to those in linguistically homogeneous groups. Finally, story type significantly affected discussion quality, underscoring the role of the text in shaping literary transactions (Gasser et al., 2022). Consistent with moral developmental theories that emphasize contextual variation in children's social reasoning (Killen & Rutland, 2011; Krettenauer & Johnston, 2011; Peplak et al., 2017), children differed in how often they referenced moral versus group-based concerns across the 3 story contexts.

Because the sample was drawn from German-speaking Swiss elementary schools and sociocultural variation was captured mainly via migration-related indicators (child born abroad, non–Swiss German home language), findings should be generalized primarily to comparable European school contexts. Switzerland has one of the highest foreign-born shares among OECD countries, so our data come from a comparatively migration-dense context (OECD, 2023).

The study is not without limitations. First, our coding system for the discussions focused on formal characteristics, rather than on the content. While moral developmental research has traditionally emphasized formal aspects of children's dialogic reasoning in moral dilemma discussions, more recent work has highlighted the importance of analyzing the content of such discussions as well (Recchia et al., 2014). For a more comprehensive assessment, future research should therefore address both formal and content-related aspects of moral dilemma discourse. Second, the study lacks a follow-up measurement. An intervention study on the 4Rs literary program demonstrated delayed effects on children's social-cognitive development (Jones et al., 2011). Therefore, it remains unclear whether the effects of the current intervention were sustained over time or whether children's moral reasoning had increased or diminished after the intervention concluded. Third, we did not include an active control group. As a result, the intervention and control conditions were not fully matched in nonspecific features such as adult attention, structured activity time, and novelty, which can themselves influence children's engagement in discussions and writing. This design choice therefore limits causal specificity: any observed differences may reflect, in part, these nonspecific factors rather than the intervention's targeted components. However, the theory-consistent gains we observed in collective and individual reasoning are unlikely to be attributable to generic attention alone. Moreover, our study used a large sample (mean $n \approx 80$ in shared-reading intervention studies; Murphy et al., 2009; Noble et al., 2019), enabling modeling of longitudinal trajectories and child-, small-group-, and teacher-level effects. Because implementing an equivalently intensive alternative at this scale is challenging, large-scale studies on moral or dialogic learning have often omitted active controls (Grøver et al., 2020; Killen et al., 2022; Wilkinson et al., 2023). Despite these constraints,

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

future large-scale trials should include an active control condition matched on attention and time on task and preregister manipulation checks of hypothesized mechanisms (alongside analysis plans) to more definitively rule out nonspecific effects. Fourth, the study design may be susceptible to experimenter bias because key roles were not fully independent: the coaching and rating of implementation fidelity involved an author of the manuscript, and 2 authors conducted the teacher training. This overlap could have influenced teachers' implementation and/or fidelity ratings (e.g., through expectancy effects or more favorable interpretations) and should therefore be considered when interpreting the findings.

Despite these limitations, the study offers important insights into how small-group discussions about narrative fiction can support children's dialogic skills and moral reasoning. These competencies can help children navigate conflicts between group loyalty and morality and stand up to peer pressure when needed (Allen et al., 2006). Narrative fiction not only fosters language development but also provides a rich context for reflecting on complex moral dilemmas that children often face in everyday life. In a time when the loudest voices often prevail, it is essential that children learn to reason together, weigh different perspectives, and make thoughtful, balanced decisions. This study shows that literary discussions can be a powerful space for cultivating these essential skills.

Supplementary material

Supplementary material is available at *Child Development* online.

Data availability

The data and analytic code necessary to reproduce the analyses presented here are publicly accessible via OSF: https://doi.org/10.17605/OSF.IO/X59DB. The materials necessary to attempt to replicate the findings presented here are not publicly accessible. Instruments are available from the first author upon request. The intervention materials were originally developed in German; sample materials and the full manual are available at https://phbern.ch/skill. The analyses presented here were not preregistered.

Author contributions

Luciano Gasser (Conceptualization, Formal analysis, Funding acquisition, Project administration, Writing—original draft [lead], Data curation, Methodology [equal]), David Preisig (Formal analysis [supporting], Writing—review & editing [equal]), Yvonne Dammert (Conceptualization, Methodology, Project administration [equal], Resources [equal], Writing—review & editing [supporting]), Anna Frei (Investigation, Resources [equal], Writing—review & editing [supporting]), Sara Egger (Data curation, Project administration, Resources [equal], Writing—review & editing [supporting]), and P. Karen Murphy (Conceptualization, Writing—review & editing [supporting])

Funding

This study was supported by the Swiss National Science Foundation [100019_200841]. The development of the intervention used in this study was supported by Stiftung Mercator Schweiz [2017-0026] and by Fachstelle für Rassismusbekämpfung (FRB), with project support from Stiftung éducation21 [17.R.066].

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

We thank the undergraduate students for their help with video coding and the participating teachers and students. We also acknowledge Neda Bodenmann, Hanspeter Inauen, and Denise Krummenacher for their coaching support for participating classrooms.

References

Allen, J. P., Porter, M. R., & McFarland, F. C. (2006). Leaders and followers in adolescent close friendships: Susceptibility to peer influence as a predictor of risky behavior, friendship instability, and depression. *Development and Psychopathology*, *18*, 155–172. https://doi.org/10.1017/S0954579406060093

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

Berkowitz, M. W., & Gibbs, J. C. (1985). The process of moral conflict resolution and moral development. *New Directions for Child Development*, *29*, 71–84. https://doi.org/10.1002/cd.23219852907

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*, 378–400. https://doi.org/10.32614/RJ-2017-066

Bruner, J. S. (2011). *Actual minds, possible worlds*. Harvard University Press.

Carroll, N. (2000). Art and ethical criticism: An overview of recent directions of research. *Ethics*, *110*, 350–387. https://doi.org/10.1086/233273

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

Christensen, R. H. B. (2018). *Cumulative link models for ordinal regression with the R package ordinal*. Retrieved November 20, 2025, from https://cran.uni-muenster.de/web/packages/ordinal/vignettes/clm_article.pdf.

Cooley, S., & Killen, M. (2015). Children's evaluations of resource allocation in the context of group norms. *Developmental Psychology*, *51*, 554–563. https://doi.org/10.1037/a0038796

Dahl, A., Martinez, M. G. S., Baxley, C. P., & Waltzer, T. (2022). Early moral development. In M. Killen, & J. G. Smetana (Eds.), *Handbook of moral development* (3rd ed, pp. 135–152). Routledge.

Dammert, Y., & Gasser, L. (2025). *SKiLL— – Sprachliche und soziale Kompetenzen mit Kinderliteratur fördern: Das Handbuch zum Förderprogramm [SKiLL—promoting language and social skills through narrative fiction: Intervention handbook]*. Bern University of Teacher Education. Retrieved April 20, 2026, from https://www.phbern.ch/forschung/projekte/soziale-und-sprachliche-kompetenzen-ueber-kinderliteratur-foerdern-0/beispiele-der-unterrichtsmaterialien.

Damon, W., & Killen, M. (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly*, *28*, 347–367.

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

Ellis, W. E., & Zarbatany, L. (2017). Understanding processes of peer clique influence in late childhood and early adolescence. *Child Development Perspectives*, *11*, 227–232. https://doi.org/10.1111/cdep.12248

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*, 1445–1475. https://doi.org/10.1007/s10648-022-09667-4

Gasser, L., Malti, T., & Buholzer, A. (2014). Swiss children's moral and psychological judgments about inclusion and exclusion of children with disabilities. *Child Development*, *85*, 532–548. https://doi.org/10.1111/cdev.12124

Grøver, V., Rydland, V., Gustafsson, J.-E., & Snow, C. E. (2020). Shared book reading in preschool supports bilingual children's second-language learning: A cluster-randomized trial. *Child Development*, *91*, 2192–2210. https://doi.org/10.1111/cdev.13348

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*, 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*, 377–380. https://doi.org/10.1126/science.1239918

Killen, M., Burkholder, A. R., D'Esterre, A. P., Sims, R. N., Glidden, J., Yee, K. M., Luken Raz, K. V., Elenbaas, L., Rizzo, M. T., Woodward, B., Samuelson, A., Sweet, T. M., & Stapleton, L. M. (2022). Testing the effectiveness of the developing inclusive youth program: A multisite randomized control trial. *Child Development*, *93*, 732–750. https://doi.org/10.1111/cdev.13785

Killen, M., & Dahl, A. (2021). Moral reasoning enables developmental and societal change. *Perspectives on Psychological Science*, *16*, 1209–1225. https://doi.org/10.1177/1745691620964076

Killen, M., Pisacane, K., Lee-Kim, J., & Ardila-Rey, A. (2001). Fairness or stereotypes? Young children's priorities when evaluating group exclusion and inclusion. *Developmental Psychology*, *37*, 587–596. https://doi.org/10.1037/0012-1649.37.5.587

Killen, M., & Rutland, A. (2011). *Children and social exclusion: Morality, prejudice, and group identity*. Wiley-Blackwell.

Krettenauer, T., & Johnston, M. (2011). Positively versus negatively charged moral emotion expectancies in adolescence: The role of situational context and the developing moral self. *The British Journal of Developmental Psychology*, *29*, 475–488. https://doi.org/10.1348/026151010X508083

Kruger, A. C. (1992). The effect of peer and adult-child transactive discussions on moral reasoning. *Merrill-Palmer Quarterly*, *38*, 191–211.

Kruger, A. C., & Tomasello, M. (1986). Transactive discussions with peers and adults. *Developmental Psychology*, *22*, 681–685. https://doi.org/10.1037/0012-1649.22.5.681

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, *22*, 545–552. https://doi.org/10.1177/0956797611402512

LaFontana, K. M., & Cillessen, A. H. N. (2010). Developmental changes in the priority of perceived status in childhood and adolescence. *Social Development*, *19*, 130–147. https://doi.org/10.1111/j.1467-9507.2008.00522.x

Lee, K., Talwar, V., McCarthy, A., Ross, I., Evans, A., & Arruda, C. (2014). Can classic moral stories promote honesty in children? *Psychological Science*, *25*, 1630–1636. https://doi.org/10.1177/0956797614536401

Lenhard, W., & Lenhard, A. (2011). *Berechnung des Lesbarkeitsindex LIX nach Björnson* [Software]. Retrieved March 5, 2024, from https://www.psychometrica.de/lix.html.

Lenhard, W., Lenhard, A., & Schneider, W. (2017). *ELFE II— - Ein Leseverständnistest für Erst- bis Siebtklässler [ELFE II—A Reading comprehension test for grade one to seven]*. Hogrefe.

Li, L., & Tomasello, M. (2022). Disagreement, justification, and equitable moral judgments: A brief training study. *Journal of Experimental Child Psychology*, *223*, 105494. https://doi.org/10.1016/j.jecp.2022.105494

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. *The British Journal of Educational Psychology*, *92*, 131–154. https://doi.org/10.1111/bjep.12442

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*, 3139. https://doi.org/10.21105/joss.03139

Malti, T., Killen, M., & Gasser, L. (2012). Social judgments and emotion attributions about exclusion in Switzerland. *Child Development*, *83*, 697–711. https://doi.org/10.1111/j.1467-8624.2011.01705.x

Mammen, M., Köymen, B., & Tomasello, M. (2019). Children's reasoning with peers and parents about moral dilemmas. *Developmental Psychology*, *55*, 2324–2335. https://doi.org/10.1037/dev0000807

Mammen, M., & Paulus, M. (2023). The communicative nature of moral development: A theoretical framework on the emergence of moral reasoning in social interactions. *Cognitive Development*, *66*, 101336. https://doi.org/10.1016/j.cogdev.2023.101336

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

Mar, R. A., Tackett, J. L., & Moore, C. (2010). Exposure to media and theory-of-mind development in preschoolers. *Cognitive Development*, *25*, 69–78. https://doi.org/10.1016/j.cogdev.2009.11.002

Mulvey, K. L. (2016). Children's reasoning about social exclusion: Balancing many factors. *Child Development Perspectives*, *10*, 22–27. https://doi.org/10.1111/cdep.12157

Mulvey, K. L., Hitti, A., & Killen, M. (2023). When fairness and group loyalty conflict: Social exclusion, prejudice, and bias in adolescence and early adulthood. In L. J. Crockett, G. Carlo, & J. E. Schulenberg (Eds.), *APA Handbook of adolescent and young adult development* (pp. 159–173). American Psychological Association.

Mulvey, K. L., & Killen, M. (2015). Challenging gender stereotypes: Resistance and exclusion. *Child Development*, *86*, 681–694. https://doi.org/10.1111/cdev.12317

Murphy, P. K. (Ed.) (2018). *Classroom discussions in education*. Routledge.

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

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026

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

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*, 740–764. https://doi.org/10.1037/a0015576

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society, Interface*, *14*, 20170213. https://doi.org/10.1098/rsif.2017.0213

Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C., Gobet, F., & Pine, J. (2019). The impact of shared book reading on children's language skills: A meta-analysis. *Educational Research Review*, *28*, 100290. https://doi.org/10.1016/j.edurev.2019.100290

Nucci, L. (2016). Recovering the role of reasoning in moral education to address inequity and social justice. *Journal of Moral Education*, *45*, 291–307. https://doi.org/10.1080/03057240.2016.1167027

Nucci, L., Creane, M. W., & Powers, D. W. (2015). Integrating moral and social development within middle school social studies: A social cognitive domain approach. *Journal of Moral Education*, *44*, 479–496. https://doi.org/10.1080/03057240.2015.1087391

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

Nussbaum, M. C. (1985). "Finely aware and richly responsible": Moral attention and the moral task of literature. *The Journal of Philosophy*, *82*, 516–529. https://doi.org/10.2307/2026358

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*, 135–198. https://doi.org/10.1207/S15326950DP3502_3

OECD. (2023). *International migration outlook 2023*. OECD Publishing.

Peplak, J., Song, J. H., Colasante, T., & Malti, T. (2017). "Only you can play with me!" Children's inclusive decision making, reasoning, and emotions based on peers' gender and behavior problems. *Journal of Experimental Child Psychology*, *162*, 134–148. https://doi.org/10.1016/j.jecp.2017.04.019

Peugh, J. L., Beal, S. J., McGrady, M. E., Toland, M. D., & Mara, C. (2021). Analyzing discontinuities in longitudinal count data: A multilevel generalized linear mixed model. *Psychological Methods*, *26*, 375–397. https://doi.org/10.1037/met0000347

Piaget, J. (1932). *The moral judgment of the child*. Free Press.

Recchia, H. E., Wainryb, C., Bourne, S., & Pasupathi, M. (2014). The construction of moral agency in mother–child conversations about helping and hurting across childhood and adolescence. *Developmental Psychology*, *50*, 34–44. https://doi.org/10.1037/a0033492

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

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development*, *76*, 451–466. https://doi.org/10.1111/j.1467-8624.2005.00856.x

Theimer, C. E., Killen, M., & Stangor, C. (2001). Young children's evaluations of exclusion in gender-stereotypic peer contexts. *Developmental Psychology*, *37*, 18–27. https://doi.org/10.1037/0012-1649.37.1.18

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

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*, 223–246. https://doi.org/10.1111/sode.12280

van der Kleij, S. W., Devine, R. T., Shapiro, L. R., Ricketts, J., & Apperly, I. (2025). Longitudinal development of theory of mind in adolescence and its associations with fiction Reading experience. *Developmental Psychology*, *61*, 1126–1135. https://doi.org/10.1037/dev0001965

Vezzali, L., Stathi, S., Giovannini, D., Capozza, D., & Trifiletti, E. (2015). The greatest magic of Harry Potter: Reducing prejudice. *Journal of Applied Social Psychology*, *45*, 105–121. https://doi.org/10.1111/jasp.12279

Walker, L. J., Hennig, K. H., & Krettenauer, T. (2000). Parent and peer contexts for children's moral reasoning development. *Child Development*, *71*, 1033–1048. https://doi.org/10.1111/1467-8624.00207

Walker, L. J., & Taylor, J. H. (1991). Family interactions and the development of moral reasoning. *Child Development*, *62*, 264–283. https://doi.org/10.2307/1131002

Wieditz, J., Miller, C., Scholand, J., & Nemeth, M. (2024). A brief introduction on latent variable based ordinal regression models with an application to survey data. *Statistics in Medicine*, *43*, 5618–5634. https://doi.org/10.1002/sim.10208

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

Wimmer, L., Currie, G., Friend, S., Wittwer, J., & Ferguson, H. J. (2024). Cognitive effects and correlates of Reading fiction: Two preregistered multilevel meta-analyses. *Journal of Experimental Psychology: General*, *153*, 1464–1488. https://doi.org/10.1037/xge0001583

Yang, M., Wang, L., & Maxwell, S. E. (2024). The role of a quadratic term in estimating the average treatment effect from longitudinal randomized controlled trials with missing data. *Psychological Methods*. https://doi.org/10.1037/met0000709

Yoo, H. N., & Smetana, J. G. (2022). Distinctions between moral and conventional judgments from early to middle childhood: A meta-analysis of social domain theory research. *Developmental Psychology*, *58*, 874–889. https://doi.org/10.1037/dev0001330

Yüksel, AŞ, Palmer, S. B., Argyri, E. K., & Rutland, A. (2025). Children's and adolescents' expectations, evaluations and reasoning about a bystander who challenges social exclusion within intragroup and intergroup peer contexts. *The British Journal of Developmental Psychology*, *43*, 67–81. https://doi.org/10.1111/bjdp.12522

Downloaded from https://academic.oup.com/chidev/advance-article/doi/10.1093/chidev/aacag089/8684900 by guest on 24 June 2026