Journal of Educational Psychology
2009, Vol. 101, No. 3, 740–764

0022-0663/09/$12.00 DOI: 10.1037/a0015576

Examining the Effects of Classroom Discussion on Students' Comprehension of Text: A Meta-Analysis

P. Karen Murphy
The Pennsylvania State University

Ian A. G. Wilkinson
Ohio State University

Anna O. Soter
Ohio State University

Maeghan N. Hennessey
University of Oklahoma

John F. Alexander
Young Scholars Charter School of Central Pennsylvania

The role of classroom discussions in comprehension and learning has been the focus of investigations since the early 1960s. Despite this long history, no syntheses have quantitatively reviewed the vast body of literature on classroom discussions for their effects on students' comprehension and learning. This comprehensive meta-analysis of empirical studies was conducted to examine evidence of the effects of classroom discussion on measures of teacher and student talk and on individual student comprehension and critical-thinking and reasoning outcomes. Results revealed that several discussion approaches produced strong increases in the amount of student talk and concomitant reductions in teacher talk, as well as substantial improvements in text comprehension. Few approaches to discussion were effective at increasing students' literal or inferential comprehension and critical thinking and reasoning. Effects were moderated by study design, the nature of the outcome measure, and student academic ability. While the range of ages of participants in the reviewed studies was large, a majority of studies were conducted with students in 4th through 6th grades. Implications for research and practice are discussed.

Keywords: reading comprehension, classroom discussion, text meta-analysis

Tremendous strides have been made in the area of reading comprehension research in recent decades (National Reading Panel, 2000). Whether recent advances in research are reflective of shifts in prevailing education-related policies (e.g., *A Nation at Risk*, National Commission of Excellence in Education, 1983 or No Child Left Behind Act of 2001), trends in funding cycles, or paradigmatic swings in literacy education is not clear. What is clear, however, is that students' reading comprehension scores on the National Assessment of Educational Progress (NAEP) have risen modestly but steadily from 1992 to 2007 in concert with these advances. Indeed, a majority (67%) of American fourth graders are now reading and comprehending at or above the Basic level on the NAEP (Lee, Grigg, & Donahue, 2007). Fourth graders who perform at the Basic level are competent at gleaning the overall meaning of what they read from developmentally appropriate literary and informational texts, can make simple inferences, and can loosely build connections between their lives and the text (Lee et al., 2007). Even more students (70%) attain this level by the eighth grade (Lee et al., 2007).

However, very few American students perform at the Proficient or Advanced levels on the NAEP assessments (Lee et al., 2007). According to data from the 2007 administration of the NAEP, only 25% of fourth graders were "able to demonstrate a strong understanding of the text . . . to extend the ideas in the text by making inferences, drawing conclusions, and making connections to their own experiences," and just 8% were able to "judge texts critically . . . and explain their judgments . . . make generalizations about the point of a story and extend its meaning by integrating personal experiences and other readings" (National Assessment Governing Board, 2007, p. 24). At the eighth grade, the percentage of students performing at the Proficient level was only 27%, and the percentage of students performing at the Advanced level was a mere 2% (Lee et al., 2007). Similarly, ACT (2006) recently reported that the majority of students tested in high school were ill prepared to read content-rich, complex college-level texts. Comprehending uncomplicated texts at such basic levels is insufficient in light of the increasing need for high-level literacy associated with rapid tech-

P. Karen Murphy, Department of Educational Psychology, The Pennsylvania State University; Ian A. G. Wilkinson and Anna O. Soter, School of Teaching and Learning, Ohio State University; Maeghan N. Hennessey, Department of Educational Psychology, University of Oklahoma; John F. Alexander, Young Scholars Charter School of Central Pennsylvania, State College, Pennsylvania.

This research was supported by Grant R305G020075 from the Institute for Education Sciences. We would also like to thank Betsy Becker, Florida State University, for advice and consultation on this meta-analytic review, as well as Makayla Potts for her assistance in manuscript preparation.

Correspondence concerning this article should be addressed to P. Karen Murphy, Department of Educational Psychology, The Pennsylvania State University, 229 CEDAR Bldg., University Park, PA 16802. E-mail: pkm15@psu.edu

nological changes in the 21st century and beyond (Bransford, Brown, & Cocking, 1999).

In response to these concerns, many educators are now directing their attention to *critical literacy*, that is, literacy that goes beyond the simple decoding of text or basic determination of meaning (Callison, 2000; Graves, Juel, & Graves, 2001). Over the last two decades, the term *critical literacy* has been variously defined and understood. For example, some theorists and researchers view critical literacy from the perspective of critical theory, positioning, and power relations (e.g., de Castell & Luke, 1987; Freebody & Luke, 1990). Luke (cited in Jongsma, 1991) stated that practices of critical literacy break down traditional teacher roles and encourage students to talk and write about their experiences, their knowledge, and their opinions on a variety of issues. In the present work, we use the term *critical literacy* as it relates to higher order thinking and critical-reflection on text and discourse. Within such an interpretation of critical literacy, the goal is to help students achieve a high-level comprehension of text, to read beyond a text's surface, and to surpass the acquisition of lower order thinking skills (Chang-Wells & Wells, 1993). The effectiveness of instructional approaches for promoting such critical literacy and high-level comprehension remain largely unexplored.

The focus of the present review was an examination of the effects of using group discussions as a tool for promoting students' high-level comprehension of text (i.e., critical literacy). The term *high-level comprehension* is used to refer to critical, reflective thinking about text. High-level comprehension requires that students engage with text in an epistemic mode to acquire not only knowledge of the topic but also knowledge about how to think about the topic and the capability to reflect on one's own thinking (cf. Chang-Wells & Wells, 1993). Related terms are *literate thinking*, *higher order thinking*, *critical thinking*, and *reasoning*.

Theoretical Underpinnings

The theoretical rationales invoked to explain the role of discussion in promoting students' reading comprehension derive largely from sociocognitive and sociocultural theory. According to Piaget (1928), social interaction is a primary means of promoting individual reasoning. Similarly, Vygotsky (1934/1986) conceived of learning as a culturally embedded and socially meditated process in which discourse plays a primary role in the creation and acquisition of shared meaning making. Essentially, Vygotsky (1978) conceptualized reading and writing as socially constructed higher order psychological processes. Within such a perspective, children develop reading skills and abilities through authentic participation in a literacy-rich environment and are apprenticed into the literate community by more knowledgeable others (e.g., parents, teachers, or more capable peers). Tharp and Gallimore (1988) stated:

> A key feature of this [sociocultural perspective] emergent view of human development is that higher order functions develop out of social interaction. Vygotsky argues that a child's development cannot be understood by a study of the individual. We must also examine the external social world in which that individual life has developed. . . . Through participation in activities that require cognitive and communicative functions, children are drawn into the use of these functions in ways that nurture and "scaffold" them. (pp. 6–7)

According to Wertsch, Del Rio, and Alvarez (1995), when students interact with others in a group in deep and meaningful ways, the outcomes or results that are produced are beyond the abilities and dispositions of the individual students who compose the group. Students bring to the discussion unique social and cultural values, background experiences, and prior knowledge and assumptions. Through the interactions, learners incorporate ways of thinking and behaving that foster the knowledge, skills, and dispositions needed to support transfer to other situations that require independent problem solving (Anderson et al., 2001; Hatano, 1993). In the context of discussion, students make public their perspectives on issues arising from the text, consider alternative perspectives proposed by peers, and attempt to reconcile conflicts among opposing points of view.

The dialogic process is negotiated and sustained through interpretations of text, high-level reasoning, and standards of interaction that govern group behavior. Similarly, Bakhtin's (1981) work suggested that reasoning is inherently dialogical; that is, one's reasoning is necessarily a response to what has been said or experienced as well as an anticipation of what will be said in response. The underlying presupposition is that reasoning is dynamic and relational. It is not so much that one cannot reason individually but rather that reasoning is mediated by prior experiences and the anticipation of future social experiences. According to Anderson et al. (2001, p. 2), "thinkers must hear several voices within their own heads representing different perspectives on the issue. The ability and disposition to take more than one perspective arise from participating in discussions with others who hold different perspectives" (see also Reznitskaya et al., 2001).

Indeed, talk is an essential feature of social–constructivist pedagogy, and researchers are beginning to understand those aspects of it that can be relied upon as either agents or signals of student learning. Select empirical and theoretical research shows that the quality of classroom talk is closely connected to the quality of student problem solving, understanding, and learning (e.g., Mercer, 1995, 2002; Nystrand, Gamoran, Kachur, & Prendergast, 1997; Wegerif, Mercer, & Dawes, 1999). This research indicates that there is sufficient reliability in language use to enable us to make valid inferences about the productiveness of talk for student learning (see also Anderson, Chinn, Chang, Waggoner, & Yi, 1997; Applebee, Langer, Nystrand, & Gamoran, 2003). The discourse–learning nexus is complex and highly situated, and the mapping between discourse and learning is imperfect. Nevertheless, empirical research in this area has begun to reach a level of maturity at which those aspects of discourse and attendant classroom norms that shape student learning can be identified and examined.

Classroom Discussions About Text

Research has identified a number of approaches to conducting intellectually stimulating discussions that appear to be effective in promoting high-level responses to text in elementary as well as high school settings (e.g., Collaborative Reasoning, Philosophy for Children, Questioning the Author, Instructional Conversations, or Book Club). These approaches serve various purposes depending on the goals teachers set for their students: to adopt a critical or analytic stance (e.g., Anderson et al., 1997), to acquire information (e.g., Beck, McKeown, Sinatra, & Loxterman, 1991), or to respond to literature on an aesthetic level (e.g., Raphael, Gavelek, & Daniels, 1998). Discussion approaches that give prominence to

interrogating or querying the text in search of the underlying arguments, assumptions, world views, or beliefs align with what Wade, Thompson, and Watkins (1994) describe as a *critical–analytic stance*. Such a stance encourages a discussion in which the reader's querying mind is engaged, prompting him or her to ask questions, and promoting a more subjective, critical response toward the text. By comparison, approaches that give prominence to knowledge acquisition are often conceptualized as being more efferent in nature (Chinn, Anderson, & Waggoner, 2001). We define an *efferent stance* as a text-focused response in which discussion gives prominence to reading to acquire and retrieve particular information. In this stance, the focus is on "the ideas, information, directions, conclusions to be retained, used, or acted on after the reading event" (Rosenblatt, 1978, p. 27).

Early in our review of relevant literature, we took issue with the term *aesthetic* as applied to discussions we observed because, in our judgment, few actually attained a truly aesthetic response (Rosenblatt, 1978). Instead, we chose to use the term *expressive stance* to describe a reader-focused response (Jakobson, 1987). In this stance, discussion gives prominence to the reader's affective response to the text or the reader's own spontaneous, emotive connection to all aspects of the textual experience (Soter & Rudge, 2005). In addition to giving prominence to a particular stance through the established goals of the discussion, each approach is also characterized by some type of instructional frame that describes the moves of the teacher, the role of the text, specific metacognitive strategies, and benchmarks of success. Although the aims of these approaches are not identical, most purport to help students develop the skills and abilities to discuss text, consider different perspectives, and provide support for arguments.

The review of research reported herein is part of a larger project whose purpose was to identify converging evidence on the use of group discussions to promote high-level comprehension of text and to advance understanding of how teachers can implement discussions and assess their effects in ways that are sensitive to instructional goals. As mentioned previously, the specific objective of this meta-analysis was to examine evidence of the effects of different approaches to conducting group discussions, including estimation of the magnitude of effects. To qualify for inclusion in our review of research, an approach to discussion had to demonstrate consistency of application and have an established place in educational research or practice on the basis of a record of peer-reviewed, empirical research conducted in the last three decades. The nine approaches we identified were Collaborative Reasoning (CR; Anderson, Chinn, Waggoner, & Nguyen, 1998), Paideia Seminar (PS; Billings & Fitzgerald, 2002), Philosophy for Children (P4C; Sharp, 1995), Instructional Conversations (IC; Goldenberg, 1993), Junior Great Books Shared Inquiry (JGB; Great Books Foundation, 1987), Questioning the Author (QtA; Beck & McKeown, 2006; McKeown & Beck, 1990), Book Club (BC; Raphael & McMahon, 1994), Grand Conversations (GC; Eeds & Wells, 1989), and Literature Circles (LC; Short & Pierce, 1990). These approaches serve various purposes depending on the goals teachers set for their students: to adopt a critical–analytic stance, to acquire information on an efferent level, or to respond to literature on an aesthetic or expressive level.

For the purposes of this review and our larger project, we categorized the approaches into one of the three stances toward text. Our categorization of the various approaches was based on the stance given prominence in the discussion as evident in the primary goals enacted by proponents or developers of the approach in the published literature. Specifically, within the critical–analytic stance, we included Collaborative Reasoning, Paideia Seminar, and Philosophy for Children because each of these approaches has an enacted goal of querying and interrogating the underlying arguments and evidence presented in the text. By comparison, we included Instructional Conversations, Junior Great Books Shared Inquiry, and Questioning the Author within the efferent stance on the basis of their enacted primary goal of searching the text for information. Finally, within the expressive stance, we included Book Club, Grand Conversations, and Literature Circles as each of these approaches encourages students to live through the text and gives prominence to highly emotive and affective responses to the text.

In making these categorizations, we fully acknowledge that some of the aforementioned approaches espouse secondary or tertiary goals that align with multiple stances. For example, many of the approaches within the critical–analytic stance also encourage readers to gather relevant information from the text and to emotively respond to the text, so as to enable the reader to take a stand or make a claim about a given issue and to support it with evidence. For this reason, as will become apparent later, we used *approach* rather than stance as our primary unit of analysis in the meta-analysis. Nonetheless, we found it useful in the larger project and in the present review to categorize the approaches into stances as a mechanism for interpreting trends across the various approaches.

Moreover, in our judgment, regardless of the primary goal as evident in the published literature, all of the discussion approaches have potential to promote students' high-level thinking and comprehension of text. For example, Collaborative Reasoning (CR) (Anderson et al., 1998) uses discussion to foster students' critical reading and thinking about text as part of reading instruction. CR is representative of approaches encouraging a critical–analytic stance toward text and is premised on the idea of reasoned argumentation as a model for critical thinking (Chinn et al., 2001). Drawing from the work of McGee (1992) on discourse as a characteristic way of talking and thinking, and the work of Kuhn (1993) and Toulmin (1958) on argumentation, CR aims to encourage students to use reasoned discourse as a means for choosing among alternative perspectives on an issue. CR discussions foster conversation among students that draws on personal experiences, background knowledge, and text for interpretive support. In the format of CR, the teacher poses a central question deliberately chosen to evoke varying points of view. Students adopt a position on the issue and are encouraged to generate reasons that support their position. Using the text, as well as personal experiences and background knowledge, students proceed to evaluate reasons and evidence, to consider alternative points of view, and to challenge the arguments of others (Waggoner, Chinn, Yi, & Anderson, 1995). Anderson et al. (1998) analyzed transcripts from CR discussions and recitation lessons to examine the quantity and quality of student participation. The authors found that students from the CR group engaged in argumentation more frequently, challenged the opinions and thoughts of others, responded to challenges, and provided evidence or information from the text in defense of their arguments more frequently than did those involved in recitation. The nature of the discourse also changed as evidenced by in-

creased student participation, decreased teacher talk, and less teacher control of topic.

Another approach to group discussion is Questioning the Author (QtA; Beck & McKeown, 2006; Beck, McKeown, Hamilton, & Kucan, 1998). QtA grew out of investigations of text characteristics that hinder or support comprehension (Black & Bern, 1981; Frederiksen, 1981; Trabasso, Secco, & van den Broek, 1984) as well as the developers' own research into problems experienced by readers when they interact with ideas presented in textbooks (Beck et al., 1991; Beck, McKeown, & Gromoll, 1989; McKeown & Beck, 1990). QtA is representative of approaches encouraging an efferent stance toward text and is premised strongly on a constructivist view of reading (e.g., Anderson, 1977). QtA aims to engage students deeply in the process of making meaning from text and encourages students to question the author's position as an "expert" and to read against the text. In fact, students are taught to "depose" the authority of the author and to realize that authors are fallible and that the difficulties they encounter when reading challenging texts are not necessarily attributable to their own inadequacies. Teachers guide students through a text, assisting them in the construction of meaning through the use of *queries*, that is, questions designed to help students focus on the meaning of an author's words. Almasi, McKeown, and Beck (1996) investigated the nature of reading engagement during QtA using data from field notes, videotapes of group meetings, and student and teacher interviews. Teachers and students viewed videotapes of classroom discussions and rated them on interpretation and levels of engagement as well as on text processing. From these data, the authors found several factors they believe contributed to increased student engagement: the classroom climate, the use of interpretive tools, and use of textual evidence to clarify and verify issues regarding characters' motives, actions, or story elements. However, the researchers did not assess whether the discussions affected students' comprehension of text.

Yet another approach to group discussion is Book Club (BC; Raphael et al., 1998; Raphael & McMahon, 1994), which we have categorized as reflective of an expressive stance toward text. This approach is based on reader-response theory (Iser, 1978; Rosenblatt, 1978), research regarding how readers derive meaning from text (Duffy, Roehler, & Mason, 1984; Pearson & Johnson, 1978), and research concerning discourse patterns (Cazden, 1988). BC is strongly rooted in a sociocultural perspective (Bakhtin, 1986; McGee, 1992; Mead, 1934) and reflects Vygotskian notions of the role of language, the zone of proximal development, and internalization (Vygotsky, 1978). BC structure comprises four elements: reading, writing, discussion, and instruction. Students read a text (reading), record their written responses to the text in journals (writing), and then use these responses to engage in small-group discussion, otherwise known as Book Clubs (BCs). The instructional element, which occurs throughout the program, is used for a variety of purposes such as discussing elements of a story, modeling strategies for students, and discussing rules for appropriate group behavior (e.g., Kong & Fitch, 2002/2003). Students also participate in *community share*, a whole-class discussion that provides opportunities for students to share information from BC discussions and enhances their awareness of issues concerning the thematic unit or historical background of the reading. Participation in BCs has been linked to increased vocabulary, as well as the metacognitive strategies of self-questioning, summarizing, and using strategies, particularly with struggling or ethnically diverse students (Kong & Fitch, 2002/2003). Raphael and colleagues (1998) have also shown that BCs enhance students' conversational competence including their ability to sustain topics and themes as well as taking the perspective of others over the course of a 3-week unit.

As a case in point, Goatley, Brock, and Raphael (1995) investigated the experiences of a diverse group of students involved in a regular education literature-based reading program. Their goal was to document the meaning-making experiences of fifth-grade students ($n = 5$) participating in a BC. Particularly important in this study is the fact that 3 of the students had previously received reading instruction in a pull-out program. Data were collected over a 3-week period and included interviews; written questionnaires on the students' perceptions of their roles in BC; field notes by the researchers; audio, video, and transcripts of BC discussions; and students' written work in response to the literature. Analyses of the data revealed that participation in BCs led to increased participation and talk on the part of all students. Moreover, the BC setting provided struggling learners with opportunities and support for learning strategies and skills related to textual interpretation and meaning construction as evidenced in students' comprehension strategy use and the ability to scaffold each other's meaning constructions. Students were able to use their reading logs for personal reactions and use these personal reactions in discussions. Evidence from discussions also showed that both peers and teachers can serve as the "more knowledgeable other" in discussions about text.

Despite differences across the purposes and goals of various approaches, evidence suggests that discussions about and around text have the potential to increase student comprehension, metacognition, critical thinking, and reasoning, as well as students' ability to state and support arguments (e.g., Bird, 1984; Reznitskaya et al., 2001). To date, however, no syntheses have quantitatively reviewed the vast body of literature on classroom discussions to examine the effects of the various approaches on students' comprehension and learning. That is, there exists no thorough quantitative examination of evidence from extant research of the effectiveness of prominent approaches to discussion. As such, the purpose of this meta-analysis was to examine the effects of the nine identified discussion approaches on students' high-level comprehension of text. We conducted a comprehensive meta-analysis of empirical studies that provided evidence of the effects of discussion on measures of teacher/student talk and on individual student comprehension and critical thinking and reasoning outcomes. Our specific purposes in this review were as follows:

- identify empirical studies that provided evidence of the effects of discussion on measures of teacher/student talk and on individual student comprehension and reasoning outcomes;
- document trends within variables that frame the reviewed studies (e.g., publications trends);
- examine effects of discussion on high-level comprehension relative to key study features including design, outcome measure, academic ability, or the number of weeks of discussion;
- examine effects of discussion on high-level comprehension by primary stance toward the text (i.e., critical–analytic, efferent, and expressive); and

model the effects of discussion on high-level comprehension using fixed-effects and random-effects models to explain variability across studies.

In addressing these specific purposes, we hoped to build a road map that individuals can use to better understand the vast discussion literature generally, as well as the effectiveness of various discussion approaches, in particular.

Method

Search Parameters

Our first step was to conduct a broad search of the literature in order to establish a beginning pool of writings from which the final body of relevant works would be chosen. To do so, we conducted exhaustive reviews of the literatures on discussion practices as they relate to the promotion of students' high-level thinking and comprehension of text by carrying out systematic searches of five major databases in the social sciences (i.e., ERIC, Education Abstracts, PsycINFO, Social Sciences Citation Index, and Digital Dissertations), keyed on the names of researchers who have played major roles in the conceptualization of a given approach (e.g., Goldenberg, 1993, or Tharp & Gallimore, 1988, for Instructional Conversation, or Raphael & McMahon, 1994, for Book Club) and titles of the approaches (e.g., Grand Conversation or Junior Great Books Shared Inquiry). We also investigated secondary citations, other printed sources, and associated Web sites. This broad search resulted in 1,001 references.

Over a 3-year period, each reference pertinent to the approaches was summarized in a highly customized EndNote library (Thomson Reuters, Carlsbad, CA) using a common set of fields. This procedure resulted in the references being initially classified into four levels. References were labeled as *Level 1* if they directly pertained to one of our nine targeted approaches (i.e., Collaborative Reasoning, Philosophy for Children, Paideia Seminar, Questioning the Author, Instructional Conversation, Junior Great Books Shared Inquiry, Literature Circles, Grand Conversation, and Book Club). The label *Level 2* was assigned to references that described other empirical research on the role of discussion in promoting students' comprehension, learning, or thinking (e.g., reviews of literature). *Level 3* references elaborated on the theoretical underpinnings of discussion as a means of promoting learning and comprehension, while references were classified as *Level 4* if they provided information on methodological tools that might be useful in understanding students' unfolding interpretations of discussion or in assessing the quality of their interpretation. Most important for the current analysis were those references classified as Level 1.

Criteria for Inclusion

To distill the research literature on classroom discussions for the meta-analysis, we established several criteria a priori to examine all Level-1 references. First, to be included in the meta-analysis, a document must be a report of an empirical study. Our initial search of the literature on classroom discussions resulted in 298 relevant Level-1 documents. Of those documents, only 127 were classified as reports of empirical studies. Second, it was imperative that the studies present quantitative data and that effect sizes could be calculated from the data presented. A number of studies were removed at this point because they either did not present quantitative data ($n = 56$) or because effect sizes could not be calculated on the basis of the data reported ($n = 11$). As a case in point, Ellis (1999) conducted a study on improving comprehension through Junior Great Books Shared Inquiry discussions in third and sixth grades. Unfortunately, only quartile scores were reported in the study, so effect sizes could not be calculated.

Given that our focus of the meta-analysis was the role of classroom discussions in promoting comprehension of text, it was also essential that the studies report effects of discussions about text. Several studies failed to meet this criterion ($n = 3$). For example, Dalton and Sison (1995) used Instructional Conversations with Spanish-speaking middle school students around the topic of problem solving in mathematics—problem solving that did not involve text. Similarly, results from several articles published by Mason (e.g., 1998, 2001) using Collaborative Reasoning in fourth- and fifth-grade science classrooms were not considered in the meta-analysis because the discussions were not about text. For example, in Mason (2001), the fourth-grade students participated in discussions about their observations of mold growing on carrots and cheese.

The final criterion for inclusion pertained to the construct under investigation. It was necessary that the articles contained quantitative data on a construct of interest. Specifically, we were interested in assessment of discourse including measurements of *teacher talk*, *student talk*, and *student-to-student talk*. These constructs were operationalized as the percentage of total utterances made by either the teacher (*teacher talk*) or student (*student talk*) within the context of the discussion. *Teacher talk* consisted of teacher utterances such as coaching or prompting, asking questions of the students, and commenting on student utterances. *Student talk* consisted of student utterances that contributed to the discussion, such as answering questions from either the teacher or other students, asking questions of classmates, or contributing opinions about the text. *Student-to-student talk* was defined as the proportion of consecutive student utterances about the discussion that were made to other students without prompting or coaching from the teacher.

We were also interested in a number of individual student outcomes. Among these outcomes were various forms of comprehension, including *text-explicit comprehension* (i.e., comprehension requiring information that is explicitly stated, usually within a sentence), *text-implicit comprehension* (i.e., comprehension requiring integration of information across sentences, paragraphs, or pages), *scriptally implicit comprehension* (i.e., comprehension requiring considerable use of prior knowledge in combination with information in text), and *general or unspecified comprehension*, in which the nature of the comprehension was unclear (Pearson & Johnson, 1978). We were also interested in individual outcome measures pertaining to *critical thinking and reasoning* (i.e., reasoned, reflective thinking that is focused on deciding what to believe or do, drawing inferences or conclusions; Ennis, 1987), *argumentation* (i.e., taking a position on an issue and arguing for that position on the basis of evidence), and *meta-cognition* (i.e., measurement of students' understanding of their own thinking). Studies that failed to measure a construct of interest were not included in the meta-analysis ($n = 10$).

Many studies that met the first two criteria failed to measure one of the aforementioned constructs of interest in the current meta-

analysis. For example, Patthey-Chavez and Clare (1996) conducted an empirical study on the use of Instructional Conversations with transitional bilingual students as a mechanism for promoting writing in a second language. The authors provided data on constructs such as fluency, syntax, and lexical variety that were not a focus of the present meta-analysis but did not assess a construct being investigated in the current review (e.g., comprehension or metacognition).

It is also important to note that several documents ($n = 3$) were reports of the same study (e.g., reported in a dissertation and refereed journal article). For example, Billings' (1999) unpublished doctoral dissertation is a report of a study of the use of Paideia Seminars with high school students, and the same study emerged in our review as an article published in the *American Educational Research Journal* (Billings & Fitzgerald, 2002). In such cases, the document reporting the most relevant data was retained for the meta-analysis.

In addition, there were two studies that met all the aforementioned criteria but were not included in the analyses (i.e., Anderson et al., 1997; Blum, Lipsett, & Yocom, 2002). Although effect sizes were calculated for each study, nuances in the way the data were presented inhibited their inclusion in the analyses. Specifically, Anderson et al. (1997) failed to present sample sizes, which made it impossible to weight and unbias the calculated effect sizes. Blum et al. (2002) provided *F* statistics for both pretest and posttest outcomes, but there were significant differences at pretest that were not controlled for at posttest, which compromised the viability of the effects sizes. Moreover, we were not able to control for these differences in a secondary analysis on the basis of data provided in the study. The employment of these criteria and the culling of Anderson et al. (1997) and Blum et al. (2002) resulted in a final pool of 42 documents upon which all subsequent analyses were conducted. Each of the documents meeting all criteria for inclusion is described in Table 1.

Although our focus in the present review is on high-level comprehension, it is important to reiterate that approaches to classroom discussion have been developed for a variety of reasons or purposes and are characterized by specific goals. Necessarily, investigations with a particular approach focus on assessment outcomes associated with the primary and secondary goals of the approach—outcomes that may or may not be associated with reading comprehension. Arguably, the design of studies and assessments is also influenced by a number of competing factors that may include predominant research paradigms (e.g., qualitative or quantitative methods), funding initiatives, or prevailing social ills or agendas (e.g., failing schools). For example, Fleener (1994) investigated the nature of the roles and levels of engagement that seventh- and eighth-grade students took on during Paedia Seminars but did not measure student comprehension in relation to the roles or level of engagement. This study was excluded from the present analysis. The exclusion of such studies in the present review is simply reflective of our comprehension focus and should in no way suggest to the reader that such approaches or studies are irrelevant to instruction or student learning.

Coding

All studies were coded using a 58-item coding manual. The manual was developed iteratively over a period of several months using one quantitative study from each discussion approach as a guide to item development, as well as a number of resources on conducting meta-analyses (e.g., Lipsey & Wilson, 2001). Major iterations of the manual were reviewed by an expert in the field of meta-analysis. Among the data collected from each study were type of publication, sample characteristics (e.g., number of students and teachers, age, predominant ethnicity, socioeconomic status, demographic area, or academic ability), research design characteristics (e.g., single group/multiple group or unit of assignment to condition), the description of the discussion (e.g., approach or prediscussion activity), characteristics of any subsamples in the study (e.g., ability level or English-language proficiency), dependent measure characteristics (e.g., group discourse or individual outcome), and effect size data.

All articles were coded using a Visual Basic interface (Microsoft Corp., Redmond, WA) with drop-down boxes and blank fields that enabled the raters to enter data directly into a Microsoft Access database. The interface minimized opportunities for disagreements between coders by placing constraints on data entry depending on the nature of the study. For instance, when single-group design was selected, the interface automatically blocked or modified the remaining available boxes for data entry. In addition, some of the simpler effect size calculations (e.g., using means and standard deviations) were automated in the interface. The programming features incorporated into the interface decreased the amount of time required to code a given article and likely reduced coding and calculation errors.

Prior to coding, raters were trained by two of the authors in consultation with the meta-analysis expert, who previously had reviewed the coding manual. All training was conducted using one study from each of the nine approaches. Specifically, the authors and meta-analysis expert explained and provided examples for each of the 58 items in the manual to the raters. Subsequently, all four individuals independently coded a single article and shared their respective codings. Any disagreements were discussed and resolved. Training continued until an 80% agreement criterion was reached among the two authors and the two raters. Having reached criterion, all studies were independently coded into separate databases by the two trained raters so that we could establish stable interpretations of the results presented in each study. Coding disagreements were resolved in weekly meetings with P. Karen Murphy. Logs were kept of all disagreements, and a third database was created containing coding representing the consensus judgment of P. Karen Murphy and the two trained raters. All analyses were performed using this consensus database. Final interrater agreement prior to consensus was 94%.

Data Preparation

A number of guiding principles were established for data preparation prior to analysis. First, all effect sizes were calculated using Hedges' $\hat{g}$. In doing so, effect sizes were unbiased to control for the influence of small samples and weighted so that studies with larger sample sizes had greater influence in the analysis. In essence, Hedges' $\hat{g}$ incorporates sample size by both computing a denominator that includes the sample sizes of the respective standard deviations (i.e., pooled standard deviation; Hedges & Olkin, 1985). To increase the stability of the effect size estimates, we also

(*text continues on page 752*)

Table 1
Summary of Key Features of Articles Included in the Meta-Analysis

Author(s) (year)	Type[a]	Sample description	Approach[b]	Weeks[c]	Design[d]	Data source[e]	Construct[f]	No. of effect sizes[g]	No. of calculations[h]	Major findings
Banks (1987)	T/D	319 students. Majority African American in a suburban area with low SES.	P4C	24	MG	California Achievement Test	Comp	6	6	Students enrolled in Philosophy for Children discussions had higher scores on the Total Language, Total Reading, and Total Mathematics components of the California Achievement Test than their counterparts in traditional instruction.
Beck, McKeown, Sandora, Kucan, & Worthy (1996)	JA	23 students (mean age 10 years); 2 teachers. Majority African American in an urban area with low SES.	QtA	Not given	SG	Teacher questions—retrieve information, construct message, extend discussion, check knowledge, percentage teacher lines of talk, percentage student lines of talk (total), student-initiated questions and comments	TT, ST	2	12	Teachers participating in the discussion shifted their questioning techniques from asking questions to retrieve information to those enabling students to construct and extend meaning. Teachers also changed the way they responded to students' comments so that discussion was promoted. Because of this, student comments became more complex, and the students collaborated more with each other.
Billings (1999)	T/D	18 students (mean age 17 years); 1 teacher. Majority White with high SES, above-average ability.	PS	20	SG	Percentage of total talk turns, percentage of total talk time	TT, ST	2	4	Teachers in Paideia Seminars predominantly used the role of director in discussions. Male students also tended to portray more dominant roles during discussion than did females, who tended to observe more often.
Bird (1984)	T/D	108 students (mean age 11 years); 10 teachers. High SES, above-average ability.	JGB	Not given	MG	Ross test—critical thinking; Worden test—critical reading	CT/R	1	2	Differences were found in critical thinking and critical reading skills between those students who participated in Junior Great Books lessons and those who did not. Positive correlations were found among critical reading, attitudes, and critical thinking skills in those students participating in the discussion activities on a full-time basis.
Biskin, Hoskisson, & Modlin (1976)	JA	30 students (mean age 8 years). Below-average ability.	JGB	Not given	MG	Recall of story details—character, events, plot, theme, total	TE	1	6	Those students who participated in the reflective strategy group performed differently than those participating in the predictive strategy and control groups on their recall of characters, events, and the full story.
Cashman (1977)	T/D	141 students (mean age 10.95 years); 6 teachers.	JGB	20	MG	Reasoning subtest total score	CT/R	3	3	There were significant differences between students participating in Junior Great Books and control groups on verbal meaning and reasoning. After IQ and pretest performance were controlled, girls participating in the discussion approach performed better than boys.
Casper (1964)	T/D	251 students (mean age 11 years); 8 teachers. Urban/suburban area, above-average ability.	JGB	30	MG	Convergent production—picture arrangement, word-group naming, vocabulary completion; divergent production—alternate uses, associational fluency, possible jobs, ideational fluency, names for stories, simile interpretations; cognition—alternate methods, similarities, word classification; evaluation—sentence selection, matched verbal relations, product choice, best word pairs, sentence sense	Comp, TE, TI, CT/R	4	18	After participating in the discussion approach, boys performed better than girls on Guilford's measures of concept naming, originality, and experiential evaluation. The only statistically significant difference between the experimental and control groups was found in a measure of associational fluency.

(*table continues*)

Table 1 (*continued*)

Author(s) (year)	Type[a]	Sample description	Approach[b]	Weeks[c]	Design[d]	Data source[e]	Construct[f]	No. of effect sizes[g]	No. of calculations[h]	Major findings
Castleberry ISD (1996)	TR	Mean age 11 years	JGB	36	MG	Oral communication, reading comprehension, and vocabulary	Comp	1	3	The percentage of students passing the Texas Assessment of Academic Skills reading test who participated in the Junior Great Books program was higher than for students not participating in the program. The authors state that this difference cannot be concluded as definitively due as to the sole influence of the discussion approach.
Chamberlain (1993)	T/D	115 students (mean age 10.67 years); 6 teachers. Majority White in an urban/suburban area with medium SES, above-average ability	P4C	12	MG	New Jersey Test of Reasoning Skills, Ross Test of Higher Cognitive Processes, percentage of critical thinking responses, agree/disagree with another student, lines of teacher discussion, student-to-student responses	TT, SST, CT/R, Arg	6	6	Students in Philosophy for Children discussions scored significantly higher on the New Jersey Test of Reasoning Skills than their counterparts in control discussions, but not on the Ross Test of Higher Cognitive Processes (possibly due to a ceiling effect). Philosophy for Children discussions saw an increase in student-to-student talk, as well as a decrease in teacher talk, while the control discussions did not.
Chesser, Gellalty, & Hale (1997)	JA	Mean age 14 years. Mixed ethnicity	PS	Not given	MG	North Carolina 8th grade Writing Test	SI	3	4	After participating in Paideia Seminars, students evidenced more gains in their writing scores on state standardized tests than those not participating, possibly because many teachers required students to write after discussions.
Chinn, Anderson, & Waggoner (2001)	JA	84 students (mean age 10 years); 4 teachers. Majority White in a suburban/rural area with medium SES, above and below-average ability	CR	4	SG	Teacher and student talk, turns, interjections, interruptions, various kinds of questions, evidence, prediction, explanations, perspectives, elaborations	TT, ST, TE TI, CT/R	5	24	After implementation of Collaborative Reasoning, students were more engaged in discussions and spoke more with fewer interruptions. Implementation was not difficult in general, although it was difficult to shift interpretive authority and control of the topic from the teachers to the students.
Davis, Resta, Davis, & Camacho (2001)	JA	21 students (mean age 10 years), 1 teacher. Majority Hispanic in an urban area with low SES, average ability	LC	10	SG	Percentage of students passing Texas Assessment of Academic Skills reading practice test	Comp	2	2	After implementation of Literature Circles, students improved their ability to discuss literature with their peers, improved their reading performance, and increased their self-motivation toward reading independently.
Echevarria (1995)	JA	5 students (mean age 8.45 years); 1 teacher. Majority Hispanic in an urban area with low SES, below-average ability	IC	24	SG	Student Outcome Measure, student utterances, self-initiated nonscripted contributions to discussion, self-initiated scripted contributions to discussion, literal recall	ST, TE	3	7	Students participating in Instructional Conversations talked more and used greater amounts of academic discourse. They initiated more of their own contributions to the discussion and evidenced better development of concepts. However, there were no differences in students' narratives or in responses to comprehension questions between the treatment and control groups.
Echevarria (1996)	JA	5 students (mean age 8.45 years); 1 teacher. Majority Hispanic in an urban area with low SES, below-average ability	IC	24	SG	Use of academic discourse, student-constructed narratives—story structure, no. and category of propositions, student utterances across two treatments	ST, SI	2	6	Same findings as the previous study.

(*table continues*)

Table 1 (*continued*)

Author(s) (year)	Type[a]	Sample description	Approach[b]	Weeks[c]	Design[d]	Data source[e]	Construct[f]	No. of effect sizes[g]	No. of calculations[h]	Major findings
Farinacci (1998)	JA	24 students (mean age 7.5 years); 1 teacher. Medium SES, average ability	LC	Not given	SG	Student talk/thinking from checklist	ST	1	3	By the end of the implementation of the discussion approach, students were able to carry on conversations without referring to their journals.
Flynn (2002)	T/D	18 students (mean age 9 years); 3 teachers. Majority White with high SES, above and below-average ability	QtA	12	SG	Retelling the story—setting, sequence, problem, world knowledge, opinion, general retelling	TE, CT/R	2	7	Participation in Questioning the Author helped students move their questions from being literal to those that were more inferential. They also increased their standardized test scores.
Geisler (1999)	T/D	10 students (mean age 6 years); 2 teachers. Majority Hispanic in a suburban area, below-average ability	IC	15	MG	Percentage of turns that were student turn, student-initiated statement, shared reading lessons	ST	1	4	Students participating in Instructional Conversations performed better on Oral Language Proficiency exams than those not participating. These students also showed an increase in their amount of talk during discussions.
Goatley, Brock, & Raphael (1995)	JA	5 students (mean age 11 years); 1 teacher. Mixed ethnicity with low SES, average ability	BC	4	SG	Turn taking, infer/explain, interpretation, judging	ST	1	5	Students were able to use their reading logs for personal reactions and use these personal reactions in discussions. Evidence from discussions also shows that both peers and teachers can serve in a role of the "more knowledgeable other" in discussions about text.
Graup (1985)	T/D	32 students (mean age 12 years). Majority White in an urban area, average ability	JGB	3	MG	Comprehension of literary material, quality of literal questions, quality of inferential questions, no. of interpretive statements made	Comp, TE, TI	3	12	Students participating in Junior Great Books discussions evidenced better comprehension than those not in discussion conditions. These students also asked more inferential questions and fewer literal questions than those not participating in discussions.
Heinl (1988)	T/D	30 students (mean age 11 years), 3 teachers. Suburban area	JGB	18	MG	Literal comprehension, inferential comprehension, Iowa Test of Basic Skills by ability	TE, TI	3	3	Lower ability students improved their comprehension of literal questions, but there were still significant differences on inferences between high- and low-ability students after participating in the discussion.
Howard (1992)	T/D	243 students (mean age 13 years); 2 teachers. Rural area, average ability	PS	1	MG	Percentage of student talk, percentage of teacher talk, proportion of students who talked; task-specific knowledge— no. of responses, quality of associations, essay scores	TT, ST, SI	3	6	Students in Paideia Seminars in the two participating schools performed differently. In one school, Paideia Seminar participants differed from other students on the quality and quantity of associations made, but students in another school did not differ on these outcomes.
Junior Great Books (1992)	O	720 students (mean age 9 years); 15 teachers. Mixed ethnicity in an urban/suburban area with medium SES, average ability	JGB	18	MG	Citing evidence from the story	TE	2	2	Students participating in Junior Great Books discussions cited textual evidence more often than control students in discussions and when writing. They also evidenced better reading vocabulary scores on standardized assessments than those students in control conditions.

(*table continues*)

Table 1 (*continued*)

Author(s) (year)	Type[a]	Sample description	Approach[b]	Weeks[c]	Design[d]	Data source[e]	Construct[f]	No. of effect sizes[g]	No. of calculations[h]	Major findings
Kim (2002)	T/D	102 students (mean age 10 years); 6 teachers. Mixed ethnicity with low SES	CR	2	MG	Total argumentative moves per minute, counterargument moves per minute, rebuttal per minute, invitations for a group member to discuss, arguments in writing, counterarguments in writing, rebuttals in writing, student ratings of group monitoring	Arg, Meta	4	21	When metacognitive group monitoring activities are completed, students are better able to exhibit multiple perspectives in the discussion. These metacognitive practices were also transferred to students' essays, resulting in essays reflecting multiple viewpoints.
Kong & Fitch (2002/2003)	JA	25 students (mean age 10.4 years); 1 teacher. Mixed ethnicity in an urban area with low SES	BC	36	SG	Metacomprehension strategy index, Slosson Oral Reading Test (SORT)	Meta	1	1	After participating in Book Club discussions, students increased their reading vocabulary, as well as the metacognitive strategies of self-questioning, summarizing, and using strategies.
Lipman (1975)	TR	40 students (mean age 11 years); 2 teachers. Mixed ethnicity in an urban area, average ability	P4C	9	MG	Iowa Test of Basic Skills	Comp	1	1	After 3 years with traditional instruction, students who had previously been enrolled in Philosophy for Children discussions had increased comprehension scores.
Martin (1998)	JA	12 students (mean age 8 years); 1 teacher. Suburban area	LC	Not given	SG	Prediction scores, Student Discussion Rubric	ST, TI	2	2	Students who participated in Literature Circles improved their abilities to use questioning and predicting strategies, as well as their abilities to connect their ideas to text.
McGee (1992)	JA	37 students (mean age 7 years); 4 teachers.	GC	Not given	SG	Total no. of responses, percentage of interpretive responses before and after interpretive questions	TI	4	12	Students were able to construct simple meanings of text, connect their personal experiences to stories, make predictions and verify their hypotheses, and evaluate and critique text.
McGee, Courtney, & Lomax (1994)	JA	12 students (mean age 7 years); 2 teachers. Urban/suburban area with medium SES, average ability	GC	Not given	SG	No. of teacher moves—facilitator, helper/nudger, responder, literary curator, reader, total no. of teacher moves	TT	2	12	Students were able to take on the roles of facilitators and helpers/nudgers in Grand Conversation discussions.
McKeown, Beck, Kucan, & Sandora (1995)	TR	93 students (mean age 10.23 years); 4 teachers. Mixed ethnicity in a rural area with medium SES	QtA	Not given	SG	Amount of teacher talk, teacher questions—retrieve information, construct message, extend discussion, check knowledge, teacher lines of talk, student lines of talk	TT, ST	2	24	There were differences in the types of questions asked and the amounts of student and teacher talk from baseline to later discussions using the Questioning the Author framework. Students were also better able to monitor their own activity during discussions.
McKeown, Beck, & Sandora (1996)	BC	Mean age 10 years	QtA	Not given	SG	Frequency of verbatim student responses, frequency of local meaning student response, frequency of integration student response, frequency of student comments	TE, TI	2	8	Teachers' questions shifted from retrieving details to extending meaning, and their comments about students' talk shifted from repeating what the student said to extending meaning in the discussion. Student talk increased, as did the student initiation of comments and questions.
Mizerka (1999)	T/D	50 students (mean age 12 years); 2 teachers. Mixed ethnicity in an urban area with low SES, average ability	LC	21	MG	California Achievement Test	Comp	2	2	Although there were no differences in the comprehension scores of students participating in student-led and teacher-led Literature Circles, students in the student-led condition participated in the discussion with much greater frequency. There was no statistically significant difference between reading attitude or the no. of pages read in the two conditions.

(*table continues*)

Table 1 (*continued*)

Author(s) (year)	Type[a]	Sample description	Approach[b]	Weeks[c]	Design[d]	Data source[e]	Construct[f]	No. of effect sizes[g]	No. of calculations[h]	Major findings
Olezza (1999)	T/D	10 students (mean age 17 years); 1 teacher. Mixed ethnicity	IC	Not given	SG	Content of the lesson, essay in response to a prompt, use of specific content lexicon in writing	Comp, SI	2	3	Students with limited English ability were able to draw on prior knowledge and construct meaning with teacher and peer support when participating in Instructional Conversations.
Pitman (1997)	O	29 students (mean age 11.79 years); 2 teachers. Low SES, average ability	LC	3	SG	Questions 10 and 11 of survey	Comp	1	1	After implementation of Literature Circle discussions, students reported that they were able to enhance their reading skills and learn from one another. They also felt they improved their oral and written communication and were able to identify themes within the literature. Researchers observed that the students were more enthusiastic and attentive during the discussions.
Reznitskaya (2002)	T/D	128 students (mean age 10.5 years); 6 teachers. Majority White in a suburban/rural area with high SES	CR	Not given	MG	Mean essay-for scores, mean essay-against scores, no. of written characters in essays, schema-articulation scores, Metropolitan Achievement Test reading scores (pretest only)	Comp, SI, Arg	6	10	Students participating in Collaborative Reasoning discussions were able to provide more supporting reasons in their persuasive essays than those students in control groups. Discussions were not effective at increasing students' recall of a persuasive text.
Reznitskaya, Anderson, McNurlen, Nguyen-Jahiel, Archodidou, & Kim (2001)	JA	115 students (mean age 10.67 years); 6 teachers. Mixed ethnicity in an suburban/rural area with medium SES	CR	5	MG	No. of idea units coded as argumentation, argumentation idea units—arguments, counterarguments, rebuttals, textual information	Arg	1	16	Students in Collaborative Reasoning classrooms gave a greater number of arguments, reasons, alternative perspectives, and story details than those in control classrooms. They were also able to generalize the common components of arguments to their writing.
Sable (1987)	T/D	83 students (mean age 11 years); 4 teachers. Suburban area, above-average ability	JGB	Not given	MG	Cloze scores, open-ended scores	Comp	1	4	No significant differences were found between students participating in Junior Great Books discussions and those who were not on reading comprehension. Experience with the discussion approach also did not equate to significant differences on these measures.
Sandora, Beck, & McKeown (1999)	JA	49 students (mean age 12.51 years); 2 teachers. Majority African American in an urban area with low SES, below-average ability	QtA	4	MG	Recall of story details, open-ended questions, response size to interpretive questions	TE, SI	2	3	Students in Questioning the Author discussions were more likely to give longer recalls of the story, including more complex story details, than those students in Junior Great Books discussions. They were more capable of stating their position and providing supporting evidence.

(*table continues*)

Table 1 (*continued*)

Author(s) (year)	Type[a]	Sample description	Approach[b]	Weeks[c]	Design[d]	Data source[e]	Construct[f]	No. of effect sizes[g]	No. of calculations[h]	Major findings
Saunders & Goldenberg (1998)	BC	27 students (mean age 10 years); 1 teacher. Majority Hispanic in an urban area with low SES, below-average ability	IC	1	MG	Total words, number of T-units, T-unit length, evidence of the tracer concept in writing, references to actions the characters would take, literal comprehension test, reviewing literal details of the story—percentage of lesson utterances, percentage of segment utterances—reviewing literal details of the story, reviewing literal details of the story, what will happen, mean teacher talk, mean student talk	TT, ST, TE, TI, SI	6	11	Students participating in Instructional Conversations were able to gain a higher level of inferential comprehension of the text without sacrificing literal comprehension. These ESL students were also able to engage in meaningful conversation with their peers.
Saunders & Goldenberg (1999)	JA	138 students (mean age 10.6 years); 20 teachers. Majority Hispanic in an urban area, with low SES, below-average ability	IC	1	MG	Factual comprehension, interpretive comprehension, theme explanation, theme exemplification	TE, TI, SI	6	8	Although Instructional Conversations alone were effective in some areas, the combined effects of literature logs and Instructional Conversations helped minority limited-English proficiency students increase various measures of comprehension.
Solomon (1990)	T/D	8 students (mean age 11 years); 1 teacher. Majority African American in a suburban area with low SES, below-average ability	JGB	10	SG	Main idea, cause and effect, making inferences, fact and opinion	TE, TI, CT/R	3	4	Participants increased their test scores, were able to internalize the creative writing process, and exhibited more positive attitudes toward the reading program.
Williams (1999)	JA	27 students (mean age 5.5 years); 1 teacher. Mixed ethnicity in an urban area with low SES	LC	Not given	SG	Percentage of teacher comments accounting for total talk, student comments elicited by teacher prompts	TT, ST	2	2	The percentage of teacher comments dropped from the beginning of the study to the end of implementation of the discussion approach. Student comments in response to teacher prompts also dropped from the beginning to the end of implementation.
Yeazell (1982)	JA	100 students (mean age 11 years); 4 teachers. Average ability	P4C	36	MG	General comprehension gain using Comprehensive Test of Basic Skills	Comp	1	1	The reading comprehension of the students participating in Philosophy for Children increased beyond those in traditional instructional settings. Students of both above- and below-average ability participating in Philosophy for Children achieved gains.

Note. SES =socioeconomic status; ISD =independent school district; ESL = English as a second language.
[a] BC = book chapter; JA = journal article; T/D = thesis or doctoral dissertation; TR = technical report; O = other. [b] CR = Collaborative Reasoning; P4C = Philosophy for Children; PS = Paideia Seminar; QtA = Questioning the Author; IC = Instructional Conversation; JGB = Junior Great Books; LC = Literature Circles; GC = Grand Conversation; BC = Book Club. [c] Weeks = weeks of discussion approach implemented in classroom. [d] MG = multiple-group design; SG = single-group design. [e] Data sources reported by the authors. [f] Construct of interest for meta-analysis: ST = Student Talk; TT = Teacher Talk; SST = Student-to-Student Talk; Comp = General Comprehension; TE = Text-Explicit Comprehension; TI = Text-Implicit Comprehension; SI = Scriptally Implicit Comprehension; CT/R = Critical-Thinking/Reasoning; Arg = Argumentation; Meta = Metacognition. [g] No. of ES used = the number of effect sizes from a given study after averaging effect sizes of the same construct from multiple researcher-made measures. [h] No. of calculations = the total number of effect sizes calculated for a given study prior to averaging.

made the decision to average effect sizes resulting from multiple researcher-made measures of the same construct within the same study. For example, if text-implicit comprehension was measured by a multiple-choice measure and by a constructed-response measure in the same study with the same sample, then the resulting effect sizes were averaged prior to analyses. Differences between the number of effect sizes calculated and the number of effect sizes used in the analysis are reported in Table 1 for each study.

Results

Framing Variables

As an aid in understanding how researchers addressed the role of classroom discussion in promoting comprehension of texts, we first examined what we have labeled as *framing variables* within these studies. Specifically, framing variables denote conditions or contextual factors pertaining to when, with whom, for how long, and in what ways (e.g., the approach) discussions took place in the reviewed studies. One thing we hoped to learn from the examination of these variables was the extent to which the number of studies examining the effects of classroom discussions has varied over the years. We were also interested in examining the nature of the samples used to explore the role of classroom discussions in text comprehension. Among the sampling features of interest were students' ages, socioeconomic statuses, racial/ethnic backgrounds, school settings, and academic ability levels.

Publication trends. The reviewed studies and concomitant data show that there was a noticeable rise in studies addressing our constructs of interest beginning in the latter part of the 1990s (see Table 1). Between 1964 and 1994, only 16 quantitative studies were conducted pertaining to discussion and comprehension, whereas 26 studies were conducted in the period from 1995 to 2002. The greatest number of studies was conducted in 1999 (n = 7). Of these 42 studies, 18 were published as journal articles, 17 were reported in unpublished dissertations or theses, 2 were published as book chapters, 3 as technical reports, and 2 came from other sources, such as ERIC documents. There appears to be little relation between the date a study was reported and the nature or type of report (e.g., journal article vs. technical report). In addition, approximately 70% of these studies were conducted by researchers who played a primary role in the creation of a given approach. Herein we refer to such individuals as proponents or developers of the approach. Only 12 studies were conducted by someone other than the individual(s) responsible for the creation of the approach. Although not directly attributable to a proponent or developer, several of the 12 studies were carried out by students of the proponent(s) or developer(s) (e.g., Collaborative Reasoning, Reznitskaya, 2002) or were conducted within a research center focusing on the approach (e.g., Philosophy for Children, Chamberlain, 1993). What is important about this finding is that it suggests that the majority of studies are conducted by the individuals who created the approach, and it is not clear whether other researchers or teachers could replicate the effects reported by the originators.

Sample characteristics. As is the case with any investigation, the characteristics of the participants can play a central role in the results. Therefore, we felt that sample characteristics were an important framing variable to examine more closely. In Table 1, we describe the sample for each study based on the information reported by the authors. In some cases, minimal information was reported (e.g., McGee, 1992), while other authors offered detailed descriptions of the participants (e.g., Beck, McKeown, Sandora, Kucan, & Worthy, 1996). The number of students participating in any given study ranged from a low of 5 (Echevarria, 1995, 1996; Goatley et al., 1995) to a high of 720 (Junior Great Books Shared Inquiry, 1992). The latter study is a report of a multiple-group design involving students of varied ethnicities from an urban/suburban area, where 15 teachers took part in Junior Great Books Shared Inquiry discussions in small classroom groups. The mean sample size across the 39 studies that reported sample information was 84.28 participants. Although there were variations in sample size across the individual studies, such differences did not affect the results of the meta-analysis. All effect sizes used in the analyses were unbiased and weighted to account for differences in sample size. The age of the participants ranged from 5.5 years (Williams, 1999) to 17 years (Billings, 1999; Olezza, 1999). The mean age of the participants was 10.39 years, and the modal age was 11 years, which is equivalent to the age of a fifth-grade student. Few studies presented results from studies with primary grade students (i.e., Geisler, 1999; Williams, 1999) or secondary education students (i.e., Billings, 1999; Olezza, 1999). As a result, data were analyzed irrespective of participant age.

Because information provided by study authors was often limited, we used broad categories to code the racial/ethnic distribution of the participants. Specifically, studies were coded on the basis of the majority racial/ethnic composition of the sample (i.e., greater than 60% White; greater than 60% African American; greater than 60% Hispanic/Latino; mixed, where no group was greater than 60%; or mixed, cannot estimate). The racial/ethnic backgrounds of the samples varied across the reviewed studies. The majority racial/ethnic composition of the samples of the reviewed studies was 16.7% mixed (i.e., no racial/ethnic group majority), 14.3% Hispanic majority, 14.3% White majority, 9.5% African American majority, and 9.5% mixed with the predominant racial/ethnic group not identified by the reporting authors.

Although these data seem to suggest that the various approaches have been examined with diverse samples, the diversity of the sampling is largely accounted for by two approaches. Specifically, five of the six studies conducted with predominantly Hispanic student samples were studies using Instructional Conversations. Two of the four studies conducted with predominantly African American student samples were studies using Questioning the Author. Moreover, more studies have been conducted with students from low socioecomonic backgrounds (35.7%) than with students from any other socioeconomic strata. Additionally, a majority of studies have been conducted with students living in urban settings (26.2%), while only 4.8% of the studies have investigated students from rural schools. Finally, the vast majority of the respondents in the reviewed studies were described by study authors as average (23.8%) or below average (19.0%) in academic ability.

Effects by Key Study Features

Prior to analyses, all effect sizes were unbiased and weighted using procedures recommended by Lipsey and Wilson (2001). Unbiasing procedures were completed because Hedges' $\hat{g}$ is biased

upward for studies with small samples. Unbiased effect sizes were obtained using the following formula:

$$\hat{g} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} * \left[1 - \frac{3}{4(n_1 + n_2) - 9}\right]$$

The resulting unbiased effect sizes were then weighted to give studies with larger sample sizes more power. Specifically, we wanted to give more weight to the effect sizes from studies with larger sample sizes because they are more likely to reveal true differences when they actually exist. The following formula was used to weight the effect sizes:

$$w = \frac{1}{se^2}$$

$$se = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{\hat{g}}{2(n_1 + n_2)}}$$

These unbiased and weighted effect sizes were used for all analyses.

Study design. Effects of the various approaches on the constructs of interest were moderated by the nature of the study design (see Table 2). As in other meta-analyses of intervention research in reading comprehension, we found that effects were weaker in multiple-group (experimental or quasi-experimental) rather than single-group (pretest–posttest) design studies. Thus, we chose to disaggregate the studies and results by design. For example, studies exhibited strong effects for text-explicit comprehension using single-group designs (e.g., Junior Great Books Shared Inquiry, effect size [ES] = 2.35; Instructional Conversations, ES = 2.99; Questioning the Author, ES = 0.95), whereas multiple-group designs using the same discussion approaches produced weaker effect sizes (e.g., Junior Great Books Shared Inquiry, ES = −0.01; Instructional Conversations, ES = 0.50; Questioning the Author, ES = 0.80). Similarly, as can be seen in Table 2, Literature Circles produced moderate effects on unspecified comprehension in single-group design studies and small effects in multiple-group design studies. The effects of Junior Great Books Shared Inquiry discussions on critical thinking/reasoning were heavily moderated by study design with single-group design studies yielding very strong effects (ES = 2.39) and multiple-group designs exhibiting

Table 2
Effect Sizes by Construct and Approach Comparing Single- and Multiple-Group Studies

Stance/approach/ grouping	Construct measured									
	TT	ST	SST	Comp	TE	TI	SI	CT/R	Arg	Meta
Critical–analytic										
CR	**−1.924**	**4.097**	**—**	***0.262***	**0.490**	**0.082**	***0.668***	**2.465**	**0.260**	**0.284**
Single	−1.924	4.097	—	—	0.490	0.082	—	2.465	—	—
Multiple	—	—	—	*0.262*	—	—	*0.668*	—	0.263	0.284
P4C	**−0.291**	**—**	**0.154**	***0.333***	**—**	**—**	**—**	**0.236**	**0.214**	—
Single	—	—	—	—	—	—	—	—	—	**—**
Multiple	−0.291	—	0.154	*0.333*	—	—	—	0.236	0.214	—
PS	**−0.343**	**0.220**	**—**	**—**	**—**	**—**	***0.428***	**—**	**—**	**—**
Single	−0.030	−0.006	—	—	—	—	—	—	—	—
Multiple	−0.655	0.446	—	—	—	—	*0.428*	**—**	**—**	**—**
Efferent										
QtA	**0.098**	**0.330**	**—**	**−0.205**	**0.899**	**—**	***0.627***	***2.499***	**—**	**—**
Single	0.098	0.330	—	−0.205	0.949	—	—	*2.499*	—	—
Multiple	—	—	—	—	*0.800*	—	*0.627*	**—**	**—**	**—**
IC	**−0.408**	**1.962**	**—**	***2.798***	**1.336**	**0.568**	***0.871***	—	—	—
Single	—	2.735	—	*2.798*	2.988	—	*1.263*	—	—	—
Multiple	−0.408	1.653	—	—	0.509	0.568	*0.610*	**—**	**—**	**—**
JGB	**—**	**—**	**—**	***0.333***	**0.331**	***1.124***	**—**	***0.718***	—	—
Single	—	—	—	—	*2.345*	*2.135*	—	*2.392*	—	—
Multiple	—	—	—	*0.176*	−0.005	*0.786*	**—**	*0.408*	**—**	**—**
Expressive										
LC	**−0.439**	**1.637**	**—**	***0.426***	**—**	***2.136***	**—**	**—**	**—**	**—**
Single	−0.439	1.637	—	*0.633*	—	*2.136*	—	—	—	—
Multiple	—	—	—	*0.114*	—	—	—	—	—	—
GC	**0.043**	**—**	**—**	**—**	**—**	**0.822**	**—**	**—**	**—**	**—**
Single	0.043	—	—	—	—	0.822	—	—	—	—
Multiple	—	—	—	—	—	—	—	—	—	—
BC	**—**	**0.050**	**—**	**—**	**—**	**—**	**—**	**—**	**—**	***1.073***
Single	—	0.050	—	—	—	—	—	—	—	*1.073*
Multiple	—	—	—	—	—	—	—	—	—	—

Note. Bolded numbers = effect sizes by approach across study design. Italicized numbers = instances in which outcomes were assessed by researchers using only individual outcome measures. Dashes = no data meeting criteria. CR = Collaborative Reasoning; P4C = Philosophy for Children; PS = Paideia Seminar; QtA = Questioning the Author; IC = Instructional Conversation; JGB = Junior Great Books; LC = Literature Circles; GC = Grand Conversation; BC = Book Club; TT = Teacher Talk; ST = Student Talk; SST = Student-to-Student Talk; Comp = General Comprehension; TE = Text-Explicit Comprehension; TI = Text-Implicit Comprehension; SI = Scriptally Implicit Comprehension; CT/R = Critical Thinking/Reasoning; Arg = Argumentation; Meta = Metacognition..

moderate effects (ES = 0.41). Due to the differences in effects attributable to study design, we have chosen to report all findings by study design.

Outcome measure. As in other meta-analyses of intervention research in reading comprehension, we found that effects were attenuated when researchers used commercially available, standardized measures rather than researcher-developed measures (see Figure 1). Several studies tested effects using commercially available, standardized measures with multiple-group designs. Those measures included the California Achievement Test, which was used in two studies (Banks, 1987; Mizerka, 1999), the Iowa Test of Basic Skills (ITBS) used by Lipman (1975), the Comprehensive Test of Basic Skills used by Yeazell (1982), the Ross Test of Critical Thinking used by both Chamberlain (1993) and Bird (1984), and the Worden Test of Critical Thinking and Reading used by Bird (1984). Among the studies employing commercially available, standardized measures with multiple-group designs, Philosophy for Children (P4C) evidenced the strongest positive effects on student comprehension. For example, Lipman (1975) compared scores on the ITBS from students who had taken part in P4C discussions during the 3 previous years with those of students who had not participated in P4C. Students in the P4C condition outperformed their comparable peer group in reading comprehension on the ITBS (ES = 0.55).

Effect size outcomes also varied depending on whether the measure assessed talk in the group or outcomes on tests given to individual students. Talk indices were used to assess teacher talk, student talk, and student-to-student talk, whereas researcher-made and commercially available measures were more commonly used to assess individual outcomes (see Figure 2). Results showed that Collaborative Reasoning, Instructional Conversations, and Literature Circles all produced very strong improvements in the amount of student talk and concomitant reductions in teacher talk (see Table 2). For example, Collaborative Reasoning was associated with an increase in student talk of 4.1 standard deviations and a decrease in teacher talk of 1.9 standard deviations. However, there were also nine studies in which talk indices were used to assess various forms of student comprehension, critical thinking and reasoning, and argumentation (e.g., Echevarria, 1995; McGee, 1992). As a case in point, Chinn et al. (2001) used student talk in the group as a measure of unspecified comprehension, text-explicit comprehension, text-implicit comprehension, and critical thinking/reasoning (see Figure 2). Of the nine studies in which talk indices served as proxies for other constructs, six measured the same construct both through talk and individual outcome measures. In every case, effect sizes were stronger for the construct when it was measured with an individual outcome measure. In most cases, the effect sizes were almost 50% larger when measured with individual outcome measures.

Effects by Primary Stance and Approach

Critical–analytic approaches. The critical–analytic stance is represented in the meta-analysis by 11 empirical studies, only 2 of which were single-group designs. As can be seen in Table 2, the approaches within this stance were effective at increasing student

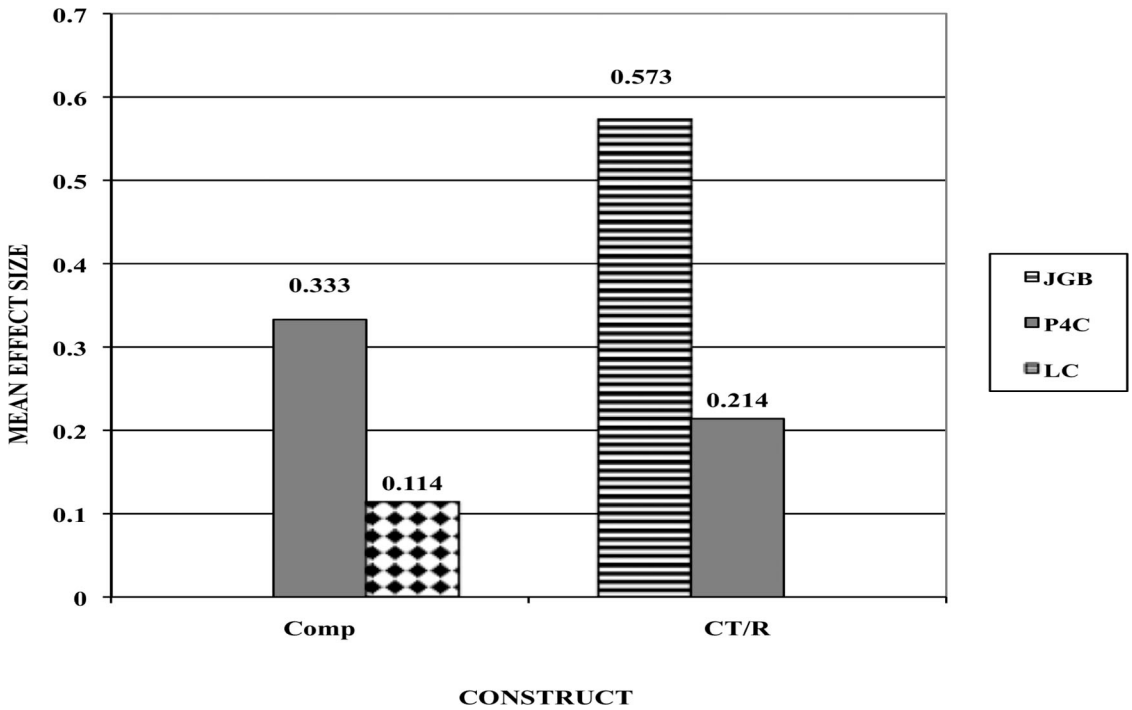
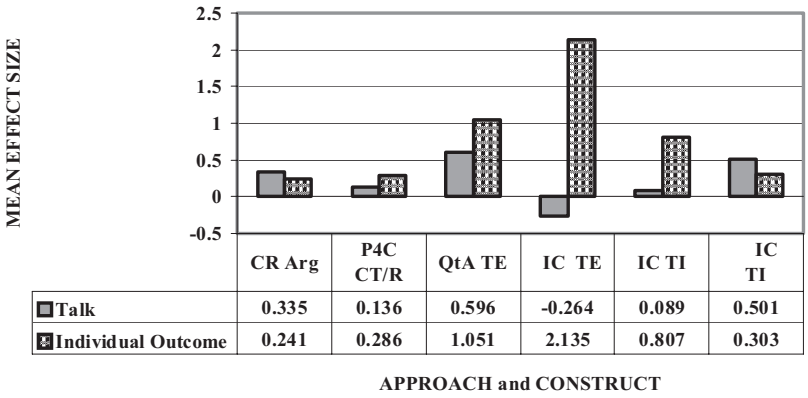


Figure 1. Effects by construct and approach for commercial assessments used in multiple-group design studies. Effect sizes were averaged within and across studies by approach. Comp = General Comprehension; CT/R = Critical Thinking/Reasoning; JGB = Junior Great Books; P4C = Philosophy for Children; LC = Literature Circles.



	CR Arg	P4C CT/R	QtA TE	IC TE	IC TI	IC TI
Talk	0.335	0.136	0.596	-0.264	0.089	0.501
Individual Outcome	0.241	0.286	1.051	2.135	0.807	0.303

Figure 2. Effects for talk versus individual outcome measures by approach and construct. CR = Collaborative Reasoning; P4C = Philosophy for Children; QtA = Questioning the Author; IC = Instructional Conversation; TE = Text-Explicit Comprehension; TI = Text-Implicit Comprehension; CT/R = Critical Thinking/Reasoning; Arg = Argumentation.

talk and decreasing teacher talk. This was especially the case for Collaborative Reasoning, in which the teacher's role is to gradually step back and let students control the flow of the discussion. Approaches within this stance were also effective at promoting low to moderate effects on various forms of student comprehension (i.e., unspecified, text-explicit, text-implicit, and scriptally implicit), although the constructs more often measured in this stance pertain to thinking and reasoning. Effect sizes for critical thinking and reasoning and argumentation ranged from 2.47 in Collaborative Reasoning (single-group study) to 0.24 in Philosophy for Children (multiple-group study). These effects make sense given that critical thinking is one of the main foci of these approaches. It is worth noting that research on Paideia Seminars have generally measured effects on constructs other than those of interest in the present meta-analysis (e.g., writing; Chesser, Gellalty, & Hale, 1997). Nonetheless, Paideia Seminars showed moderate effects on student talk, teacher talk, and scriptally implicit comprehension in multiple-group studies.

Efferent approaches. The efferent stance is represented by the highest number of empirical studies ($n = 21$) and the highest number of multiple-group studies ($n = 13$) relative to the other two stances. As shown in Table 2, the efferent approaches were successful at increasing student talk. Specifically, Instructional Conversations were highly effective at promoting student talk and decreasing teacher talk, whereas Questioning the Author exhibited only a moderate negative effect on student talk and minimally increased teacher talk. This latter finding makes sense in light of the centrality of the teacher in Questioning the Author discussions.

Studies conducted within the efferent stance had a predominant focus on comprehension and showed strong effects in single-group design studies with effect sizes ranging from a high of 2.79 standard deviations for Instructional Conversations (IC) to a low of −0.21 standard deviations for Questioning the Author (QtA) when general comprehension was measured. IC exhibited very strong effects on individual student outcomes, especially general comprehension (ES = 2.80, single-group design) and text-explicit comprehension (ES = 2.99, single-group design). As mentioned previously, most of these IC studies were conducted with predominantly low socioeconomic, high limited-English proficiency Hispanic populations attending urban schools. As such, these findings suggest that ICs are particularly effective at helping these struggling readers better comprehend narrative texts. Similarly, Junior Great Books Shared Inquiry discussions exhibited moderate to strong effects on text-explicit (ES = 2.35, single-group design) and text-implicit comprehension (ES = 2.14, single-group design), as well as critical thinking and reasoning (ES = 2.39, single-group design). Finally, QtA discussions resulted in weak effects on general comprehension (ES = −0.21), strong effects on text-explicit (ES = 0.90, average for single-group and multiple-group design) and scriptally implicit comprehension (ES = 0.63, multiple-group design), and very strong effects on students' critical thinking and reasoning (ES = 2.50, single-group design). As one might expect, the strength of these more efferent approaches lies in their powerful influence on individual student performance, especially as evidenced in pretest–posttest design studies.

Expressive approaches. The expressive stance is represented by 10 empirical studies, only 1 of which was a multiple-group design study. It is important to note that 6 of the 10 studies were conducted using Literature Circles (LC), and studies of this approach provided more quantitative information than was provided in studies on Grand Conversations or Book Club. In terms of student and teacher talk, LC exhibited strong positive effects on student talk (ES = 1.64, single-group design) and were moderately effective at decreasing teacher talk (ES = −0.44, single-group design). Given that the focus of this stance is on fostering students' affective, emotive responses to text through discussion and the fact that much of the discussions are student led, we were surprised by the lack of data regarding the assessment of student talk.

Similarly, very few studies within the expressive stance reported effects on individual student outcomes. Literature Circles produced moderate effects on unspecified comprehension (ES = 0.43, averaged across single-group and multiple-group designs) and very strong effects on text-implicit comprehension (ES = 2.14) in single-group design studies. By comparison, studies employing Grand Conversations showed moderate effects on text-implicit comprehension (ES = 0.82) in single-group designs. It is interesting that in our review, the only study using Book Club in which

individual student outcomes were reported measured effects on students' meta-cognition (Kong & Fitch, 2002/2003, ES = 1.07, single-group design) but not comprehension. Overall, very few quantitative studies have been conducted on the effects of the more expressive discussion approaches. That being said, a lack of quantitative data does not imply that the aforementioned approaches failed to affect students' comprehension of text. It merely indicates that quantitative data relative to these constructs of interest was not assessed in the reviewed studies. Despite the limited findings on the effects of some of the expressive approaches (e.g., Grand Conversation and Book Club), the available data for Literature Circles indicated strong effects on talk and comprehension resulting.

Student academic ability. Given that effect sizes for single-group designs were not comparable to those for multiple-group designs, all analyses relative to ability were disaggregated by study design (Table 3). Results showed that the approaches exhibited greater effects with below-average-ability students and weaker effects with average or above-average students. For single-group designs, there were no significant differences in comprehension between students who were below-average or average in ability, $F(1, 12) = 3.07$, $p = .11$, and no students characterized as above average in academic ability were investigated in single-group design studies. However, this difference was practically significant (ES = 0.88). Similarly, there were no significant differences in comprehension for below-average, average, or above-average students when participating in a multiple-group design, $F(2, 28) = 2.20$, $p = .13$. However, the differences between the comprehension of below-average students was practically different from the performance of students with average ability (ES = 1.03) and those with above-average ability (ES = 3.20). For multiple-group designs, there was a minimal difference in the comprehension of students of average and above-average ability (ES = 0.20). These findings suggest that the reviewed approaches were particularly beneficial at promoting talk and increasing performance on individual outcome measures for students characterized by authors as below average in academic ability (Table 3). For example, the use of Instructional Conversation discussions increased both the text-explicit (ES = 0.76) and text-implicit (ES = 0.81) comprehension of below-average ability students with limited-English proficiency (e.g., Saunders & Goldenberg, 1999). Evidence of gains in text-explicit comprehension for below-average-ability students who were not considered to be limited-English proficient was also found by multiple researchers when using approaches that give prominence to an efferent stance (e.g., Biskin, Hoskisson, & Modlin, 1976, ES = 0.72; Sandora, Beck, & McKeown, 1999, ES = 0.80). It is important to note, however, that these trends were derived from cross-sectional data and therefore must be interpreted with some caution.

Weeks of discussion. We were also interested in the extent to which effects were moderated by the number of weeks that students participated in discussions (see Table 4). Again, due to differences in effect sizes attributable to study design, we analyzed the data collected from single-group and multiple-group design studies separately. The mean number of weeks of discussion was 13. Categories were created for the number of weeks of discussion used in each study. Specifically, the standard deviation of the number of weeks (SD = 10 weeks) was added to and subtracted from the mean number of weeks of discussion to create the categories. Due to variability in the distribution of number of weeks among the studies, this technique resulted in two groups below the mean and three groups above the mean. Subsequently, categories of 1–3 weeks, 4–13 weeks, 14–23 weeks, 24–33 weeks, and 34–36 weeks resulted from this procedure. Trends for indices of talk suggest that student talk increased the longer discussions were implemented (e.g., Howard, 1992; Saunders &

Table 3
Effect Sizes by Construct and Student Ability Comparing Single- and Multiple-Group Studies

	Construct measured						
Design/measure/ability	TT	ST	Comp	TE	TI	SI	CT/R
Single group							
Individual							
Below	—	—	—	3.267	2.135	1.261	2.446
Average	—	—	0.633	—	—	—	—
Above	—	—	—	—	—	—	—
Talk							
Below	—	2.735	—	0.490	—	—	—
Average	−0.613	1.466	—	−0.179	0.082	—	2.465
Above	−0.030	−0.006	—	—	—	—	—
Multiple group							
Individual							
Below	—	—	—	0.781	0.807	0.614	—
Average	—	—	0.333	−1.032	0.531	0.235	—
Above	—	—	0.040	−0.002	0.043	—	0.299
Talk							
Below	−0.408	1.653	—	−0.349	0.089	—	—
Average	−0.655	0.446	—	0.501	—	—	—
Above	−0.291	—	—	—	—	—	0.136

Note. Dashes = no data meeting criteria. TT = Teacher Talk; ST = Student Talk; Comp = General Comprehension; TE = Text-Explicit Comprehension; TI = Text-Implicit Comprehension; SI = Scriptally Implicit Comprehension; CT/R = Critical Thinking/Reasoning.

Table 4
Effect Sizes by Construct and Weeks of Discussion Comparing Single- and Multiple-Group Studies

Design/measure/ no. of weeks	Construct measured						
	TT	ST	Comp	TE	TI	SI	CT/R
Single group							
Individual							
1–3	—	—	1.077	—	—	—	—
4–13	—	—	0.412	1.823	2.135	—	2.446
14–23	—	—	—	—	—	—	—
24–33	—	—	—	6.155	—	1.261	—
34–36	—	—	—	—	—	—	—
Talk							
1–3	—	—	—	—	—	—	—
4–13	−1.924	2.074	—	0.490	0.082	—	2.465
14–23	−0.030	−0.006	—	—	—	—	—
24–33	—	2.735	—	−0.179	—	—	—
34–36	—	—	—	—	—	—	—
Multiple group							
Individual							
1–3	—	—	0.433	−0.060	0.715	0.516	—
4–13	—	—	0.551	0.800	—	0.627	0.286
14–23	—	—	.0397	0.689	1.785	—	0.430
24–33	—	—	0.199	−0.002	0.043	—	0.014
34–36	—	—	0.430	—	—	—	—
Talk							
1–3	−0.532	0.426	—	−0.349	0.089	—	—
4–13	−0.291	—	—	—	—	—	0.136
14–23	—	—	—	0.501	—	—	—
24–33	—	—	—	—	—	—	—
34–36	—	—	—	—	—	—	—

Note. Dashes = no data meeting criteria. TT = Teacher Talk; ST = Student Talk; Comp = General Comprehension; TE = Text-Explicit Comprehension; TI = Text-Implicit Comprehension; SI = Scriptally Implicit Comprehension; CT/R = Critical-Thinking/Reasoning.

Goldenberg, 1998). Similarly, teacher talk decreased the longer implementation continued. These trends were seen for both single- and multiple-group designs (Table 4).

The trends were different, however, for individual student outcomes. For single-group design studies, large effect sizes were observed for general comprehension in studies in which discussions lasted less than 4 weeks (ES = 1.08), and it is interesting that effect sizes decreased for studies in which discussions lasted longer (ES = 0.41). In studies investigating general comprehension with multiple-group designs, effect sizes remained in a moderate range (0.20–0.55) regardless of the number of weeks that a discussion approach was employed. These results suggest that the greatest effect of the implementation of a discussion approach on general comprehension occurs in the first 3 weeks (e.g., Graup, 1985). In the case of multiple-group design studies investigating text-explicit and text-implicit comprehension, it appears that there was a ceiling effect on student gains. Specifically, effect sizes for these two types of comprehension were moderate for discussions lasting less than 24 weeks (e.g., Heinl, 1988; Sandora et al., 1999; Saunders & Goldenberg, 1999), yet negligible for multiple-group studies lasting longer than 24 weeks (text-explicit comprehension, −0.002; text-implicit comprehension, 0.04). The benefits of using discussion to increase text-explicit and text-implicit forms of comprehension can be seen early in implementation of the discussion approach, but such increases tend to dissipate over time. This ceiling effect can also be seen in multiple-group designs assessing students' critical thinking or reasoning. Specifically, before 24 weeks of discussion, effect sizes are moderate (0.29–0.43), and after 24 weeks of implementation, effects on students' critical thinking are negligible (ES = 0.01). Again, we must note that these trends were derived from cross-sectional data and therefore must be interpreted with some caution.

Modeling Variability in Effect Sizes

Finally, beyond descriptive outcomes, we were also interested in modeling the effects of discussion on high-level comprehension using fixed-effects and random-effects models to explain variability across studies. To address the issue of variability across studies, we completed an investigation of the effect sizes (Hedges' $\hat{g}$) obtained from the studies assessing student comprehension. Due to the lack of comparability between single-group and multiple-group study outcomes (i.e., attenuated effects in multiple-group designs), only multiple-group design studies were included in the analyses. We focused the modeling on comprehension because although other constructs were reviewed herein, the largest proportion of the quantitative data was on comprehension outcomes. For this analysis, data from the four types of comprehension (i.e., general/

unspecified comprehension, text-explicit comprehension, text-implicit comprehension, and scriptally implicit comprehension) were analyzed. Of the reviewed studies, multiple effect sizes were reported for only six of the nine discussion approaches (i.e., Collaborative Reasoning, Philosophy for Children, Questioning the Author, Instructional Conversation, Junior Great Books Shared Inquiry, and Literature Circles). As such, only data reported in studies pertaining to these six approaches were included in the present analysis.

Fixed-effects model. Having calculated the weighted effect sizes, a homogeneity analysis was conducted to test the assumption that all effect sizes are estimating the same population. Specifically, we wanted to determine whether the comprehension effect sizes obtained from the six discussion approaches were similar or whether variability was inherent in the effect size measures. The homogeneity analysis was tested using a Q statistic and is distributed as a chi-square distribution.

$$Q = \sum(\hat{g})^2 w - \frac{\left[\sum \hat{g} w\right]^2}{\sum w}$$

The test of this assumption was significant, Q_{total} (38, 0.05) = 318.06, $p < .0001$, suggesting that the variation in the effect sizes was due to more than random sampling error.

Next, we were interested in determining the extent to which various features of the studies influenced the variability of the effect sizes. We first partitioned the data by discussion approach because the varying goals and structure of the discussion approaches may contribute to differences in comprehension increases. A Q statistic was calculated for each of the sets of effect sizes representing each discussion approach. These Q statistics were added to determine the Q_{within} groups.

$Q_{\text{within}} = Q_{\text{CR}} + Q_{\text{P4C}} + Q_{\text{QtA}} + Q_{\text{IC}} + Q_{\text{JGB}} + Q_{\text{LC}}$

$Q_{\text{within}} = 3.474 + 186.361 + 0.169 + 12.123 + 98.466 + 0.062$

$Q_{\text{within}} = 300.655$

The Q_{within} value gives an indication that much of the variance in the effect sizes was due to the discussion approach used in the specific studies from which the effect sizes were calculated. However, this supposition must be tested to determine whether discussion approach alone accounts for the variability in the effect sizes or whether a significant portion of variance remains unexplained. To do this, we then subtracted the Q_{within} value from Q_{total} to determine Q_{between}.

$Q_{\text{between}} = Q_{\text{total}} - Q_{\text{within}}$

$Q_{\text{between}} = 318.060 - 300.655$

$Q_{\text{between}} = 17.405$

$Q_{between}$ represents the variance remaining in the effect sizes after accounting for variance due to discussion approach. This value was significant at the .05 level, Q_{between} (6, 0.05) = 17.41, $p = .01$, suggesting that there is variability in the effect sizes above and beyond that explained by discussion approach. In other words, different models must be tested to determine the factors best explaining the variance in effect sizes.

The same procedure was used to determine whether the variability in Q_{total} was due to the differences in the types of comprehension effect sizes that were analyzed. In other words, the total variance in the effect sizes may be due to the type of comprehension assessed (i.e., general [COMP], text-explicit [TE], text-implicit [TI], or scriptally implicit [SI] comprehension) rather than the discussion approaches used in the studies. Thus, the data were partitioned due to the type of comprehension assessed.

$Q_{\text{within}} = Q_{\text{Comp}} + Q_{\text{TE}} + Q_{\text{TI}} + Q_{\text{SI}}$

$Q_{\text{within}} = 18.663 + 83.150 + 20.102 + 177.732$

$Q_{\text{within}} = 299.647$

$Q_{\text{between}} = 318.060 - 299.647$

$Q_{\text{between}} = 18.413$

This value was significant at the .05 level, Q_{between} (3, 0.05) = 18.41, $p < .0001$, suggesting that there is variability in the effect sizes above and beyond that which is explained by the differences in the types of comprehension assessed.

It may be that discussion approach and type of comprehension assessed explain different variability in the effect sizes. Thus, the model including both factors was tested. The data were partitioned by both the approach used and the type of comprehension assessed. Table 5 provides the Q statistics for each of the 13 partitioned groups that contribute to Q_{within}. The sum of all the Q statistics for the partitioned groups was 291.64. Thus:

$Q_{\text{between}} = 318.060 - 291.643$

$Q_{\text{between}} = 26.417$

This value was again significant at the .05 level, Q_{between} (12, 0.05) = 26.41, $p = .01$, suggesting that variability exists in the effect sizes beyond that explained by the differences in both the discussion approach and the type of comprehension assessed.

Random-effects model. Due to the lack of fit of the fixed-effects models, the fit of random-effects models was assessed. Whereas the fixed-effects model assumes that the variability be-

Table 5
Q Statistics Partitioned by Both Discussion Approach and Type of Comprehension Assessed

Approach/comprehension assessed	Q statistic
CR	
Comp	0.406
SI	1.629
P4C	
Comp	8.775
SI	174.258
QtA	
TE	0.000
SI	0.002
IC	
TE	3.200
TI	0.853
SI	0.184
JGB	
Comp	6.516
TE	73.937
TI	15.760
LC	
Comp	0.123
Q_{within}	291.643

Note. CR = Collaborative Reasoning; Comp = General Comprehension; SI = Scriptally Implicit Comprehension; P4C = Philosophy for Children; QtA = Questioning the Author; TE = Text-Explicit Comprehension; TI = Text-Implicit Comprehension; JGB = Junior Great Books; LC = Literary Circles.

tween effect sizes is due only to sampling error, the random-effects model presupposes that the variability between effect sizes is due to sampling error plus variability in the populations being assessed. The excess variability seen in the fixed-effects models may be derived from random differences in the populations that cannot be measured. For example, the discussion approaches may have been implemented in ways that differ across studies. In addition, the comprehension assessments may not be comparable across studies. Thus, due to these unexplained differences in the factors, a random-effects model may be more appropriate. For the random-effects model, a constant is added to the calculation of the weights to represent the variability across the population effects:

$$w = \frac{1}{se^2 + \hat{v}_\theta}$$

In order to assess the fit of random-effects models, we calculated the random-effects variance component:

$$\hat{v}_\theta = \frac{Q_{total} - k - 1}{\sum w - \left(\frac{\sum w^2}{\sum w}\right)},$$

where k is equal to the number of effect sizes. This formula is based upon Q_{total} calculated previously. The variance component for the random-effects models tested was calculated to be .16.

Using the new weights, we performed the previously conducted analysis again to assess the variability among the effect sizes. First, a homogeneity analysis was conducted again with no partitioning of the effect sizes. The results were significant, Q_{total} (39, 0.05) = 60.87, $p = .01$, suggesting that there is a need for partitioning the effect sizes in some way to account for variability in the population.

Thus, the effect sizes were again partitioned by approach.

$Q_{\text{within}} = Q_{\text{CR}} + Q_{\text{P4C}} + Q_{\text{QtA}} + Q_{\text{IC}} + Q_{\text{JGB}} + Q_{\text{LC}}$
$Q_{\text{within}} = 1.484 + 5.108 + 0.061 + 2.130 + 46.601 + 0.046$
$Q_{\text{within}} = 55.430$
$Q_{\text{between}} = Q_{\text{total}} - Q_{\text{within}}$
$Q_{\text{between}} = 60.872 - 55.430$
$Q_{\text{between}} = 5.442$

This result was not statistically significant at the .05 level, Q_{between} (5, 0.05) = 5.44, $p = .36$, suggesting that the variance in discussion approaches accounts for the variability in the effect sizes using a random-effects model. In other words, all of the variability between effect sizes observed in the homogeneity analysis was explained by the differences in the discussion approaches.

As a comparison, we also partitioned the effect sizes by the type of comprehension assessed because this variable may explain more variance in the effect sizes than the variance in the discussion approaches.

$Q_{\text{within}} = Q_{\text{Comp}} + Q_{\text{TE}} + Q_{\text{TI}} + Q_{\text{SI}}$
$Q_{\text{within}} = 3.759 + 42.645 + 6.551 + 3.328$
$Q_{\text{within}} = 56.283$
$Q_{\text{between}} = 60.872 - 56.283$
$Q_{\text{between}} = 4.589$

This result was also not statistically significant at the 0.05 level, Q_{between} (3, 0.05) = 4.59, $p = .20$, suggesting that the differences in the types of comprehension assessed also accounted for the variability in the effect sizes using a random-effects model.

However, in conceptually evaluating the outcomes of these two random-effects models, we deemed the random-effects model partitioning the data by discussion approach to be a better explanation for variation in the outcomes than the model partitioning the data by the type of comprehension assessed. In essence, the nature and structure of a particular discussion *approach* should arguably serve as the theoretical and operational underpinning of a given study, while *type of comprehension* would be confounded with study, and possibly, approach. As a case in point, researchers examining the effects of Philosophy for Children assessed only general or unspecified comprehension, whereas researchers examining Literature Circles assessed general or unspecified comprehension and text-implicit comprehension. As such, we conclude that approach statistically and conceptually accounts for maximum variance in comprehension effect sizes across the reviewed studies.

Discussion

A key presupposition within the vast body of literacy research is that discussions about and around text enhance students' comprehension, thinking, and reasoning (e.g., Almasi et al., 1996; Cazden, 1988). Such a perspective is situated within a rich sociocultural tradition emerging from the classic work of scholars such as Vygotsky (1978) and more contemporary theorists like Bakhtin (1981, 1986) who suggest that thinking and reasoning are inherently dialogical. Moreover, literacy researchers have amassed approximately 300 manuscripts and studies and created more than a dozen discussion approaches aimed at understanding and enhancing the fundamental role of classroom discourse in comprehension and learning. Yet, no researcher, to date, has closely examined the effects of the various discussion approaches on students' text comprehension and learning. The purpose of this review was to analyze the effects of selected approaches to discussion on students' high-level comprehension of text using meta-analytic techniques. In doing so, we paid careful attention to characteristics of the nature and design of the study, including sample characteristics, research design, dependent measures, and results.

A number of key findings emerged in our review of relevant literature. First, we found that many of the approaches were highly effective at promoting students' literal and inferential comprehension, especially those that we categorized as more efferent in nature, and that relatively few of the approaches were particularly effective at promoting students' critical thinking, reasoning, and argumentation about and around text. Another major finding was that most approaches were effective at increasing student talk and decreasing teacher talk. However, increases in student talk did not necessarily result in concomitant increases in student comprehension. Finally, we found that effectiveness of an approach at increasing student comprehension, critical thinking and reasoning, and argumentation were substantively attenuated due to study design and the nature of measures employed. In the paragraphs that follow, we expand upon each of these issues as well as several other findings.

Prior to offering more specific concluding remarks and implications emerging from this review of relevant literature, we felt it was important to recognize some of the limitations of the present work. First, we purposefully constrained the literature included in the review in ways that may have influenced the results. For example, we included only empirical works pertaining to one of

the nine selected discussion approaches. Certainly, there exist discussion formats beyond these nine approaches (e.g., Point–Counterpoint, Rogers, 1990/1991), and the inclusion of data from different approaches might have altered our outcomes. It is important to reiterate, however, that we attempted to be as exhaustive as possible in our selection of approaches but felt it was equally imperative that any included discussion approach be substantiated by a record of published, peer-reviewed research.

Our selection of constructs of interest might also be seen as a limiting factor. As we mentioned previously, for example, some approaches (e.g., Paedia Seminar) measured outcomes not aligned with our constructs of interest (e.g., writing). That being said, our goal in the present review was to focus on proximal indicators of high-level comprehension such as various forms of comprehension, critical thinking and reasoning, or argument. Certainly the review of other, more distal indicators seems to be ripe fodder for future research. Finally, the fact that we chose meta-analytic techniques over other types of syntheses may give the appearance of a limitation. In essence, the selection of meta-analytic techniques meant that no purely qualitative results could be incorporated into the analyses. Although we conceptually agree with those who might take such a position, we felt it was important to look primarily at quantitative outcomes due to their importance in educational decision making. Anna O. Soter is simultaneously conducting a best-evidence synthesis of pertinent qualitative study outcomes. Despite these limitations, we feel that the present meta-analysis of the classroom discussion literature bears important results—results that have the potential to influence educational research and practice.

Perhaps the most substantive theoretical and educational contribution of this study was the finding that the various approaches to discussion differentially promoted high-level comprehension of text. This result was manifest in our descriptive analysis as well as in our random-effects model. Many of the approaches were effective at promoting students' comprehension in multiple-group design studies, especially those that we categorized as more efferent in nature, namely, Questioning the Author, Instructional Conversation, and Junior Great Books Shared Inquiry. As would be expected, effects of discussion on comprehension were even greater in the single-group design studies. Some of the approaches were particularly effective at promoting students' critical thinking, reasoning, and argumentation about and around text in multiple-group design studies (i.e., Collaborative Reasoning, Philosophy for Children, and Junior Great Books Shared Inquiry) and in single-group design studies (i.e., Collaborative Reasoning, Questioning the Author, and Junior Great Books Shared Inquiry). Relatively few approaches were effective at increasing literal or basic comprehension and high-level comprehension (i.e., critical thinking and reasoning about or around text) in multiple-group design studies (i.e., Collaborative Reasoning, Philosophy for Children, and Junior Great Books Shared Inquiry).

Very few studies examined the effects of classroom discussion on metacognition with the notable exceptions of studies using Collaborative Reasoning and Book Club. Book Club discussions were highly effective at promoting students' metacognition in single-group design studies. It is important to again note that effects on measures of comprehension, critical thinking and reasoning, argumentation, and metacognition were all attenuated in multiple-group design studies.

Another major finding from the review was that the various discussion approaches were extremely effective at increasing student talk and decreasing teacher talk in both single-group designs and in multiple-group designs. In essence, it would appear that these classroom discussion formats allowed students to have more classroom time to share their thoughts, while in many of the approaches, teachers took more of a facilitative role (e.g., Collaborative Reasoning or Philosophy for Children). What was surprising, however, was that increases in student talk did not necessarily result in concomitant increases in student comprehension. For example, in Collaborative Reasoning discussions, student talk increased by almost 4 standard deviations and teacher talk decreased by approximately 2 standard deviations, but students' comprehension gains were generally less than 0.5 of a standard deviation. It would seem that increasing talk is not enough; rather, a particular kind of talk is necessary to promote comprehension. Moreover, it would appear that teachers do not necessarily have to talk less in order to enhance student comprehension. As a case in point, teacher talk actually increased slightly after Questioning the Author was implemented, yet students in these studies showed gains in text-explicit comprehension, scriptally implicit comprehension, and critical thinking and reasoning in single-group design studies.

As mentioned previously, we also found that the effects of discussion were moderated by the study design and nature of the outcome measures. As in other intervention research in reading comprehension, effects were weaker in multiple-group than in single-group (pretest–posttest) design studies. Effects on individual outcomes were also attenuated when researchers used commercially available, standardized measures rather than researcher-developed measures. Only five studies tested effects of commercially available, standardized measures using multiple-group designs. Among the studies employing commercially available, standardized measures with multiple-group designs, the strongest effect on student comprehension relative to a control condition was recorded for Philosophy for Children (P4C) over 30 years ago (Lipman, 1975). In that study, scores on the Iowa Test of Basic Skills from students who had taken part in P4C discussions 3 years previously were compared with scores from comparable students with no P4C experiences.

The findings also revealed that use of these discussion approaches appears to be more potent for students of below-average ability than for students of average or above-average ability, possibly due to the fact that students of higher ability levels already possess the skills needed to comprehend narrative text. The number of weeks of discussion played a role in the effects that discussions had on comprehension and critical thinking, particularly when multiple-group designs were employed. Specifically, before 24 weeks of discussion, effect sizes were moderate; minimal changes in students' comprehension outcomes occurred after this period.

A number of conclusions can be drawn about the framing variables and demographics of participants in the reviewed studies. First, it seems that more and more attention is being paid to the important role of discussion in text comprehension, particularly on the part of students completing doctoral theses and dissertations. Although empirical research in this area dates back to the early 1960s (Casper, 1964), more than half of the studies we included in the review were conducted in the last decade. Moreover, of the 42 studies, almost half were doctoral dissertations. Also of interest

was that more than 70% of the included studies were conducted by researchers who played a primary role in the creation of each discussion approach or whom we have referred to as proponents or developers of the approaches. In essence, interest in the role of classroom discussions is rising among literacy researchers and their students. Given that so many of the studies have been conducted by proponents or developers of given approaches and/or their students, it is difficult to gauge the portability of the effects of the approach. Reasons for the increased attention to the relations between classroom discussions and student comprehension are not clear. As mentioned previously, the increase in the number of investigations in this area could be attributed to a number of sources including policy legislation and concomitant funding initiatives, social agendas and ills, or paradigmatic shifts in the field of literacy. While linking particular approaches or studies to any one of the previously mentioned sources is beyond the scope of the present study, we feel strongly that such empirical research is necessary and important.

We were pleasantly surprised by the diversity of the students participating in the studies. The reviewed approaches were implemented and exhibited robust effects with students of diverse ethnicities, abilities, and, to a limited extent, ages. Ages of participants ranged from very young (e.g., first grade) to adolescents (e.g., college), and the selection of approach did not seem to be linked in any way to the age of participants. In contrast, we did find that some approaches were more often implemented with ethnic minorities than were other approaches. Specifically, five of the six studies conducted with predominantly Hispanic student samples were studies using Instructional Conversations. Two of the four studies conducted with predominantly African American student samples were studies using Questioning the Author. Additionally, most of the respondents across the various studies attended schools in urban settings and were characterized as having low socioeconomic backgrounds. As such, the aggregate outcomes seem to represent the effects one might expect for a sample pool of relatively poor, ethnically diverse, 11-year old students with low to average reading ability.

Several implications for research and practice can be drawn from these findings and conclusions. First, it would appear that many more quantitative studies are needed to further examine the various approaches to discussion. It would be particularly helpful if the approaches were examined by individuals other than their proponents or developers and with students from suburban and rural schools. In considering the design of future studies, many more multiple-group studies are needed, particularly ones in which commercially available assessments are employed as outcome measures. Finally, we would urge researchers to provide as many indicators of comprehension as possible, including individual outcome measures as well as talk.

Given the powerful influence of the type of approach in influencing student comprehension, it seems particularly important that practicing educators pay careful attention to the goals of the approach. In the end, not all of the classroom discussion approaches have the same goals nor result in the same kinds of effects. A teacher keenly interested in enhancing a particular kind of comprehension or critical thinking and reasoning should consider how such an instructional goal aligns with the goals and outcomes reported for a given approach. It also would seem that most of the approaches would be robust to moderate variations in student characteristics including ethnicity and ability. Finally, it seems a powerful message that teacher talk decreased for almost all of the approaches (the exception was Questioning the Author). It would appear that most of the approaches require teachers to yield the floor to students to some extent while mindfully attending to the nature of the discourse.

In the end, this meta-analysis revealed that not all discussion approaches are created equal, nor are they equally powerful at increasing students' high-level comprehension of text. In fact, very few approaches were effective at increasing literal or inferential comprehension *and* critical thinking and reasoning about text. Nonetheless, talk appears to play a fundamental role in text-based comprehension. In effect, what this extensive analysis reminded us was that talk is a *means* and not an *end.* It is one thing to get students to talk to each other during literacy instruction but quite another to ensure that such engagement translates into significant learning. Simply putting students into groups and encouraging them to talk is not enough to enhance comprehension and learning; it is but a step in the process.

References

References marked with an asterisk indicate studies included in the meta-analysis.

ACT, Inc. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading.* Iowa City, IA: Author.

Almasi, J. F., McKeown, M. G., & Beck, I. L. (1996). The nature of engaged reading in classroom discussions of literature. *Journal of Literacy Research, 28,* 107–146.

Anderson, R. C. (1977). The notion of schemata and the educational enterprise. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 415–431). Hillsdale, NJ: Erlbaum.

Anderson, R. C., Chinn, C., Chang, J., Waggoner, J., & Yi, H. (1997). On the logical integrity of children's arguments. *Cognition and Instruction, 15,* 135–167.

Anderson, R. C., Chinn, C., Waggoner, M., & Nguyen, K. (1998). Intellectually stimulating story discussions. In J. Osborn & F. Lehr (Eds.), *Literacy for all: Issues in teaching and learning* (pp. 170–186). New York: Guilford Press.

Anderson, R. C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., et al. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction, 19,* 1–46.

Applebee, A., Langer, J., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Education Research Journal, 40,* 685–730.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays by M. M. Bakhtin* (M. Holquist, Ed. & Trans. & C. Emerson, Trans.). Austin, TX: University of Texas Press.

Bakhtin, M. M. (1986). *Speech genres and other late essays* (V. W. McGee, Trans.). Austin, TX: University of Texas Press.

*Banks, J. C. R. (1987). *A study of the effects of the critical thinking skills program, Philosophy for Children, on a standardized achievement test.* Unpublished doctoral dissertation, Southern Illinois University, Carbondale.

Beck, I. L., & McKeown, M. G. (2006). *Improving comprehension with Questioning the Author: A fresh and expanded view of a powerful approach.* New York: Scholastic.

Beck, I. L., McKeown, M. G., & Gromoll, E. W. (1989). Learning from social studies texts. *Cognition and Instruction, 6,* 99–158.

Beck, I. L., McKeown, M. G., Hamilton, R. L., & Kucan, L. (1998). Getting at the meaning. *American Educator, 66, 71,* 85.

*Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the Author: A year-long classroom implementation to engage students with text. *The Elementary School Journal, 96,* 385–414.

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 26,* 251–276.

*Billings, L. (1999). *Discussion in the high school classroom: An examination of sociolinguistic roles.* Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.

Billings, L., & Fitzgerald, J. (2002). Dialogic discussion and the Paideia Seminar. *American Educational Research Journal, 39,* 907–941.

*Bird, J. B. (1984). *Effects of fifth graders' attitudes and critical thinking/reading skills resulting from a Junior Great Books program.* Unpublished doctoral dissertation, Rutgers–The State University of New Jersey, New Brunswick.

*Biskin, D. S., Hoskisson, K., & Modlin, M. (1976). Prediction, reflection, and comprehension. *The Elementary School Journal, 77,* 131–139.

Black, J. B., & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior, 20,* 267–275.

Blum, H., Lipsett, L., & Yocom, D. (2002). Literature Circles: A tool for self-determination in one middle school inclusive classroom. *Remedial and Special Education, 23,* 99–108.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Callison, D. (2000). Critical literacy. *School Library Media Activities Monthly, 16,* 34–36.

*Cashman, R. K. (1977). *The effects of Junior Great Books program at the intermediate grade level (4–5–6) on two intellectual operations, verbal meaning, and reasoning ability*. Unpublished doctoral dissertation, Boston College.

*Casper, T. P. (1964). *Effects of the Junior Great Books program at the fifth grade level on four intellectual operations and certain of their component factors as defined by J. P. Guilford.* Unpublished doctoral dissertation, Saint Louis University.

*Castleberry Independent School District (1996). *Junior Great Books: Summary of program implementation and evaluation, 1995–1996.* Castleberry, TX: Author.

Cazden, C. (1988). *Classroom discourse: The language of teaching and learning.* Portsmouth, NH: Heinemann.

*Chamberlain, M. A. (1993). *Philosophy for Children program and the development of critical thinking of gifted elementary students.* Unpublished doctoral dissertation, University of Kentucky, Lexington.

Chang-Wells, G. L. M., & Wells, G. (1993). Dynamics of discourse: Literacy and the construction of knowledge. In E. A. Forman, N. Minick, & C. A. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children's development* (pp. 58–90). New York: Oxford University Press.

*Chesser, W. D., Gellalty, G. B., & Hale, M. S. (1997). Do Paideia Seminars explain higher writing scores? *Middle School Journal, 29,* 40–44.

*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.

Dalton, S., & Sison, J. (1995). *Enacting Instructional Conversation with Spanish-speaking students in middle school mathematics* (Report No. 12). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED382030)

*Davis, B. H., Resta, V., Davis, L., I., & Camacho, A. (2001). Novice teachers learn about Literature Circles through collaborative action research. *Journal of Reading Education, 26,* 1–6.

de Castell, S., & Luke, A. (1987). Literacy instruction: Technology and technique. *American Journal of Education, 95,* 413–430.

Duffy, G., Roehler, L., & Mason, J. (1984). *Comprehension instruction: Perspectives and suggestions.* New York: Longman.

*Echevarria, J. (1995). Interactive reading instruction: A comparison of proximal and distal effects of instructional conversations. *Exceptional Children, 61,* 536–552.

*Echevarria, J. (1996). The effects of instructional conversations on the language and concept development of Latino students with learning disabilities. *The Bilingual Research Journal, 20,* 339–363.

Eeds, M., & Wells, D. (1989). Grand Conversations: An exploration of meaning construction in literature study groups. *Research in the Teaching of English, 23,* 4–29.

Ellis, N. (1999). *Improving reading comprehension by the use of Junior Great Books in Grades 3 and 6 at Simon Guggenheim School.* Unpublished doctoral dissertation, Loyola University, Chicago.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–25). New York: Freeman.

*Farinacci, M. (1998). "We have so much to talk about": Implementing Literature Circles as an action–research project. *The Ohio Reading Teacher, 32,* 4–11.

Fleener, A. M. (1994). *An analysis of rhetorical vision constructed by seventh and eighth grade participants in Paideia Seminars.* Unpublished doctoral dissertation, University of Minnesota, Twin Cities.

*Flynn, P. A. M. (2002). *Dialogic approaches toward developing third graders' comprehension using Questioning the Author and its influence on teacher change.* Unpublished doctoral dissertation, Fordham University, New York.

Frederiksen, J. R. (1981). Sources of process interactions in reading. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 361–386). Hillsdale, NJ: Erlbaum.

Freebody, P., & Luke, A. L. (1990). "Literacies" programs: Debates and demands in cultural context. *Prospect, 5*(3), 7–17.

*Geisler, D. L. (1999). *The influence of Spanish instructional conversations upon the oral language and concepts about print of selected Spanish-speaking kindergarten students.* Unpublished doctoral dissertation, Texas Woman's University, Denton.

*Goatley, V. J., Brock, C. H., & Raphael, T. E. (1995). Diverse learners participating in regular education "Book Clubs." *Reading Research Quarterly, 30,* 352–380.

Goldenberg, C. (1993). Instructional conversations: Promoting comprehension through discussion. *The Reading Teacher, 46,* 316–326.

*Graup, L. B. (1985). *Response to literature: Student generated questions and collaborative learning as related to comprehension (Cognitive, Essays, Junior Great Books*). Unpublished doctoral dissertation, Hofstra University, Hempstead, New York.

Graves, M. F., Juel, C., & Graves, B. B. (2001). *Teaching reading in the 21st Century* (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Great Books Foundation. (1987). *An introduction to shared inquiry.* Chicago: Author.

Hatano, G. (1993). Time to merge Vygotskian and constructivist conceptions of knowledge acquisition. In E. A. Forman, N. Minick, & C. A. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children's development* (pp. 153–166). New York: Oxford University Press.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

*Heinl, A. M. (1988). *The effects of the Junior Great Books program on literal and inferential comprehension.* Unpublished master's thesis, University of Toledo.

*Howard, J. B. (1992). *Seminar discussion and enlarged understanding of*

ideas. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.

Iser, W. (1978). *The act of reading: A theory of aesthetic response.* Baltimore: John Hopkins University Press.

Jakobson, R. (1987). The poetry of grammar and the grammar of poetry. In K. Pomorska & S. Rudy (Eds.), *Language in literature* (pp. 121–144). Cambridge, MA: The Belnap Press of Harvard University.

Jongsma, K. S. (1991). Critical literacy. *The Reading Teacher, 44,* 518–519.

*Junior Great Books. (1992). *The Junior Great Books curriculum of interpretive reading, writing and discussion.* Chicago: Great Books Foundation.

*Kim, S. (2002). *The effects of group monitoring on transfer of learning in small group discussions.* Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign.

*Kong, A., & Fitch, E. (2002/2003). Using Book Club to engage culturally and linguistically diverse learners in reading, writing, and talking about books. *The Reading Teacher, 56,* 352–362.

Kuhn, D. (1993). *The skills of argument.* Cambridge, UK: Cambridge University Press.

Lee, J., Grigg, W. S., & Donahue, P. L. (2007). *Nation's report card: Reading* (NCES 2007–496). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

*Lipman, M. (1975). *Philosophy for Children.* (ERIC Document Reproduction Service No. ED103296)

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

*Martin, J. (1998). Literature Circles. *Thresholds in Education, 24,* 15–19.

Mason, L. (1998). Sharing cognition to construct scientific knowledge in school context: The role of oral and written discourse. *Instructional Science, 26,* 359–389.

Mason, L. (2001). Introducing talk and writing for conceptual change: A classroom study. *Learning and Instruction, 11,* 305–329.

*McGee, L. M. (1992). An exploration of meaning construction in first graders' Grand Conversations. In C. K. Kinzer & D. J. Leu (Eds.), *Literacy research, theory, and practice: Views from many perspectives* (Vol. 41, pp. 177–186). Chicago: National Reading Conference.

*McGee, L. M., Courtney, L., & Lomax, R. G. (1994). Teachers' roles in first graders' Grand Conversations. In National Reading Conference (Ed.), *Multidimensional aspects of literacy research, theory, and practice* (Vol. 43, pp. 507–526). Chicago: National Reading Conference.

McKeown, M. G., & Beck, I. L. (1990). The assessment and characterization of young learners' knowledge of a topic in history. *American Educational Research Journal, 27,* 688–726.

*McKeown, M. G., Beck, I. L., Kucan, L., & Sandora, C. A. (1995). *Second-year classroom implementation of questioning the author: New teachers, new site.* Pittsburgh, PA: Author.

*McKeown, M. G., Beck, I. L., & Sandora, C. A. (1996). Questioning the Author: An approach to developing meaningful classroom discourse. In M. F. Graves, P. van der Broek, & B. M. Taylor (Eds.), *The first R: Every child's right to read* (pp. 97–119). New York: Teachers College Press.

Mead, G. H. (1934). *Mind, self, and society from the standpoint of a social behaviorist.* Chicago: University of Chicago Press.

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners.* Philadelphia: Multilingual Matters.

Mercer, N. (2002). The art of interthinking. *Teaching Thinking, 7,* 8–11.

*Mizerka, P. M. (1999). *The impact of teacher-directed literature circles versus student-directed literature circles on reading comprehension at the sixth-grade level.* Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign.

National Assessment Governing Board. (2007). *Reading framework for the 2007 National Assessment of Educational Progress.* Washington, DC: U.S. Department of Education.

National Commission of Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office.

National Reading Panel. (2000). *Teaching children to read: Reports of the subgroups* (NIH Publication No. 00–4754). Washington, DC: U.S. Department of Health and Human Services, National Institute of Child Health and Human Development.

No Child Left Behind Act of 2001. P.L. 107–110, 107th Cong. (2001).

Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom.* New York: Teachers College Press.

*Olezza, A. M. (1999). *An examination of effective instructional and social interactions to learn English as a second language in a bilingual setting.* Unpublished doctoral dissertation, University of Connecticut, Storrs.

Patthey-Chavez, G. G., & Clare, L. (1996). Task, talk, and text: The influence of Instructional Conversation on transitional bilingual writers. *Written Communication, 13,* 515–563.

Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension.* New York: Holt, Rinehart and Winston.

Piaget, J. (1928). *The child's conception of the world.* London: Routledge and Kegan Paul.

*Pitman, M. (1997). *Literature Circles.* (ERIC Document Reproduction Service No. ED416503)

Raphael, T. E., Gavelek, J. R., & Daniels, V. (1998). Developing students' talk about text: An initial analysis in a fifth-grade classroom. *National Reading Conference Yearbook, 47,* 116–128.

Raphael, T. E., & McMahon, S. I. (1994). Book Club: An alternative framework for reading instruction. *The Reading Teacher, 48,* 102–116.

*Reznitskaya, A. (2002). *The influence of group discussions and explicit instruction on the acquisition and transfer of argumentative knowledge.* Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign.

*Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32,* 155–175.

Rogers, T. (1990/1991). A point, counterpoint response strategy for complex short stories. *Journal of Reading, 34,* 278–282.

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

*Sable, E. D. (1987). *The effects of Junior Great Books literature discussion on reading comprehension achievement of gifted fifth graders: Application of general linear model for cross-level inferences.* Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg.

*Sandora, C., Beck, I., & McKeown, M. (1999). A comparison of two discussion strategies on students' comprehension and interpretation of complex literature. *Journal of Reading Psychology, 20,* 177–212.

*Saunders, W., & Goldenberg, C. (1998). The effects of an instructional conversation in Latino students' concepts of friendship and story comprehension. In R. Horowitz (Ed.), *The evolution of talk about text: Knowing the world through classroom discourse.* Newark, DE: International Reading Association.

*Saunders, W. M., & Goldenberg, C. (1999). Effects of instructional conversations and literature logs on limited- and fluent-English-proficient students' story comprehension and thematic understanding. *The Elementary School Journal, 99,* 277–301.

Sharp, A. M. (1995). Philosophy for Children and the development of ethical values. *Early Child Development and Care, 107,* 45–55.

Short, K. G., & Pierce, K. M. (Eds.) (1990). *Talking about books: Creating literate communities.* Portsmouth, NH: Heinemann.

*Solomon, S. (1990). *Developing higher order reading comprehension skills*

in the learning disabled student: A non-basal approach. Unpublished master's thesis, Nova Southeastern University, Fort Lauderdale, Florida.

Soter, A. O., & Rudge, L. (2005, April). *What the discourse tells us: Talk and indicators of high-level comprehension.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Ontario, Canada.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context.* Cambridge, UK: Cambridge University Press.

Toulmin, S. E. (1958). *The uses of argument.* Cambridge, UK: Cambridge University Press.

Trabasso, T., Secco, T., & van der Broek, P. W. (1984). Causal cohesion and story coherence. In H. Mandi, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Erlbaum.

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

Vygotsky, L. S. (1986). *Thought and language.* Cambridge, MA: The MIT Press.

Wade, S., Thompson, A., & Watkins, W. (1994). The role of belief systems in authors' and readers' constructions of texts. In R. Garner & P. A. Alexander (Eds.), *Beliefs about text and instruction with text* (pp. 265–293). Hillsdale, NJ: Erlbaum.

Waggoner, M., Chinn, C., Yi, H., & Anderson, R. C. (1995). Collaborative reasoning about stories. *Language Arts, 72,* 582–589.

Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: An empirical investigation of a possible sociocultural model of cognitive development. *Learning and Instruction, 9,* 493–516.

Wertsch, J. V., Del Rio, P., & Alvarez, A. (Eds.). (1995). *Sociocultural studies of mind.* Cambridge, UK: Cambridge University Press.

*Williams, B. G. (1999). Emergent readers and Literature Circle discussions. In J. R. Dugan, P. E. Linder, W. M. Linek, & E. G. Sturtevant (Eds.), *Advancing the world of literacy: Moving into the 21st century* (pp. 44–53). Carrollton, GA: The College Reading Association.

*Yeazell, M. I. (1982). Improving reading comprehension through philosophy for children. *Reading Psychology: An International Quarterly, 3,* 239–246.

Received January 18, 2008
Revision received February 9, 2009
Accepted February 16, 2009 ■